Professional and Practice-based Learning

Michael Goller
Eva Kyndt
Susanna Paloniemi
Crina Damşa *Editors*

Methods for Researching Professional Learning and Development

Challenges, Applications and Empirical Illustrations

MOREMEDIA

Springer

Professional and Practice-based Learning

Volume 33

Professional and practice-based learning brings together international research on the individual development of professionals and the organisation of professional life and educational experiences. It complements the Springer journal *Vocations and Learning: Studies in vocational and professional education*.

Professional learning, and the practice-based processes that often support it, are the subject of increased interest and attention in the fields of educational, psychological, sociological, and business management research, and also by governments, employer organisations and unions. This professional learning goes beyond, what is often termed professional education, as it includes learning processes and experiences outside of educational institutions in both the initial and ongoing learning for the professional practice. Changes in these workplaces requirements usually manifest themselves in the everyday work tasks, professional development provisions in educational institution decrease in their salience, and learning and development during professional activities increase in their salience.

There are a range of scientific challenges and important focuses within the field of professional learning. These include:

- understanding and making explicit the complex and massive knowledge that is required for professional practice and identifying ways in which this knowledge can best be initially learnt and developed further throughout professional life.
- analytical explications of those processes that support learning at an individual and an organisational level.
- understanding how learning experiences and educational processes might best be aligned or integrated to support professional learning.

The series integrates research from different disciplines: education, sociology, psychology, amongst others. The series is comprehensive in scope as it not only focuses on professional learning of teachers and those in schools, colleges and universities, but all professional development within organisations.

* * *

Please contact Grace Ma at grace.ma@springer.com if you wish to discuss a book proposal.

Michael Goller • Eva Kyndt
Susanna Paloniemi • Crina Damşa
Editors

Methods for Researching Professional Learning and Development

Challenges, Applications and Empirical Illustrations

Springer

Editors
Michael Goller
Institute of Educational Science
Paderborn University
Paderborn, Germany

Eva Kyndt
Centre for Transformative Innovation
Swinburne University of Technology
Melbourne, VIC, Australia

Susanna Paloniemi
Department of Education
University of Jyväskylä
Jyväskylä, Finland

Crina Damşa
Department of Education
University of Oslo
Blindern, Oslo, Norway

This work contains media enhancements, which are displayed with a "play" icon. Material in the print book can be viewed on a mobile device by downloading the Springer Nature "More Media" app available in the major app stores. The media enhancements in the online version of the work can be accessed directly by authorized users.

ISSN 2210-5549 ISSN 2210-5557 (electronic)
Professional and Practice-based Learning
ISBN 978-3-031-08517-8 ISBN 978-3-031-08518-5 (eBook)
https://doi.org/10.1007/978-3-031-08518-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Series Editors' Foreword

This book series addresses a variety of research approaches on professional and practice-based learning. Thus, it uncovers the plethora of possibilities of how working can be conceived as a rich source for learning by assembling different theoretical perspectives with different methodological paradigms from various areas of professional practices.

For most individuals, work is an essential part of life that affords and provides learning experiences. Schools, universities, and vocational education and training can only provide a starting point for an individual's lifelong professional career. As societal, economic, and technological developments continuously permeate one's work life, individuals need to adapt to changes and develop their skills and capabilities, most frequently beyond educational settings. Learning through practice and work is of crucial significance for individuals as a tool to cope with changing work environments and requirements.

A side effect of the strong development of research on learning and professional development is that a multitude of new research perspectives and innovative research methods emerged within the last decade. One of the significant triggers of these developments certainly is that new technologies have been increasingly used to explore new and effective methods of data collection. Accelerated progress can especially be observed in the collection and analysis of complex online measures regarding learning process methods. Thus, data can be extracted that directly captures the execution of occupational practices. Novel kinds of data often require novel methods of analysis, and it is an ongoing challenge to connect those online measures to more traditional kinds of data.

This volume provides an overview on the current state of empirical and methodological approaches of researching professional learning. In separate sections, new developments concerning data collection, data analysis, and research paradigms are outlined. In these sections, the empirical methods and procedures will not only be characterised but also provided with examples from research in the field of professional learning. Hence, this volume mainly addresses researchers from all disciplines who are interested in investigating professional learning and development.

Griffith University, Brisbane, Australia Stephen Billett

University of Regensburg Hans Gruber
Regensburg, Germany

Paderborn University Christian Harteis
Paderborn, Germany

Contents

Chapter 1
Addressing Methodological Challenges in Research on Professional Learning and Development

Michael Goller, Eva Kyndt, Susanna Paloniemi, and Crina Damşa

Abstract Research in the area of professional learning and development is faced with particular empirical and methodological challenges due to its nature and contexts. This chapter introduces and briefly describes these challenges. It then gives an overview of each of the methods/approaches (i.e., chapters) in this book in relation to the identified challenges. The chapter ends with a presentation of the overall structure of the book.

Keywords Professional learning and development · Workplace learning · VET · Research methods

1.1 Professional Learning and Development as a Research Field

Professional learning and development (PLD) is concerned with the processes through which professionals or future professionals acquire, maintain or update their personal capacities to adequately deal with tasks and problems at their current or future workplaces (e.g., Grosemans et al., 2017; Gruber & Harteis, 2018; Hilkenmeier et al., 2021). This learning and development can take place through

M. Goller (✉)
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: michael.goller@upb.de

E. Kyndt
Centre for Transformative Innovation, Swinburne University of Technology, Melbourne, VIC, Australia
e-mail: ekyndt@swin.edu.au

S. Paloniemi
Department of Education, University of Jyväskylä, Jyväskylä, Finland
e-mail: susanna.paloniemi@jyu.fi

C. Damşa
Faculty of Educational Sciences, University of Oslo, Oslo, Norway
e-mail: crina.damsa@iped.uio.no

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_1

participation in designated learning activities that are explicitly planned for competence acquisition (e.g., vocational and professional education, trainings, workshops, simulations) as well as engagement in work-related practices that focus primarily on work results (e.g., problem-solving, reflection of work situations, collaborating and interacting with other actors at work) (see Hilkenmeier et al., 2021; Tynjälä, 2013). The latter is often described as workplace learning (Billett, 2004; Kyndt & Beausaert, 2017; Tynjälä, 2008). PLD outcomes can be located both at the level of the individual, such as knowledge, skills, abilities, and motivational orientations (sometimes summarised as competencies), or that of the employing organisation, such as improved quality or productivity of work processes and organisational development (e.g., Smet et al., 2022; Tynjälä, 2013).

The field of PLD is broad and ranges from vocational education and training (VET) to various workplace learning contexts. Scholars interested in PLD generally aim to understand (a) how PLD mechanisms and processes are manifested and unfold; (b) what phenomena trigger PLD; (c) what individual or contextual factors foster or hinder PLD; (d) how PLD is manifested in the relations and transitions between VET and working life; (e) how PLD can be best supported by purposefully designing learning activities or restructuring workplaces; (f) how PLD requirements are affected by societal, economic and technological changes, and (g) the outcomes it yields for individuals, organisations, and society more broadly. Such research questions are approached from the perspectives of a range of research paradigms and disciplines (e.g., Billett et al., 2014; Noe & Ellingson, 2017), leading to a wide variety of conceptual and theoretical frameworks being employed to describe and explain PLD (e.g., Billett, 2001; Cairns, 2021; Hager, 2013; Tynjälä, 2013). Although this diversity can be seen as an advantage, it also raises challenges as research strands are only loosely connected, and similar phenomena are described using disparate terminology (e.g., Fenwick, 2006).

Besides these conceptual and theoretical issues, PLD research also faces particular empirical and methodological challenges due to its nature and contexts. The majority of PLD-related empirical investigations still tend to use a rather limited set of established research methods. On the qualitative research side, interviews have been the predominant choice for data collection and have been analysed through thematic or content analysis (e.g., Boud & Middleton, 2003; Sandberg, 2010; Taylor, 2017). Although new qualitative methodologies have emerged in the field in the last few years, such as design-based research studies, ethnographic studies and video-based studies, they have not been mainstreamed. Quantitative researchers, on the other hand, most frequently employ (cross-sectional) questionnaire studies, which they analyse using a range of classical statistical correlation methods based on either manifest or latent constructs (e.g., Beausaert et al., 2013; Goller et al., 2020; Renta Davis et al., 2017). Although these methods continue to be of great relevance, other more recent methods have not been extensively deployed in PLD research. The development of newer methods has mainly been triggered by the availability of new digital technologies and computational means – which allow the collection and processing of complex sets of both small and large amounts of data – or emerging theorisations (e.g., posthumanism). Other methodological advancements can be traced back to the general trend of digitalisation in work settings that requires new ways through which to investigate how employees learn and develop under such altered circumstances (e.g., by adapting existing methods to digital contexts).

The scientific community concerned with PLD has gradually adopted these methods, although not at the same pace as other related research areas (see, e.g., Cortina & Landis, 2013; Donche et al., 2015). We argue that scholars working in the field of PLD have not yet exploited the full range of available empirical research methods. This may be seen as unfortunate since these less established and, in certain cases, more sophisticated methods are suitable to address empirical phenomena and challenges at the core of this field, which are yet to be solved by the predominant research methods.

In fact, a number of existing methods could be useful in PLD research but have only been applied by a handful of PLD researchers. With this book, we aim to introduce and discuss these methods. We do not claim to provide an exhaustive overview but aim to present at least one alternative or innovative approach to certain challenges. For this purpose, Sect. 1.2 in this introduction describes and discusses the challenges connected to research in the PLD context. Section 1.3 then gives an overview of each of the methods/approaches (i.e., chapters) in this book in relation to the identified challenges, and finally, Sect. 1.4 presents the overall structure of the book.

1.2 Challenges Connected to PLD Research

Research in the PLD context focuses on learning and development that allow professional actors to adequately tackle highly specific work-related tasks and problems. As foreshadowed, such learning and development processes take place in a variety of settings that span from school-like apprenticeship classes in the context of VET, smaller training courses that follow adult-education frameworks, simulation settings or designated training centres (e.g., skillslabs in hospitals), virtual learning contexts (e.g., massive open online courses: MOOCs) to the workplace itself. The combination of such a high variety of learning and development settings alongside a focus on diversity regarding the distinct competencies required in different work domains inevitably leads to a research field that is highly heterogeneous in nature. This becomes clear in comparison to learning and development occurring in school contexts in which all students of a certain grade level follow the same curriculum and are taught by teachers who themselves received a similar professional education. Generally speaking, the population of students attending schools is well-defined, and teaching happens in a comparable way (classroom settings, usually one teacher per class, about 15–30 students, similar educational goals). In stark contrast, PLD is specific to particular professional domains and sometimes even organisations, making the population of potential research participants much smaller and less well-defined. In addition, PLD research also has to deal with a variety of organisations that are often unwilling to allow researchers to conduct research, let alone implement more complex research designs such as experiments with control groups.

Furthermore, learning and development processes that take place directly at work (i.e., workplace learning) are usually tacit, diverse and complex in nature and, therefore, difficult to measure. Finally, the world of work is usually characterised by dynamism that leads to ever-changing work circumstances and corresponding competence requirements (e.g., Harteis & Goller, 2014). In the following, we outline the empirical challenges faced by researchers in the field of PLD and how they are currently being dealt with. We acknowledge that the challenges mentioned here are not exclusive to PLD research but are at its very core and that similar research challenges might, for instance, occur in research on industrial, work and organisational psychology, organisational behaviour, ergonomics and human factors (see, e.g., Golombiewksi, 2001; Ones et al., 2018a, b, c; Wilson & Sharples, 2015). In what follows, the identified challenges will be numbered to make them more accessible to readers. However, this numbering does not imply a level of importance or even an exhaustive description.

The first challenge faced by PLD researchers is the difficulty involved in obtaining large samples, which are often a requirement for drawing inferences and generalising findings when using traditional methods. *Small samples* can result from a small number of potential participants, for example, when researching small and mid-sized organisations (which make up the majority of organisations in many industrialised countries) or training which is typically delivered to smaller groups in professional settings (in contrast to higher education settings). However, even when collaborating with larger organisations or groups, it remains challenging to obtain a high number of responses. The first hurdle is to convince the organisation to invest time (both management and employees), which is not an easy task given that employee time is often one of the biggest organisational expenditures. Second, and especially in Europe, the General Data Protection Regulation (GDPR) makes it difficult to work with existing data sources such as performance results. Finally, the individual employee also needs to be convinced to spend time providing data while they themselves often do not receive a direct return that is of value to them. It might even be the case that data collection could hinder employees from doing their jobs for a certain period, requiring them to work harder after the study, or they could be impeded from making relevant work progress. This challenge, along with others discussed below, partly explains the fact that qualitative studies, and especially interview studies, are popular in the field of PLD.

Second, it is very common to find that observations are not independent of each other as participants are clustered in teams, organisations, training courses, sectors, industries or a combination thereof. As such, PLD research often needs to deal with *nested samples* and either account for this nestedness or explain its effects. Current approaches include both (qualitative or quantitative) multiple case studies and multilevel analyses of quantitative data. However, these methods still pose challenges when it comes to understanding interactions between actors, especially for quantitative studies that require sufficiently large sample sizes at all levels of analysis. For example, to reap the true benefits of multilevel analyses, having a sufficiently high number of individuals is not enough on its own; these individuals also need to stem from a sufficient number of randomly sampled organisations that are often difficult to acquire (for a basic introduction to multilevel analysis in the context of PLD, see Kyndt & Onghena, 2014).

Third, the effectiveness and outcomes of PLD are an important area of interest in the field and are best studied using *experimental designs that include control groups* (with randomised controlled trials as the pinnacle). Unfortunately, for similar reasons, PLD researchers face difficulty collecting data from large samples because gaining access to a control group can be challenging. In addition, and even more importantly, PLD and especially workplace learning cannot be separated from the work context or activities. As such, even if it is practically possible to set up a randomised controlled trial, it would not be suitable to study topics such as workplace learning due to the lack of (ecological) validity. As such, many PLD researchers often opt for a pre-post-test design and the use of surveys.

Fourth, as work and professional learning are intertwined, *workplace learning processes* are often *tacit in nature*. While traditional in-depth interviews, focus groups, etc. have yielded very interesting insights into workplace learning, it remains difficult to truly measure tacit learning processes as participants are often (no longer) unaware that they are/were in fact learning (Eraut, 2004; Rausch, 2014). In some cases, participants might even be fully convinced that any development based on work experience is not something to be subsumed under the umbrella of learning. Consequently, researchers cannot use everyday vocabulary connected to learning and development within their measurement instruments, which could easily threaten the validity of PLD studies.

Fifth, next to being tacit, *(workplace) learning processes are complex and dynamic in nature*. In other words, not only is it unclear when and how learning processes unfold in work contexts, they do not have clear start and end points and might be strongly intertwined with all kinds of processes taking place at work. While longitudinal survey studies are suitable for studying dynamics over time, to date, a lack of insights offering clear guidance regarding when, how often and with what time lag to measure PLD is still a typical concern for both qualitative and quantitative researchers. After all, workplace learning might occur for an employee several days in a row due to a combination of emerging problems and sufficient time to deal with them but might be non-existent for a few weeks thereafter.

Sixth, *professional domains and workplaces are inherently heterogeneous*. For instance, the work of a carpenter is fundamentally different from that of someone on an assembly line or someone working in retail. Connected to this heterogeneity is the fact that the required competences are naturally different between professional domains and that workplaces and learning processes are characterised by major differences (see also Billett, 2001). While workplace learning in some domains is based predominantly on observing more experienced peers and mimicking their behaviour, this is not possible in other domains in which the work being done is hardly observable (e.g., when employees work alone or the work-related phenomena are opaque or completely hidden from the learner; see Billett, 2014; Goller, 2017).

Another example is learning through active experimentation (i.e., trial and error), which might be possible in some domains but not in others (e.g., aeroplane pilots). Researchers have tried to tackle these issues using multiple case-studies or multilevel analyses with random level-2 effects.

Seventh, work itself is often directly affected by *current technological changes*. Due to economic reasons, organisations acting in free markets are under pressure to rapidly adopt technological developments to remain competitive. It follows that work, generally speaking, has become increasingly less routine as product and process life cycles have shortened significantly over the last 30 years (Billett, 2009; Green, 2007; Harteis & Goller, 2014). This development poses two connected challenges. First, PLD research needs to account for these changes. On one hand, existing methods need to be adapted to new working contexts (e.g., digitalised work). One the other hand, new or existing methods need to tap into newly emergent research potentials (e.g., big data; see next challenge). Second, PLD research needs to support organisations and employees by generating insights into how certain developments will change the world of work, including competence requirements, learning potential and so on.

The eighth and final challenge is emergent but rapidly gaining prominence. Due to new types of professional practices enabled by technology, PLD researchers are confronted with an *abundance of (unused and often ill-structured) data* gathered by many organisations, but it is not always clear what they mean or how to use them. However, traditional statistical analyses (such as covariance-based structural equation modelling, ordinary least square regression, analysis of variance, etc.) are not equipped to deal with high volumes of data, and established theories might not be able to grasp the complexity of the reality that is being captured with these data, leaving potential insights undetected.

1.3 Methodological Advancements: Alternative Approaches to Challenges

As mentioned earlier, we believe that the field of PLD might benefit from adopting emerging and alternative methods (borrowed from other disciplines or developed within the field itself) that enable researchers to address empirical phenomena and challenges at the core of this field and that have yet to be solved by more common research methods. In a similar vein, the lack of methodological innovativeness within PLD research is increasingly acknowledged and discussed within the relevant scientific community (Harteis, 2017; Harteis et al., 2018; Kyndt & Goller, 2017). Therefore, this book brings together innovative, emerging and/or less common methods and approaches in the field of PLD. Below, we briefly discuss the different methods and/or approaches included in this volume because they have the potential to address the challenges presented above.

The first challenge pertained to the fact that it is common to have *small samples and an absence of normality*, which means that data sets are often unsuitable for a range of analyses that have certain assumptions or that it is difficult to make generalisations. However, both the Bayesian approach and PLS-based structural equation modelling (SEM) enable researchers to quantitatively *analyse* small datasets or datasets that are not multivariate normally distributed. *Bayesian statistics* (Nokelainen et al., 2022, Chap. 10) can be used both as an alternative to more classical analysis techniques (e.g., t-tests) and as an estimation method for more complex models (e.g., hierarchical linear models) with comparably smaller datasets than classic frequentist methods. *PLS-based SEM* (Goller & Hilkenmeier, 2022, Chap. 12) is an approach used to test complex relationships between different unobserved constructs; it does not rely on certain distributional assumptions, such as classical SEM, and has various other advantages. In terms of *research approaches*, *Q method* (Leidig et al., 2022, Chap. 20) makes use of very small samples to find clusters of similarly thinking people – that is, individuals with similar subjectivities.

The second challenge often faced by PLD researchers is how to deal with *nested samples*. As outlined earlier, quantitative studies often use multilevel analyses when dealing with nested samples; however, this is not always possible due to sample-size requirements at each level (e.g., sufficient numbers of individuals AND teams AND organisations in a three-level design). However, when these requirements are not met, both *Bayesian statistics* (Nokelainen et al., 2022, Chap. 10) and *PLS-based SEM* (Goller & Hilkenmeier, 2022, Chap. 12) might provide an alternative as the former does not rely on large-sample theory and the latter is relatively robust to violations of independence. Nevertheless, the nestedness of samples can also be a strength and not just something to account for, as it might present an opportunity to study interpersonal interactions in greater depth, for example, by using *social network analysis* (Palonen, 2022, Chap. 22). In addition, social network analysis might help make visible the nested structures within an organisation. Furthermore, David et al. (2022, Chap. 9) make the case for a temporal research approach to studying *interaction dynamics* among members nested within teams.

A third challenge in PLD research is the *reduced possibility for experiments and control groups*. However, there are several potential alternatives that are still relatively rare in PLD research that do allow researchers to make the contribution of interventions visible. In the context of questionnaire studies, the *vignette technique* (Anselmann & Mulder, 2022, Chap. 4) allows researchers to define different work-related stimuli, which are then presented to study participants. In this way, research designs can be developed where participants react to a simulated control and an experimental setting. More on the qualitative side, *(virtual) ethnography* (Lemmetty et al., 2022, Chap. 18) allows researchers to capture learning processes at work as they unfold in their natural (digital) environments; *design-based research* (Gerholz & Wagner, 2022, Chap. 23) offers a structured yet flexible way in which to construct interventions with a focus on the iterative development of practices; and *change laboratory* (Kajamaa & Hyrkkö, 2022, Chap. 24) equips researchers with a toolkit for in-depth examinations of processes and tensions that are not usually revealed through the simple use of post-factum self-reports. Besides focusing on the core elements of learning and development in work organisations (e.g., the nature of learning, work practices and processes, the learning culture), developmental aspects have become the foci of qualitative research. So far, design-based research and cultural historical activity theory-based change laboratories have been the main approaches in the research and development of work practices and work organisations. They are now being accompanied by novel applications such as interventionist designs combined with ethnographic approaches. This, we believe, is also partly related to the challenges of gaining research collaboration and developing professional learning practices in work organisations. Interestingly, PLD research itself can become an important site for the practice of PLD.

In recent years, many researchers have focused on tackling the fourth challenge and made progress in rendering the *tacit nature of (workplace) learning processes* more visible. While there is no quick fix to this challenge, there are several methods or research approaches that can contribute to making implicit processes more explicit. Two methods and approaches discussed in this volume aim to tackle this challenge through repeated measurements that take place directly in work contexts: *experience sampling* (Seifried & Rausch, 2022, Chap. 2) and *diaries* (Rausch et al., 2022, Chap. 3). Two other methods, namely *video-based interaction analysis* (Filliettaz et al., 2022, Chap. 19) and *video annotations* (Steffen & Pouta, 2022, Chap. 13), use videographic data to understand how learners perceive work-related situations and coordinate their professional actions in the context of social encounters at work. Three other methods tap into data resulting directly from learners' doing. In the chapters about *eye tracking* (Jossberger, 2022, Chap. 21) and *log-file data* (Spliethoff & Abele, 2022, Chap. 8), such doing is directly connected to problem-solving. In contrast, in the chapter on *learning analytics* (Littlejohn et al., 2022, Chap. 25), the data are generated from individuals' learning based on their activities in the context of digital learning management systems. In addition, data collection methods involving psychophysiological indicators, such as *electrodermal activity* (Paloniemi et al., 2022, Chap. 5), *cortisol levels* (Kärner & Sembill, 2022, Chap. 6) and *ECG and EEG* (Silvennoinen et al., 2022, Chap. 7), are being explored as a way to capture the multimodality of PLD. What is common in these rare research designs is that various and selected psychophysiological indicators are integrated in the context of research on experiential learning processes either in laboratory settings or everyday work contexts. Importantly, the development of technological devices and applications has paved the way for this development.

There has been growing criticism of cross-sectional studies as it has increasingly been recognized that it is important to capture the *dynamic and complex nature of (workplace) learning processes*. In terms of data collection, *log-file data* (Spliethoff & Abele, 2022, Chap. 8), *learning analytics* (Littlejohn et al., 2022, Chap. 25), *experience sampling* (Seifried & Rausch, 2022, Chap. 2) and *diaries* (Rausch et al., 2022, Chap. 3) are aimed at capturing both the dynamics and complexity of learning. In terms of data analysis, person-centred analyses such as *latent profile and latent* *cluster analysis* (Bauer, 2022, Chap. 11) are specifically interesting when analysing individual differences in prerequisites, processes or outcomes of learning, while *visual analytics* (Kyndt & Aerts, 2022, Chap. 15) are particularly suitable for identifying unknown patterns in complex data sets. Research approaches that are considered to tackle this challenge in a holistic way include *longitudinal multiple case studies* (Cuyvers et al., 2022, Chap. 26) and a temporal research approach to studying *interaction dynamics* among team members (David et al., 2022, Chap. 9).

Workplaces and connected PLD processes are *inherently heterogeneous* in nature. While a few approaches are built on this heterogeneity and are interested in generating rich descriptions about it, namely the *narrative approach* (Vähäsantanen & Arvaja, 2022, Chap. 17), *virtual ethnography* (Lemmetty et al., 2022, Chap. 18) and *video-based interaction analysis* (Filliettaz et al., 2022, Chap. 19), others are mainly aimed at making this heterogeneity visible and/or explaining it, including *diaries* (Rausch et al., 2022, Chap. 3), *experience sampling* (Seifried & Rausch, 2022, Chap. 2), *log-file data* (Spliethoff & Abele, 2022, Chap. 8), *Q method* (Leidig et al., 2022, Chap. 20), *latent profile and latent cluster analysis* (Bauer, 2022, Chap. 11), *data mining* (Ifenthaler, 2022, Chap. 14), *visual analytics* (Kyndt & Aerts, 2022, Chap. 15) and *learning analytics* (Littlejohn et al., 2022, Chap. 25). Other methods allow study participants to be exposed to standardised stimuli such as work situations that reduce heterogeneity. These include *vignettes* (Anselmann & Mulder, 2022, Chap. 4) and *video annotations* (Steffen & Pouta, 2022, Chap. 13). Finally, *Bayesian statistics* (Nokelainen et al., 2022, Chap. 10) can be used to deal with heterogeneous data, even in the context of small sample sizes.

Technological and economic developments have the potential to quickly *change workplaces* and learning and development requirements for many professional actors. Information about such unknown futures is important but difficult to come by. One approach that aims to generate data on potential future developments is to ask experts and knowledgeable stakeholders through the *Delphi method* (Harteis, 2022, Chap. 16).

Finally, we identified the *abundance of (unused) data* stemming from new types of professional practices enabled by technology as an emerging challenge. While there are many challenges around this, including privacy laws and validity issues, in terms of data analysis, *data mining* (Ifenthaler, 2022, Chap. 14), *visual analytics* (Kyndt & Aerts, 2022, Chap. 15), and *learning analytics* (Littlejohn et al., 2022, Chap. 25) all offer opportunities to seize the potential offered by these data. Table 1.1 gives an overview of the content of each chapter in terms of the above-identified research challenges.

1.4 Structure of the Book

The book is structured in three main parts comprising chapters on the various research methods and/or approaches identified above. The first part contains chapters that are mainly concerned with methods of data collection (e.g., diary method,

Table 1.1 Research challenges and the research methods that might be suited to tackling them

Challenge	Chapters
Small sample sizes	**10 Bayesian statistics** (allow researchers to deal with small samples in an elaborated way); **12 PLS-based SEM** (allows researchers to estimate SEM in relatively small samples); **20 Q method** (makes use of very small samples to find groups of similarly thinking people)
Nested sample structure	**9 Interaction dynamics in team learning** (explicitly focuses on teams as nested data structures); **10 Bayesian statistics** (allows researchers to deal with small nested samples); **12 PLS-based SEM** (relatively robust against violations of assumptions); **22 Social network analysis** (can make the nested structure within an organisation visible);
Reduced possibilities to use experimental designs with control groups	**4 Vignettes** (introduce standardised theoretical work settings that study participants can react to); **18 Virtual ethnography** (provides insights into processes that take place online); **23 Design-based research** (reveals iterations between and formative aspects of interventions); **24 Change laboratory** (reveals features of formative interventions)
Tacit nature of workplace learning	**2 Experience sampling method** (makes implicit processes explicit and explains them); **3 Diaries** (make implicit processes explicit and explain them); **5 EDA** (makes emotions and their relationship with learning visible); **6 Psychophysiological assessments** (make stress and its relationship with learning visible); **7 Physiological measurements** (make emotions and their relationship with learning visible); **8 Log-file data** (generate objective process data about learners' problem-solving); **13 Video annotations** (combine advantages from concurrent reporting and retrospective recall to make implicit processes visible); **19 Video-based interaction analysis** (makes implicit interaction processes visible); **21 Eye tracking** (generates objective process data about learners' problem-solving); **25 Learning analytics** (make implicit learning processes visible)
Dynamic and complex nature of (workplace) learning processes	**2 Experience sampling method** (captures dynamics and complexity); **3 Diaries** (capture dynamics and complexity); **8 Log-file data** (capture dynamics and complexity); **9 Interaction dynamics in team learning** (capture dynamics and complexity); **11 Latent profile and class analysis** (capture individual differences in prerequisites, processes or outcomes of learning); **15 Visual analytics** (find unknown patterns in large data sets); **25 Learning analytics** (capture dynamics); **26 Longitudinal multiple case study** (collects holistic longitudinal data)
Heterogeneity between workplaces and professional learning and development processes	**2 Experience sampling method** (make heterogeneous learning processes visible); **3 Diaries** (make heterogeneous learning processes visible); **4 Vignettes** (introduce standardised theoretical work settings that study participants can react to); **8 Log-file data** (make heterogeneity visible); **10 Bayesian statistics** (allow researchers to deal with heterogeneous data based on small samples); **11 Latent profile and class analysis** (make heterogeneity visible); **13 Video annotations** (introduce standardised theoretical work settings that study participants can react to); **14 Data mining** (makes heterogeneity visible); **15 Visual analytics** (make heterogeneity visible); **17 Narrative approach** (takes heterogeneity into account and extracts patterns); **19 Video-based interaction analysis** (allows researchers to take heterogeneity into account); **20 Q method** (makes heterogeneity visible); **25 Learning analytics** (make heterogeneous learning processes visible)

(continued)

Table 1.1 (continued)

Challenge	Chapters
Technological changes that have to be accounted for	**16 Delphi technique** (Generates information about unknown futures)
Abundance of (unused) data	**14 Data mining** (makes use of existing data); **15 Visual analytics** (find unknown patterns in large data sets); **25 Learning analytics** (make use of existing data and data sources)

vignettes in surveys, psychophysiological measures). The second part includes chapters describing methods for analysing different types of data (e.g., Bayesian analysis, PLS-based SEM, data mining). The third part comprises chapters describing research approaches. Research approaches are systematic and holistic ways of gathering data that are then used in specific data analysis techniques (e.g., design-based research, eye tracking or the Q method). Thus, while the first two parts of the book follow the general structure of an empirical research project (i.e., initial data collection and subsequent data analysis), the third part presents methodological approaches that cannot be categorised as data collection or data analysis and, instead, holistically define how to both gather data and make sense of it. Overall, the structure of the book illustrates the different epistemological and ontological premises within the research field.

Each of the chapters in this volume explains a particular method from a conceptual perspective and offers empirical examples (often by presenting a study conducted using this method). The focus is on the purpose, advantages, disadvantages, and practical principles of the method being introduced. The empirical examples contribute to illustrating how the more technical and conceptual descriptions can be applied within the field of PLD. In this way, each chapter is similarly accessible to readers of a range of knowledge backgrounds. Finally, two experienced researchers from the domain of PLD offer their reflections on the featured approaches and identify their potential future methodological applications in the two concluding chapters of this book. Monika Nerland's contribution (2022, Chap. 27) comprehensively discusses the different chapters of the book from a qualitative research perspective. She also discusses PLD research challenges in the context of emerging trends such as digitalisation tendencies. Similarly, Erno Lehtinen (2022, Chap. 28) comments on the challenges of PLD research, albeit with a focus on workplace learning. In his contribution, he also comments on how the different methods described in the book might help overcome some of the challenges he perceives within the research field.

To sum up, this book aims to offer a comprehensive collection of methods that are well equipped to tackle particular research challenges specific to PLD-related investigations. By doing so, it develops future directions for research on PLD. The book supports scholars in understanding upcoming empirical research and methods and encourages novice as well as experienced scholars to adopt new empirical strategies beyond well-established and widely used ones. This is especially important since certain methods are becoming increasingly more complex and differentiated; as such, even scholars within a certain research account might not be able to easily catch up with certain developments and, therefore, fully understand what other researchers within their field are doing. It follows that the book will be a useful resource for both students and scholars interested in empirical research in the field of PLD.

References

Anselmann, V., & Mulder, R. (2022). Using the vignette technique to increase insight into professional development at work. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 71–86). Springer.

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 243–268). Springer.

Beausaert, S., Segers, M., Fouarge, D., & Gijselaers, W. (2013). Effect of using a personal development plan on learning and development. *Journal of Workplace Learning, 25*(3), 145–158.

Billett, S. (2001). Knowing in practice: Re-conceptualising vocational expertise. *Learning and Instruction, 11*(6), 431–452. https://doi.org/10.1016/S0959-4752(00)00040-2

Billett, S. (2004). Workplace participatory practices: Conceptualising workplaces as learning environments. *Journal of Workplace Learning, 16*(6), 312–324. https://doi.org/10.1108/13665620410550295

Billett, S. (2009). Changing work, work practice: The consequences for vocational education. In R. Maclean & D. Wilson (Eds.), *International handbook of education for the changing world of work. Bridging academic and vocational learning* (pp. 175–187). Springer.

Billett, S. (2014). *Mimetic learning at work: Learning in the circumstances of practice*. Springer.

Billett, S., Harteis, C., & Gruber, H. (Eds.). (2014). *International handbook of research in professional and practice-based learning*. Springer.

Boud, D., & Middleton, H. (2003). Learning from others at work: Communities of practice and informal learning. *Journal of Workplace Learning, 15*(5), 194–202.

Cairns, L. (2021). Learning and work: Theories and developments. In M. Malloch, L. Cairns, K. Evans, & B. O'Connor (Eds.), *The SAGE handbook of learning and work* (pp. 5–33). SAGE.

Cortina, J. M., & Landis, R. S. (2013). *Modern research methods for the study of behavior in organizations*. Routledge.

Cuyvers, K., Van den Bossche, P., & Donche, V. (2022). Longitudinal case study research to study self-regulation of professional learning: Combining observations and stimulated recall interviews throughout everyday work. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 579–600). Springer.

David, L., Endedijk, M., & Van den Bossche, P. (2022). Investigating interaction dynamics: A temporal approach to team learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 187–209). Springer.

Donche, V., De Maeyer, S., Gijbels, D., & van den Bergh, H. (2015). *Methodological challenges in research on student learning*. Garant Publishers.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273. https://doi.org/10.1080/158037042000225245

Fenwick, T. (2006). Tidying the territory: Questioning terms and purposes in work-learning research. *Journal of Workplace Learning, 18*(5), 265–278. https://doi.org/10.1108/13665620610674953

Filliettaz, L., Garcia, S., & Zogmal, M. (2022). Video-based interaction analysis: A research and training method to understand workplace learning and professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 419–440). Springer.

Gerholz, K.-H., & Wagner, A. (2022). Design-based research – Grounding, understanding and empirical illustration in the context of vocational education. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 513–534). Springer.

Golembiewski, R. T. (Ed.). (2001). *Handbook of organizational behavior* (2nd rev. expanded ed.). Marcel Dekker.

Goller, M. (2017). *Human agency at work: An active approach towards expertise development*. Springer.

Goller, M., & Hilkenmeier, F. (2022). PLS-based structural equation modelling: An alternative approach to estimating complex relationships between unobserved constructs. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 269–292). Springer.

Goller, M., Harteis, C., Gijbels, D., & Donche, V. (2020). Engineering students' learning during internships: Exploring the explanatory power of the job demands-control-support model. *Journal of Engineering Education, 109*(2), 307–324. https://doi.org/10.1002/jee.20308

Green, F. (2007). *Demanding work: The paradox of job quality in the affluent economy*. Princeton University Press.

Grosemans, I., Coertjens, L., & Kyndt, E. (2017). Exploring learning and fit in the transition from higher education to the labour market: A systematic review. *Educational Research Review, 21*, 67–84. https://doi.org/10.1016/j.edurev.2017.03.001

Gruber, H., & Harteis, C. (2018). *Individual and social influences on professional learning: Supporting the acquisition and maintenance of expertise*. Springer.

Hager, P. (2013). Theories of workplace learning. In M. Malloch, L. Cairns, K. Evans, & B. O'Conner (Eds.), *The SAGE handbook of workplace learning* (pp. 17–31). Sage.

Harteis, C. (2017, August 31). *New measurements of professional learning: On the explanatory power of new data collection methods*. An invited symposia at the EARLI2017 conference, Tampere, Finland.

Harteis, C. (2022). Delphi-technique as a method for research on professional learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 351–371). Springer.

Harteis, C., & Goller, M. (2014). New skills for new jobs: Work agency as a necessary condition for successful lifelong learning. In T. Halttunen, M. Koivisto, & S. Billett (Eds.), *Promoting, assessing, recognizing and certifying lifelong learning* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-8694-2_3

Harteis, C., Kok, E., & Jarodzka, H. (2018). The journey to proficiency: Exploring new objective methodologies to capture the process of learning and professional development. *Frontline Learning Research, 6*(3), 1–5.

Hilkenmeier, F., Goller, M., & Schaper, N. (2021). The differential influence of learner factors and learning context on different professional learning activities. *Vocations and Learning, 14*(3), 411–438. https://doi.org/10.1007/s12186-021-09266-4

Ifenthaler, D. (2022). Data mining and analytics in the context of workplace learning: Benefits and affordances. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 313–327). Springer.

Jossberger, H. (2022). Eye tracking in professional learning and development: Uncovering expertise development among residents in radiology. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 467–489). Springer.

Kajamaa, A., & Hyrkkö, S. (2022). Change laboratory method for facilitating transformative agency and collective professional learning – Case from a Finnish elementary school. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 535–556). Springer.

Kärner, T., & Sembill, D. (2022). Multimodal stress assessment in working and learning contexts: Using physiological, observational and experience-based data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 111–136). Springer.

Kyndt, E., & Aerts, J. (2022). Addressing 'wicked problems' using visual analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 329–348). Springer.

Kyndt, E., & Beausaert, S. (2017). How do conditions known to foster learning in the workplace differ across occupations? In R. Noe & J. Ellingson (Eds.), *Autonomous learning in the workplace* (pp. 201–218). Taylor & Francis.

Kyndt, E., & Goller, M. (2017, April 30). *Methodological advancements in quantitative research on work-related learning and professional development.* A thematic roundtable at the AERA annual meeting, San Antonio, USA.

Kyndt, E., & Onghena, P. (2014). Hierarchical linear models for researching professional learning: Relevance and implications. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional practice-based learning* (pp. 337–368). Springer.

Lehtinen, E. (2022). How to deal with the complexity in research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 619–627). Springer.

Leidig, S., Köhler, H., Caruso, C., & Goller, M. (2022). Q Method: Assessing subjectivity through structured ranking of items. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 441–466). Springer.

Lemmetty, S., Collin, K., Glăveanu, V., & Paloniemi, S. (2022). Capturing actions of communities: Towards virtual ethnography and digital tools in researching organizations and workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 397–418). Springer.

Littlejohn, A., Kennedy, E., & Laurillard, D. (2022). Professional learning analytics: Understanding complex learning processes through measurement, collection, analysis, and reporting of MOOC data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 557–578). Springer.

Nerland, M. (2022). Researching professional learning in complex environments: Opportunities and challenges from a qualitative research perspective. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 603–618). Springer.

Noe, R., & Ellingson, J. (2017). *Autonomous learning in the workplace*. Taylor & Francis.

Nokelainen, P., Aldahdouh, T. Z., & Aldahdouh, A. A. (2022). Bayesian statistics in the research field of professional learning and development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 213–241). Springer.

Ones, D. S., Anderson, N., Viswesvaran, C., & Sinangil, H. K. (Eds.). (2018a). *The Sage handbook of industrial, work & organizational psychology: Personal psychology and employee performance* (Vol. 1, 2nd ed.). SAGE.

Ones, D. S., Anderson, N., Viswesvaran, C., & Sinangil, H. K. (Eds.). (2018b). *The Sage handbook of industrial, work & organizational psychology: Organizational psychology* (Vol. 2, 2nd ed.). SAGE.

Ones, D. S., Anderson, N., Viswesvaran, C., & Sinangil, H. K. (Eds.). (2018c). *The Sage handbook of industrial, work & organizational psychology: Managerial psychology and organizational approaches* (Vol. 3, 2nd ed.). SAGE.

Palonen, T. (2022). Seeing workplaces from a social network analysis (SNA) approach. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 491–512). Springer.

Paloniemi, S., Penttonen, M., Eteläpelto, A., Hökkä, P., & Vähäsantanen, K. (2022). Integrating self-reports and electrodermal activity (EDA) measurement in studying emotions in professional learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 87–109). Springer.

Rausch, A. (2014). Using diaries in research on work and learning. In C. Harteis, A. Rausch, & J. Seifried (Eds.), *Discourses on professional learning* (Vol. 9, pp. 341–366). Springer. https://doi.org/10.1007/978-94-007-7012-6_17

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.

Renta Davids, A. I., Van den Bossche, P., Gijbels, D., & Fandos Garrido, M. (2017). The impact of individual, educational, and workplace factors on the transfer of school-based learning into the workplace. *Vocations and Learning, 10*(3), 275–306.

Sandberg, F. (2010). Recognising health care assistants' prior learning through a caring ideology. *Vocations and Learning, 3*(2), 99–115.

Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.

Silvennoinen, M., Parviainen, T., Malinen, A., Karjalainen, S., Manu, M., & Vesisenaho, M. (2022). Combining physiological and experiential measures to study the adult learning experience. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 137–164). Springer.

Smet, K., Grosemans, I., De Cuyper, N., & Kyndt, E. (2022). Outcomes of informal work-related learning behaviours: A systematic literature review. *Scandinavian Journal of Work and Organizational Psychology, 7*(1–2), 1–18. https://doi.org/10.16993/sjwop.151

Spliethoff, L., & Abele, S. (2022). Measuring professional competence using computer-generated log data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 165–186). Springer.

Steffen, B., & Pouta, M. (2022). Participant's video annotations as a database to measure professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 293–312). Springer.

Taylor, A. (2017). Service-learning programs and the knowledge economy: Exploring the tensions. *Vocations and Learning, 10*(3), 253–273.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*, 130–154. https://doi.org/10.1016/j.edurev.2007.12.001

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*(1), 11–36. https://doi.org/10.1007/s12186-012-9091-z

Vähäsantanen, K., & Arvaja, M. (2022). The narrative approach to research professional identity: Relational, temporal, and dialogical perspectives. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 373–395). Springer.

Wilson, J. R., & Sharples, S. (2015). *Evaluation of human work* (4th ed.). CRC Press.

Michael Goller (PhD) is a research associate at the Institute of Educational Science at Paderborn University. Between October 2018 and March 2020, he served as visiting professor at the University of Bamberg and between April and September 2021 as a visiting professor at the University of Jena. His current research interests comprise expertise development and workplace learning in general, including in the particular context of digitalisation, as well as learning and development of students at the boundary of institutionalised and less institutionalised learning environments such as internships or particular learning arrangements in VET.

Eva Kyndt (PhD) is full Professor Human Resource Development at Swinburne University of Technology (Centre for Transformative Innovation) in Melbourne, Australia. Her research focuses on the intersection between learning and working and includes topics such as the transition from education to work, workplace learning, sustainable careers, and social networks in organisations. She is the founding Editor-in-Chief of Teaching and Teacher Education: Leadership and Professional Development.

Susanna Paloniemi (PhD) is a senior lecturer in the Department of Education, University of Jyväskylä, Finland. Her current research and teaching interests include professional agency, workplace learning, and emotions in work organisations. She has utilised various research methods in her research, and thus holds an interest in methodological elaborations within the field. Her recent publications include articles in Frontline Learning Research, Studies in Continuing Education, and Professional Development in Education. She acts as an associate editor of Vocations and Learning journal.

Crina Damșa (PhD) is an Associate Professor at the Department of Education and currently the Vice of Innovation and Digitalisation at the Faculty of Educational Sciences, University of Oslo, in Norway. Her main research interests focus on learning and teaching in higher education, the role of digital technologies in learning and platformization of education and learning. Currently, Crina leads a large research project on teamwork analytics, where analytics-based dashboards are explored as potential support for team-based learning of students and professionals. Crina has published widely in journals and volumes within the field of higher education and of relevance to workplace learning, such as Vocation and Learning and the The Wiley Handbook of Vocational Education and Training.

Part I
Methods for Data Collection

Chapter 2
Applying the Experience Sampling Method to Research on Workplace Learning

Jürgen Seifried and Andreas Rausch

Abstract Most learning in the workplace is informal and remains at least partly unconscious. Therefore, retrospective measurements of such learning are prone to memory bias. Applying the Experience Sampling Method (ESM) to research workplace learning can reduce this bias and provide additional opportunities to capture contextual factors of workplace learning. ESM has a long tradition of collecting data on everyday experiences. It was developed in the 1970s and has increasingly established itself as a tool for capturing everyday work experiences as well as learning processes in formal school contexts. However, literature research shows that ESM is rarely used in research on learning at work. This chapter aims to describe variations of ESM using exemplary studies. In addition, we discuss selected research questions and corresponding designs to explore workplace learning through the application of ESM.

Keywords Experience sampling method · Workplace learning · Process data · Research design

2.1 Introduction

Schwarz (2012, p. 22) stated that "every aspect of human cognition, emotion, motivation, and behaviour is situated and highly context-sensitive", and therefore research methods are needed to learn more about "the situated and embedded nature of human experience" (p. 23). From a research perspective, capturing the subjective

J. Seifried (✉)
Mannheim Business School (MBS), Economic and Business Education – Professional Teaching and Learning, University of Mannheim, Mannheim, Germany
e-mail: seifried@bwl.uni-mannheim.de

A. Rausch
Mannheim Business School (MBS), Economic and Business Education – Workplace Learning, University of Mannheim, Mannheim, Germany
e-mail: rausch@uni-mannheim.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_2

Table 2.1 An overview of intensive longitudinal methods (Connor & Lehman, 2012, p. 92)

	Experience	Behaviour	Physiology
	Phenomenology like mood, pain, fatigue, cognitions, perceptions, and appraisals	*Actions that are observable to others like drinking, smoking, exercise, talking, eating, interactions, and location*	*Internal workings of the body and brain like temperature, breathing, heart rate, blood pressure, and hormone levels*
Active	Experiences are self-reported by participants (e.g., mood ratings, pain ratings, stress ratings).	Behaviours are self-reported by participants (e.g., self-reported alcohol use, food/exercise diary, event-recording).	Physiology is measured by participants (e.g., participant takes saliva samples which are assayed for cortisol by experimenters).
	Experience sampling, daily diaries, event sampling	Experience sampling, daily diaries, event sampling	Neuroendocrine sampling, physiological sampling
Passive	Experience is inferred through observation (e.g., unobtrusive auditory sampling with the electronically activated recorder [EAR]).	Behaviours are measured with no intervention or reporting necessary (e.g., pedometer or actigraph to infer physical activity, unobtrusive auditory sampling with the EAR, GPS to measure location).	Physiology is measured with no intervention by participant (e.g., passive sampling of heart rate and blood pressure, continuous glucose monitoring, temperature tracking, measurement of breathing).
	Acoustic sampling	Activity sampling, acoustic sampling, passive telemetrics, context sampling	Neuroendocrine sampling, physiological sampling

experience in real-time and in different life situations – be it work or leisure – is a challenging endeavour (e.g., Beal, 2015; Connor & Lehman, 2012; Reis, 2012; Schwarz, 2012). Technological advances are helping to refine data collection methods and develop new approaches to research questions and research designs. Nowadays, eye-tracking, logging data, data mining, or observational videos are used to capture learning processes in real-time. From our perspective, however, the potential of collecting data in real-time lies in the holistic view of physiological and psychological data, taking into account the contextual factors. By linking different strands of data, such as experiential data, observational data, and behavioural data, deep insights into learning and development processes in different settings (e.g., informal learning at the workplace) can be gained. Table 2.1 provides an overview of different longitudinal data collection methods for analysing processes at the micro-level of daily life (Connor & Lehman, 2012, p. 92). Experience sampling requires participants to actively report experiences and behaviours. Passive approaches to capturing experiences or behaviours include observational methods or recording activities, such as using digital devices (e.g., GPS trackers). Additional methods such as measuring heart rate or recording temperature are suitable for recording physiological data.

The key features of the Experience Sampling Method (ESM), as one of these intensive longitudinal methods (Connor & Lehman, 2012), can be summarised as follows: ESM enables the capturing of daily experiences, behaviours, or physiological states of individuals in their natural environment. Repeated measures (daily, multiple times per day) of subjects' perceptions of various constructs are used to collect data (Beal, 2015). It enables the exploration of dynamically changing processes and experiences in the real environment or everyday life (Connor & Lehman, 2012).

A look at selected reviews on workplace learning (e.g., Billett et al., 2014; Gruber & Harteis, 2018; Malloch et al., 2011; Mikkonen et al., 2017; Rintala et al., 2019; Tynjälä, 2013) shows that experience sampling does not seem to play a prominent role in the context of researching professional and workplace learning. Questionnaires and interviews are widely used in workplace learning research (e.g., Berings et al., 2006; Guskey, 2014; Sawchuk, 2009). In some cases, these approaches are less effective because experiences are collected retrospectively and aggregated over a longer period. Workplace learning and professional development often comprise more informal processes, so it can be quite difficult to get a comprehensive picture using surveys. Therefore, online measures are becoming increasingly important. However, ESM studies are still rare in this context. This is surprising, as ESM has proven itself in many research areas (e.g., psychology, classroom research, clinical research, mental health research; see also Hektner et al., 2006) and is becoming increasingly popular. For example, ESM has been widely used in organisational research over the past decade (Beal, 2015; Gabriel et al., 2019) to learn more about well-being at work (Dimotakis et al., 2011; Sonnentag, 2015), stress in the workplace (Bono et al., 2013; Daniels et al., 2006), flow at work (Fullagar & Kelloway, 2009), the role of positive affect and personality (empathy and altruism) as predictors of workplace helping (Conway et al., 2009), the spill-over of workplace experiences on employees' home lives (Lim et al., 2018), organisational citizenship behaviour (Koopman et al., 2016), mistrust at work (Lanaj et al., 2018), and many other topics.

Unlike methods such as retrospective questionnaires or interviews – which are better suited for longitudinal studies where research questions concern developments over months, years, or decades – ESM has the advantage of minimising retrospective bias that can occur in a single assessment at a single point in time and between different people. Several variants can be attributed to ESM, which are referred to differently in some places (e.g., momentary assessment, ambulatory assessment, daily diary method or everyday experience method) (Beal, 2015; Stone, 2018). Ambulatory or ecological assessment is the overarching, more comprehensive term. In addition, it includes physiological and behavioural measurements (e.g., GPS data to locate individuals) (Wrzus, 2020). ESM and the daily diary method are similar in principle, but the daily diary method usually requires only one report per day over several weeks (once-per-day assessment approach), whereas in ESM participants report their current thoughts, feelings, emotions, and behaviours several times a day (Fisher & To, 2012, see also Beal & Gabriel, 2019). However, it is difficult to distinguish ESM from the diary method. Rausch et al. (2022) contrast typical features of the experience sampling method and the diary method by using criteria such as the aim of the study, the underlying research questions, measurement density, or the chosen sampling schedule. In contrast to ESM and the daily diary method, descriptive experience sampling (Hurlburt, 2006) asks study participants to describe their current thoughts at each prompt. Some signals are followed by an in-depth interview with the researcher (Fisher & To, 2012).

In summary: ESM is concerned with self-reported experiences, feelings, or behaviours: The use of ESM is appropriate when the constructs to be collected vary over time and the research question addresses intrapersonal issues. ESM designs involve intensive repeated interviews at short intervals (e.g., four reports per day) and short study duration of regularly 1–2 weeks (Beal, 2015). ESM can provide more valid information about people's daily lives than traditional interviewing techniques, where subjects can only characterise their daily lives retrospectively and in a blanket manner, which can lead to retrospective bias that occurs in point-in-time surveys (Gabriel et al., 2019). ESM is becoming increasingly popular in various disciplines, e.g., in learning and classroom research. A prominent field of application is the recording of emotional states in learning situations. For example, ESM can be used to record emotional processes (e.g., boredom), to analyse the emotional experience of different teaching methods (e.g., Goetz et al., 2014, 2020) or to research students' engagement (Xie et al., 2019a, b). In VET research, Detlef Sembill's research group has conducted several longitudinal studies with apprentices. In these studies, teaching experiences were continuously recorded in very short temporal cycles (between five and ten minutes; Continuous State Sampling Method, CSSM: Sembill et al., 2002; for further developments see Kaerner & Warwas, 2015; Kaerner et al., 2017). More recently, further variants have emerged in this context: Rausch et al. (2016, see also Rausch et al., 2019) developed the Embedded Experience Sampling Method (see below) for a holistic competence assessment.

2.2 The Experience Sampling Method (ESM) – A Brief Characterisation

2.2.1 Conceptual Elements of ESM

For a long time, researchers have been searching for an ecologically valid method to capture the subjective experience of individuals in everyday situations (Nowlis & Cohen, 1968). In this context, Csikszentmihalyi and colleagues (e.g., Csikszentmihalyi et al., 1977; Csikszentmihalyi & Larson, 1987; Larson & Csikszentmihalyi, 1983) established the Experience Sampling Method (ESM), which was in the methodological tradition of time-use studies and the survey of "subjective well-being".[1] This method was intended to capture activities, subjective

[1] ESM has a long tradition in research of daily life. As Beal (2015) points out, Hersey's 1932 study, which collected data (including blood pressure, sleep quality, work outcomes) from 12 men four times a day over an entire year, can be seen as the starting point of ESM research. However, no such studies can be found in the following decades, and it is not until the 1970s that ESM reappears.

Table 2.2 Conceptual elements of experience sampling methods (Beal, 2015, p. 385)

Key elements	Description
Natural environment	Capturing experiences as closely as possible to how they would naturally occur
Immediacy of experience	Prioritising concrete and immediate experiences over abstract or recalled experiences
Representative sampling	Assessing a range of experiences that accurately reflects an individual's daily life

qualities of experience, and their patterns, and to determine the extent to which external life circumstances could be related to subjective experience (Csikszentmihalyi & Larson, 1987, 527). Also in the 1970s, Brandstätter developed the time-sampling diary (TSD), which is very similar to ESM (Brandstätter, 2007). Hormuth (1986), following Csikszentmihalyi's group, developed a signalling device that was programmed via a portable Epson computer and could send 128 signals over 8 days. After being prompted by a signal tone, the test subjects each had to fill out questionnaires, which they carried with them in written form and sufficient quantity. These contained questions about the time of processing, the location and the activity just performed, and considered cognitive, emotional, and motivational items (Csikszentmihalyi & Larson, 1987, p. 527; Hormuth 1986, p. 269 f.). This example illustrates the basic principles of ESM: The "experience sampling methodology involves repeated measures of the same participants in everyday life, with a focus on assessing variables that fluctuate in the short term" (Fisher & To, 2012, p. 865). ESM studies thus collect repeated measures (daily, multiple times daily) of individuals' (e.g., staff, learners) perceptions of various constructs to capture lived, daily experience within-person. Thus, it can be said that ESM is typically used to capture constructs that fluctuate over time within an individual (within-person variability, e.g., well-being; Sonnentag, 2015). Beal (2015, p. 384) also agrees that "experience sampling designs involve intensive repeated assessments with short intervals and study durations", although "short" is difficult to determine. Table 2.2 summarises the core conceptual elements of ESM, following Beal (2015, p. 385).

2.2.2 Level 1 and Level 2 Variables (Within-Person and Between-Person Processes)

When considering ESM, a within-person and a between-person perspective can be distinguished (e.g., Beal, & Gabriel, 2019; Hamaker, 2012), and it is useful to draw on the terminology of multilevel modelling. There, variables that are measured repeatedly over time for a person, as well as the relationships of the variables within the person, are referred to as level 1 (analysis of processes within the person). In contrast, level 2 refers to stable person-level variables that are typically measured once per participant (analysis of processes between persons). These include, for example, demographic data, attitudes, traits (e.g., Big Five) and stable work environment characteristics, as well as the relationships between these variables. In this context, level 2 relationships cannot automatically be transferred to similar level 1 variables or vice versa. The underlying processes and the strength and direction of the relationships may be different at different levels and need to be examined at each level (Fisher & To, 2012, see also Beal, 2015).

2.2.3 *Study Designs and Schedules for Data Collection*

Connor and Lehman (2012, p. 89 ff.) elaborate on some key decisions to be made when designing a study to explore daily life. First, the research question must be determined. In ESM, the focus is on micro-level processes in daily life, which are collected in real-time. ESM is best suited for asking correlative questions. Causal questions are less suitable for investigation. For the target variables, variables that address experiencing, thinking or behaviour can be considered. For physiological variables, other methods are suitable (see above, Table 2.1). The target group and the sample size must then be determined. It should also be considered that different target groups may be familiar with different data collection technologies (e.g., digital natives vs seniors). These decisions influence the choice of the data collection strategy (schedules for data collection, see below), the technology used for data collection (e.g., delivery of the questionnaire via personal digital assistants, cell phones, tablets, paper-based), pilot testing of the questionnaires and assessment tools, as well as the concrete implementation of the study (see Fig. 2.1). For studies in work contexts, it is also important to clarify data protection issues with those affected in advance and to involve employee representatives from the beginning. In addition, it is important to consider how to motivate participants to take part in the study on an ongoing basis.

Concerning the timing of data collection (designing sampling protocol), three approaches are commonly discussed (e.g., Conner & Lehman, 2012, Fisher & To, 2012; Myin-Germeys et al., 2018), namely interval-dependent reporting, signal-dependent reporting, and event-dependent reporting. (1) In interval-dependent reporting, participants respond at predetermined times that generally remain constant over the survey period (i.e., daily at 10:00 am, 2:00 pm, and 4:00 pm). (2) In contrast,

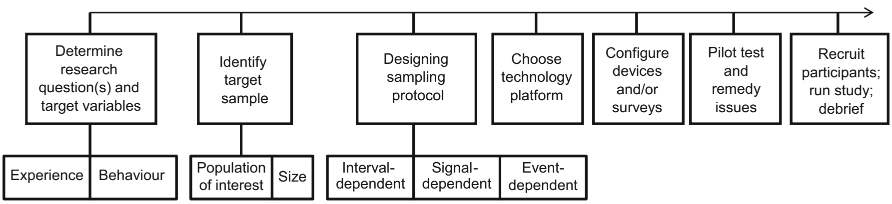

Fig. 2.1 Key questions concerning the design of an ESM study. (Adapted from Connor & Lehman, 2012, 90)

Table 2.3 Schedules for ESM reporting (following Fisher & To, 2012, p. 869)

Schedule	Possibilities	Constraints
Interval contingent	No need for specific technological support Works paper-pencil based, PDA, cell phone, or via personal computer Less intrusive for participants as they can adjust well to data collection intervals Useful for recording all phenomena that typically occur cyclically (e.g., regular increase in stress during the day)	The risk that fluctuations are not recorded with sufficient accuracy if the time intervals are chosen "incorrectly" If the period between the signal and the data input is too long, distortions or forgetting effects may occur With many signals, it becomes difficult to motivate participants to participate steadily
Signal contingent	Possibility of covering the entire range of experience Possibility of capturing a random sample of highly variable states Capturing of the experience with only a small memory error A larger number of signals/measurements per day are possible	Technological support is necessary The risk that signals are not perceived in noisy environments (e.g., workplace) The risk that the lack of predictability of signals will be stressful for participants
Event contingent	Allows concentration on specific events of particular interest Information about the events of interest can be captured on time and without memory loss	The risk that events of interest are not recognised as such by the participants Need for the events of interest to have a trigger that is identifiable to the participants Participants know exactly which events are of interest If the events of interest are rare (or do not happen), only a little data is collected

the signal-dependent variant asks randomly for responses at different times during the day. The schedules can be stratified, e.g., with two randomly selected signals within a morning and the condition that there must be a certain period between the signals. This method is always useful when a representative sample of events and experiences is needed. (3) Finally, there is the possibility of requesting a report after the occurrence of predefined events (event-dependent). This means that participants report their experiences each time a certain event occurs (e.g., feedback from supervisor, noticing an error in the work process, a stressful situation at work). Of course, some of the approaches can be combined. Table 2.3 shows the advantages and disadvantages of the different approaches (Fisher & To, 2012, p. 869, see also Conner & Lehman, 2012, p. 106 f., and Myin-Germeys et al., 2018, p. 127 f.).

2.2.4 Data Analysis

There are several things to consider when analysing ESM data. First, time-series analyses provide the opportunity of mapping processes that occur within individuals

(Box & Jenkins, 1976; Box et al., 2008; Chatfield & Xing, 2019; Hamaker, 2012). A prerequisite for time series analysis is the presence of a sufficient number of data points[2] – a requirement that is usually met in ESM studies. Time series analysis can be conducted under different objectives: Usually, the first step is to identify the characteristics of a time series by describing the empirical time series. The first steps are usually the plot of diagrams and the calculation of usual measures such as arithmetic mean and standard deviation. Splines[3] can be determined to show the structure of the development over time. Regression models allow the identification of trends. Furthermore, the formulation and adjustment of stochastic models serve to conclude an empirical time series to an underlying theoretical model. The empirical time series is understood here as the realisation of a stochastic process, i.e., a set of dependent random variables. The modelling is done with the help of ARIMA processes (Autoregressive Integrated Moving Average). The structure of the data or the location of the data at different levels must then be considered in multi-level analyses. In ESM, we usually have repeated assessments nested in days, which in turn are nested in subjects, and sometimes subjects are nested with a workgroup or an organisation (Myin-Germeys et al., 2018). Therefore, 2-level models are often used, with level 1 being within people and level 2 being between people. When signal-level reports are nested within days and days are nested within individuals, or when people belong to different workgroups, 3-level models are required (Fisher & To, 2012). In addition, mixed-effects models (inclusion of additional random effects), autoregressive models or mixed latent Markov models are considered appropriate to analyse ESM data (Myin-Germeys et al., 2018).

2.2.5 Potentials and Limitations of ESM

ESM offers many advantages, but it also has some limitations (Atz, 2013; Beal, 2015; Fisher & To, 2012; Gabriel et al., 2019; Napa Scollon et al., 2009; Uy et al., 2010; Wrzus, 2020): An important advantage of ESM is that real-world measurements are possible (ecological validity, e.g., Reis, 2012). Therefore, many events that cannot be replicated in the laboratory or cannot be replicated with sufficient validity (e.g., social interactions in the workplace) can be captured and analysed. The

[2]There are no clear guidelines for a minimum number of measurement points. As a rule, it is assumed that the reliability of the prediction increases with the number of measurement points. Depending on the research question and complexity of the underlying model, at least 50 (better 100) measurements are recommended. The determination of the distance between two data entries also proves to be relevant. If the distance between the measurement points is too large, there is a risk that the underlying processes cannot be adequately recorded. If the time interval between data entry is very short and the subjects are asked to enter data very frequently, they might feel disturbed and refuse to enter data.

[3]A spline is a function that connects a certain number of points "smoothly". Splines are mainly used for interpolation and approximation.

approach is thus "ecological" in that it examines study participants in their usual environment and activities. ESM is appropriate when large amounts of data need to be collected for individuals. In typical ESM studies, for example, subjects are asked to respond to ESM signals four or five times a day over one or two working weeks, which means that between 20 and 50 data points per participant are possible. This abundance of data enables the analysis of dynamic processes within a person over time. ESM is therefore suitable for analysing meaningful within-person variations over short periods in constructs such as job satisfaction, work stress, emotions, motivation, flow, engagement, agency. It is also possible to capture work behaviour, effort, performance, workload, task characteristics and so on (Beal, 2015; Fisher & To, 2012). Another important aspect is that ESM minimises errors that exist due to recall errors. This is a serious disadvantage in global retrospective experience reports. ESM enables "online" reporting of current thoughts and experiences. ESM is further suitable for research questions that require within-person and between-person analysis. In addition, an analysis of the moderation of Level 1 relationships to Level 2 becomes possible, i.e., how individuals differ in their perception of environmental conditions. Finally, with today's technical equipment, data collection (e.g., with mobile phones) is quite easy, but of course, ESM can also be conducted on a paper-pencil basis, so ultimately no expensive technical equipment is necessary.

A key limitation is that ESM can be perceived as burdensome by participants (intrusive ESM). Therefore, one tries to keep the questionnaires as short as possible. In ESM, it is common to use one-point assessments. Repeating a detailed assessment several times can lead to the selectivity of the samples (e.g., only highly motivated people participate) and other biases (e.g., assessments may be skipped at inappropriate times). In addition, measurement reactivity can occur, e.g., because people observe their experience and behaviour more closely, reporting moods also influences them in the long term, or study participants report experiences in a socially desirable way.

Finally, it is important to note that the ESM is not necessarily superior to other types of self-report (e.g., retrospective, focusing on trait constructs). Connor and Lehman (2012, p. 91) state that "experiential and retrospective reports provide different types of information – one is not necessarily better than the other".

2.3 Examples for the Use of ESM to Research Workplace Learning

Nowadays, ESM is used in many domains. However, studies with a strong focus on workplace learning processes are still rare. In Example 1, we describe one of the few studies that use ESM to analyse workplace learning. However, if we broaden the perspective, we can find many ESM studies related to the workplace. For example, ESM is often used in organisational research (e.g., knowledge sharing behaviour, see Example 2). In addition to the exemplary studies from the field of organisational research, we show how ESM can be used in VET research. We report on a study from the school sector (analysis of cognitive and emotional processes in VET-classrooms, see Example 3). Finally, we will outline a new approach to using ESM for the technology-based assessment of competencies in vocational education (see Example 4). In the following, we present each of the selected studies in a brief profile and characterise them in terms of the study objective, data collection method, sample and response rate, evaluation method and selected key findings.

2.3.1 Example 1: Workplace Learning (Daniels et al., 2009, 2011)

Daniels et al. (2009, 2011) conducted different ESM studies to learn more about the extent to which work control and social support enable workers to solve problems at work and thus promote learning processes (study 1). The aim of study 2 was to analyse the relationships between workplace characteristics and idea generation and implementation. Table 2.4 provides an overview of the profile of the ESM studies, and Fig. 2.2 shows the operational definitions and relationships hypothesised by Daniels et al. (2009, 2011). Both studies are based on Karasek and Theorell's (1990) model of demand control and support, which addresses three key components, namely work demands (e.g., time pressure, difficult work), work control (e.g., degree of autonomy), and social support (e.g., interactions with colleagues and supervisors, see Johnson & Hall, 1988).

Study 1 focused on analysing the relationship between job control and social support and workers' ability to solve problems at work, assuming that the relationship is mediated by learning processes (Daniels et al., 2009). It is further hypothesised that learning occurs at work when both demands and control and support are high (active learning hypothesis), while high demands and low control and support should have negative effects (stress hypothesis). It is also assumed that learning promotes well-being.

The samples comprised two subsamples of 78 and 106 workers from different occupations and companies, respectively, who provided information about their experiences four times a day (interval-contingent) during one working week. Personal digital assistants were used for data collection. The extent to which work control was used to solve problems was operationalised by measuring the extent to which participants changed aspects of their work activities to solve problems (two items: "In the past hour, did you change your work objectives for the hour to solve the issues?" "In the past hour, did you change the order in which you normally do work tasks to solve the issues?"). The extent to which social support was used to solve problems was assessed by measuring the extent to which participants discussed problems to solve them (two items: "In the past hour, did you discuss the issues to help you solve them?" "In the past hour, did you ask for other people's views to help solve the issues?"). Learning was measured with three items: "Have

Table 2.4 Profile of two ESM studies on workplace learning (Daniels et al., 2009, 2011)

Study objective	Study 1: Analysing the relationship of job control and social support and the workers' ability to solve problems at work. The relationship is mediated by learning processes Study 2: Analysing the relationships between workplace characteristics and idea generation and implementation
Method of data collection	Personal digital assistants (PDAs), time sampling, four times a day, one working week
Samples and response rates	Study 1: Two samples; 78 and 106 workers from different professions and companies in the UK, compliance rates of 66 and 71%. Study 2: 89 employees from different organisations in the UK, compliance rates of 66%.
Data analysis	Multilevel regression analysis (HLM 6, Raudenbush et al., 2004)
Key findings	Study 1: Learning mediated the relationship between changing aspects of work activities to solve problems (job control) and well-being as well as the relationship between discussing problems to solve problems (social support) and affective well-being. Study 2: The extent to which individuals "changed aspects of their work activities to solve problems" was associated with higher levels of idea generation among individuals with high levels of initiative. The extent to which people "discussed problems to solve problems" was associated with higher levels of idea implementation for people with high levels of personal initiative.

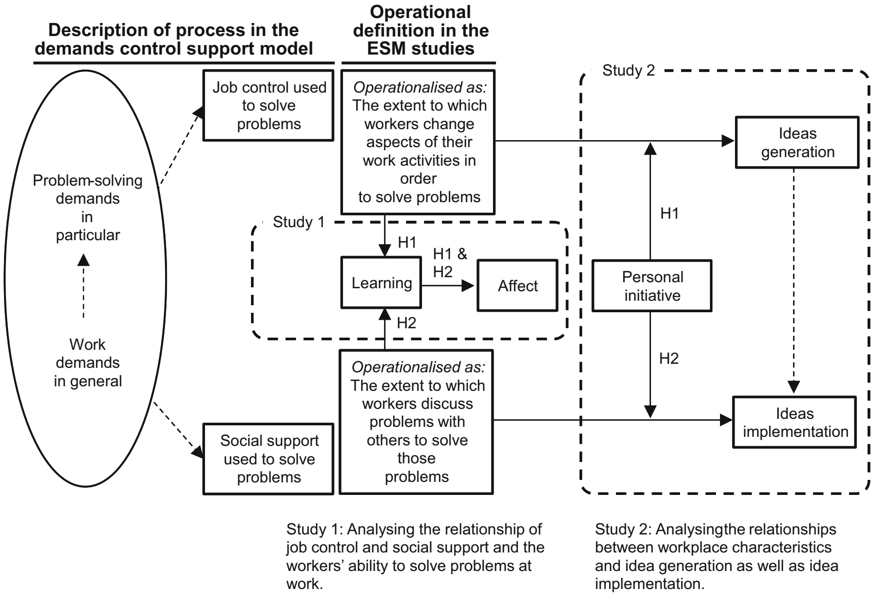

Fig. 2.2 Hypothesised relationships between job control, social support, and the target variables. (Following Daniels et al., 2009, p. 1005, and Daniels et al., 2011, p. 585)

you learnt anything in the past hour that would help your work performance?" "Have you learnt anything in the past hour that would help you deal with difficult issues more efficiently?" "Have you learnt anything in the past hour that would help you solve work problems more quickly?" Learning acts as a mediator on both the relationship between changing aspects of work activity to solve problems and well-being (Hypothesis 1) and the relationship between discussing problems to solve problems and well-being (Hypothesis 2). Well-being (affect) was assessed by asking workers how they felt in the moment (e.g., motivated, worried, excited).

Daniels et al. (2009) used multilevel regression (HLM 6, Raudenbush et al., 2004) to analyse the data. In both samples, changing aspects of work activities to solve problems was associated with learning. The same was true for discussing problems with others, and Daniels et al. found associations between hourly learning and momentarily activated pleasant affect (well-being). All in all, the findings support the assumption that the way individuals use control and support to deal with problem-solving challenges is associated with workplace learning and well-being.

Daniels et al. (2011) also conducted an ESM study to analyse the relationships between workplace characteristics and idea generation and implementation. It was expected that more agentic individuals (employees with high personal initiative) would be more likely to use workplace controls to solve problems in ways that generate new and useful ideas (H1) as well as implementing ideas (H2). Again, the ESM was used and again staff provided data up to four times a day for up to five working days ($N = 89$). Idea generation was assessed using three-item scales (example item: "In the past hour, have you had any ideas that could improve your work performance?") and idea implementation (example item: "In the past hour, have you implemented new ideas that could help you deal with difficult issues more efficiently?") in situ. The authors predicted cross-level interaction between initiative and problem solving, and multilevel regressions were run (levels: hourly responses, embedded in participants, embedded in different organisations). They found that the extent to which individuals "changed aspects of their work activities to solve problems" was associated with higher levels of idea generation among individuals with high levels of initiative. The extent to which people "discussed problems to solve problems" was associated with higher levels of idea implementation for people with high levels of personal initiative.

2.3.2 *Example 2: Analysing Knowledge Sharing Behaviour in Business Centres (Weijs-Perrée et al., 2020)*

A central component of organisational knowledge management and an important resource for learning in organisations is knowledge sharing (Ipe, 2003; North & Kumta, 2018). This refers to the provision of an employee's knowledge within an organisation to create new knowledge. In this context, face-to-face interactions between employees play a significant role.

Weijs-Perrée et al. (2020) have conducted an analysis of knowledge sharing behaviour in the workplace using ESM. Their study focuses on the users of business centres, as the opportunity to network and share knowledge with people from other organisations is considered one of the main benefits of such facilities. Data were collected from 100 users (22 self-employed workers, freelancers, or entrepreneurs and 78 employees) of seven business centres in the Netherlands. Questionnaires were used to collect information about the users, the working environment, and the working conditions. Regarding ESM, the signal contingent sampling method was used. Respondents were asked three times a day over 10 working days about the characteristics of face-to-face interactions (e.g. pre-planned, intentional unscheduled or coincidental interaction; business or social interaction) at the workplace in the past hour. In addition, it was asked whether knowledge was shared during the interaction. If this was affirmed, it was also asked whether the knowledge would also be available elsewhere or not (explicit vs. tacit knowledge). Finally, information about the people involved in the interactions was collected. Table 2.5 shows the profile of the study.

Altogether, 658 face-to-face interactions during ten workdays were included in the analyses. Most of the interactions were work-related (61%), and most interactions were characterised as discussions/debates (about 40%) or chats (24%). Knowledge was shared (no knowledge sharing: 37%, explicit knowledge shared: 8%, tacit knowledge shared: 55%) in more than 60% of the interactions. A mixed multinomial logit model (MMNL) was used for data analysis and to learn more about the interaction effects between the knowledge sharing types and type of interaction, pre-planned interaction, inter-organisational interaction, office-concept, and so forth). The results show, among other things, that tacit knowledge is more often shared during discussions/debates, formal meetings and when receiving or giving advice. In contrast, explicit knowledge is more often shared during pre-planned interactions than during unplanned interactions. In addition, it could be shown that the office concept influences the sharing of knowledge: a negative effect of the cellular office concept (vs. open workspace) on the sharing of tacit knowledge was found.

Table 2.5 Profile of the study on knowledge sharing behaviour in business centres (Weijs-Perrée et al., 2020)

Study objective	Analysing the knowledge sharing behaviour in the workplace
Method of data collection	Personal digital assistants (PDAs), time sampling, three times a day, ten workdays
Sample and response rates	100 users of seven business centres in the Netherlands. The compliance rate for ESM is 39%.
Data analysis	Mixed multinomial logit model
Key findings	Interaction characteristics have significant effects on the type of knowledge (explicit vs. tacit knowledge) shared in business centres.

2.3.3 Example 3: Analysing Motivational Processes in Vocational School Classrooms (Seifried & Klüber, 2006)

From the perspective of the design of learning environments, it is of immediate interest what significance central design features (e.g., learning content, lesson organisation) have for the experience of teaching-learning situations. Regarding motivational experience, for example, it can be assumed that the experience of autonomy and interest varies depending on the selected didactic form of work and the learning content to be worked on (Deci & Ryan, 1985). Corresponding questions can be effectively investigated with the help of the ESM (e.g., Goetz et al., 2014, 2020). For work-related learning in vocational schools, several studies have been conducted that researched situational experience in self-organised learning (Sembill et al., 2002; Seifried & Klüber, 2006). In each of the studies, treatment groups (students who learn in a self-organised manner) were compared with control groups that received teacher-led instruction. Table 2.6 shows the key facts of the ESM study on motivation processes in vocational education.

A basic assumption was that learners in self-organised learning perceive a higher degree of autonomy and are therefore more motivated than learners in the control group (e.g., Deci & Ryan, 1985). Furthermore, it was hypothesised that in both groups (treatment and control group) the level of perceived autonomy would be higher in instructional phases that offered learners more freedom (there were phases of independent learning in both the experimental and control groups) than the comparative values of teacher-led instructional phases. To clarify this question, the ESM data were coupled with observation or video data (extent of self-directed learning in different didactic settings: teacher-centred class work discourse vs. student-centred group work). ESM data were collected with a personal

Table 2.6 Profile of the ESM study on motivational processes in VET classrooms (Seifried & Klüber, 2006)

Study objective	Analysing motivational processes of learners on different instructional settings (self-organised learning vs. teacher directed instruction)
Method of data collection	Personal digital assistants (PDAs), time sampling in five minutes intervals, 80 lessons (45 minutes each)
Sample and response rates	$n = 15$ students (treatment group), $n = 15$ students (control group) form a German vocational school, time series for both groups with more than 600 measurement time points. Participants with more than 50% missing values were excluded from the analysis. This applied to five participants in the treatment group and three in the control group. The compliance rate in total was in both groups between 50% and 60%.
Data analysis	Time series analysis, two-factor analysis of variance with a repeated-measures factor
Key findings	In the treatment group, students reported higher levels of autonomy than in the control group. In both groups, the perceived level of autonomy was determined through the instructional setting within the class.

digital assistant, in short time intervals (five-minute intervals: Continuous State Sampling Method, CSSM). The motivational items were each operationalised using a single item as follows: “I can participate” (autonomy); “I am interested” (interest) (continuous scale from 0 to 100). On this basis, time series were conducted with up to 600 measurement points per participant, which were first evaluated for each study participant regarding trends, mean values, standard deviations, and so forth. Subsequently, a mean value was calculated for each study participant for each experience item depending on the respective type of classwork (teacher-centred- vs. student-centred phases). To identify possible differences, a two-factor analysis of variance was then conducted for each category with a repeated-measures factor (instructional work form: teacher-centred class work discourse vs. student-centred group work) and an independent factor (treatment group vs. control group). The analysis of the motivation state variables led to the expected results. In the treatment group, students reported a significantly higher level of autonomy than in the control group. Furthermore, in both groups, the perceived level of autonomy during the work phases that could be described as student-centred was significantly higher than the values for frontal teaching. The analysis of the subjective experience during teaching showed that by distinguishing between different forms of teaching, a considerable degree of variability in the experience values over time becomes apparent. Figure 2.3 shows the results for the two motivational items (autonomy and interest).

2.3.4 Example 4: Measuring Non-cognitive Facets of Competence in Assessment Situations (Rausch et al., 2016, 2019; Seifried et al., 2020)

In competence diagnostics, the measurement of non-cognitive competence facets has been largely ignored for a long time. However, work-related competencies are regarded as a multi-dimensional construct that is not limited to cognitive facets such as domain knowledge or problem-solving strategies but also includes non-cognitive facets in the sense of domain-specific emotional and motivational dispositions such as interest and self-concept. If considered at all, these competence facets are measured with self-report questionnaires that are detached from the performance assessment. Instead, a new method called Embedded Experience Sampling (EES) (see Table 2.7 for key facts of this ESM study example) enables the measurement of non-cognitive processes during competence assessment.

In the context of computer-based competence tests, subjects are requested to pause at specific moments in time and spontaneously answer short questions (EES items) that relate to their experience of the current situation. These EES items are embedded in an event that resembles typical social interactions in the workplace. To evaluate the feasibility and validity of the method, EES was implemented in a series of studies in the context of commercial vocational education and training (Rausch et al., 2016, 2019; Seifried et al., 2020). Figure 2.4 shows an example of one of the EES items.

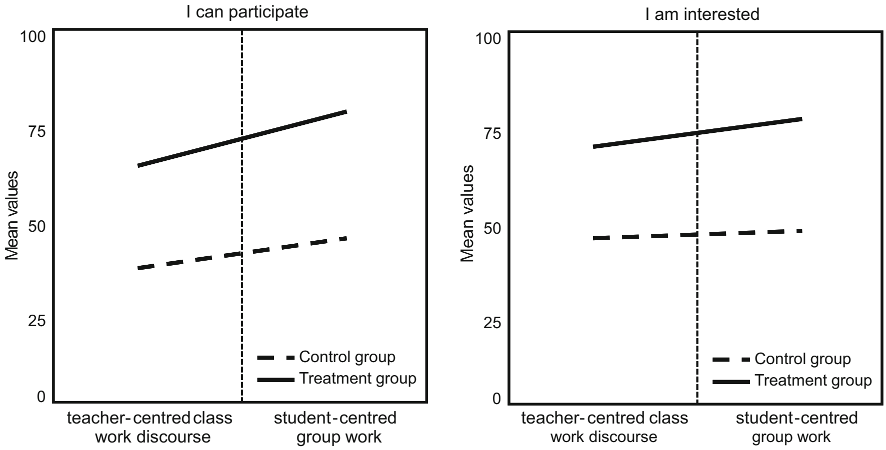

Fig. 2.3 Classroom experience as a function of the treatment and teaching activities

Table 2.7 Profile of the study on capturing state emotions in assessment situations (Rausch et al., 2016, 2019; Seifried et al., 2020)

Study objective	Integrative assessment of professional competencies of trainees in the German VET system. For assessment issues, authentic problem scenarios were developed. And the test takers worked in an open-ended problem space
Method of data collection	EES was integrated into an office simulation, four EES items within three task scenarios
Sample and response rates	$n = 782$ trainees in the dual German VET system. The compliance rate for the EES items in total is 100%.[a]
Data analysis	Item Response Theory: Estimation of plausible values, correlation analysis
Key findings	The correlations between the cognitive and the non-cognitive facets were all positive and showed moderate effect sizes. EES is seen as a valid and reliable approach for the measurement of non-cognitive competence facets.

[a]The test takers had to make an entry in EES items to continue working on the item. There was no option not to complete the EES items

The computer-based assessment of domain-specific problem-solving competence including the EES events was implemented in a larger study with almost 800 participants in vocational schools in six federal German states (see Rausch et al., 2016; Seifried et al., 2020). The data were collected in computer-equipped classrooms in vocational schools. Four EES events were defined (simulation of typical events in office work: short email response after receiving the task, phone call from the sender of the task, short visit by a colleague, short request from the sender of the task after receiving the solution). The test takers reported that they considered the EES to be authentic and, on average, did not evaluate EES as an additional burden when

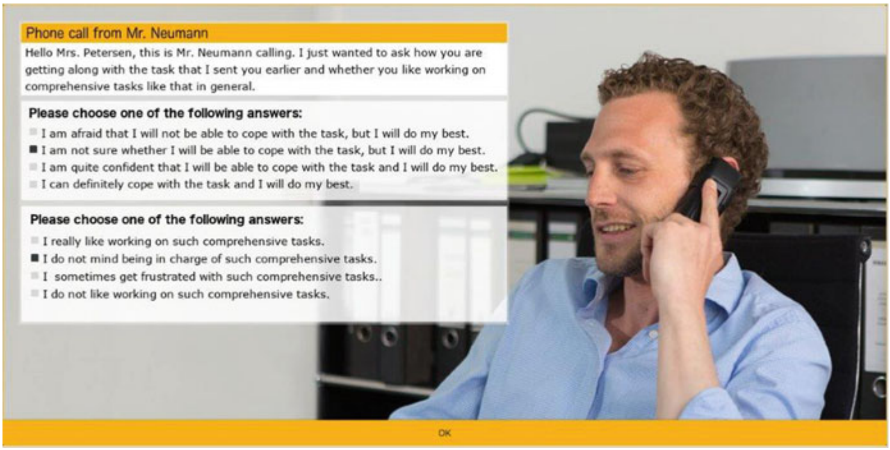

Fig. 2.4 Computer-based EES event “phone call” with two EES items (translated from German). (Rausch et al., 2019)

working on the tasks within the office simulation. Furthermore, substantial correlations between EES data and test motivation support the assumption of convergent validity. Furthermore, latent correlations between the cognitive facets and the non-cognitive facets based on plausible values were calculated. In general, the correlations are all positive and of small to medium size. All in all, one can say EES is a valid and reliable approach for the measurement of non-cognitive competence facets and allows for the integration of non-cognitive facets of competencies in the assessment of professional competencies.

2.4 Conclusion and Outlook

ESM is a promising method for learning more about processes in the workplace. There are ESM studies on topics such as knowledge sharing, help-seeking, stress at the workplace and so forth (see above). Unfortunately, this method is rarely used for research on learning processes in the workplace. In our contribution, we have shown the central features of ESM, which advantages and disadvantages are associated with the ESM method and exemplified possible applications of ESM. In our opinion, the potential for analysing workplace learning has not yet been fully exploited.

However, there are also challenges in ESM studies (e.g., Beal, 2015; Myin-Germeys et al., 2018). Challenges can be identified on the one hand in the questionnaire development or the definition and operationalisation of the target constructs. Here it is important to develop reliable and valid instruments and then use them in replication studies to strengthen the empirical basis and to be able to make evidence-based statements based on ESM studies. The usually rather small sample size is also seen as problematic. In this respect, a careful selection of study participants is necessary to minimise selection bias. In addition, measures must be taken to motivate the study participants to participate in the ESM studies permanently. Finally, from a methodological point of view, a pluralism of methods is advisable, whereby in particular the combination of active and passive approaches to data collection should be seen as promising.

There are several future directions of ESM research. Beal (2015) points to three developments that could be significant for the future use of ESM, namely the use of ESM for trait assessment (see ESM study example 4 above: assessment of vocational competencies), the expansion to higher levels of analysis and finally the inclusion of Big Data (when combining ESM data with other data sources such as heart rate monitors or GPS locators). The empirical coupling of experiential data (using ESM), physiological data (e.g., heart rate) and observational data (e.g., activities at work) enables the identification of correlation patterns between the visible level of action and the subjective, not directly observable experience.

The potential of ESM studies (like diary studies, cf. Rausch et al., 2022) lies primarily in researching informal learning in the workplace. This is especially true if the following applies: (1) The phenomena or processes to be studied are rather unnoticed in everyday work. (2) The phenomena of interest or their characteristics are difficult or impossible to reconstruct in retrospect. (3) The analyses aim at investigating correlations of changed situation characteristics on learning or development processes. When planning for the level 1 variables, it might be helpful to adapt Tynjälä's (2013, see also Gruber & Harteis, 2018) 3-P-model of workplace learning to the situational perspective of ESM. Variables should capture relevant contextual and individual preconditions of the actual situation (presage), process variables on behaviour and emotional experiences during the respective situation (process), and variables on the results of the situation (product) as for instance, the perceived learning. Table 2.8 provides examples of different research contexts. Regarding level 2 variables, the respective trait constructs such as general self-efficacy (example 1), general product knowledge (example 2) or the general openness to critique (example 3) would be interesting individual factors.

Overall, we can conclude that the Experience Sampling Method is a promising method for capturing people's subjective experience in everyday situations. The work on ESM can undoubtedly be seen as an important development for process studies, which also have direct relevance for learning and training contexts, and which open the possibility for new and deeper insights into interesting phenomena of workplace learning and professional development.

Table 2.8 Exemplary level 1 variables for ESM studies adapted from the 3-P-model of workplace learning

Research context	Situational preconditions (presage)	Processes	Situational products
Learning from errors	Individual preconditions: situational self-efficacy, mood Contextual preconditions: task characteristics, time pressure	Coping approaches such as problem-oriented coping, emotion-oriented coping, self-blaming, experienced support from colleagues	Results of error handling, perceived learning
Learning from sales talks	Individual precondition: product knowledge Contextual preconditions: customer characteristics	Applied conversational strategies, applied sales aids, talk time	Purchase decision, perceived learning
Learning from feedback	Individual precondition: satisfaction regarding the considered performance Contextual preconditions: characteristics of the feedback provider	Evaluative message of the feedback (positive vs. negative), justification of the feedback, social tone	Emotional states, perceived learning

References

Atz, U. (2013). Evaluating experience sampling of stress in a single-subject research design. *Personal and Ubiquitous Computing, 17*, 639–652. https://doi.org/10.1007/s00779-012-0512-7

Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior, 2,* 383–407. https://www.annualreviews.org/doi/10.1146/annurev-orgpsych-032414-111335

Beal, D. J., & Gabriel, A. S. (2019). Looking within: An examination, combination, and extension of within-person methods across multiple levels of analysis. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 305–327). American Psychological Association. https://doi.org/10.1037/0000115-014

Berings, M. G. M. C., Doornbos, A. J., & Simons, P. R.-J. (2006). Methodological practices in on-the-job learning research. *Human Resource Development International, 9*(3), 333–363. https://doi.org/10.1080/13678860600893557

Billett, S., Harteis, C., & Gruber, H. (Eds.). (2014). *International handbook of research in professional and practice-based learning*. Springer.

Bono, J. E., Glomb, T. M., Shen, W., Kim, E., & Koch, A. J. (2013). Building positive resources: Effects of positive events and positive reflection on work stress and health. *Academy of Management Journal, 56*(6), 1601–1627. https://doi.org/10.5465/amj.2011.0272

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series: Forecasting and control* (4th ed.). Wiley.

Brandstätter, H. (2007). The Time sampling diary (TSD) of emotional experience in everyday life situations. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 318–331). Oxford University Press.

Chatfield, C., & Xing, H. (2019). *The analysis of time series. An introduction with R* (7th ed.). CRC Press/Taylor & Francis.

Connor, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89–107). Guilford Press.

Conway, J. M., Rogelberg, S. G., & Pitts, V. E. (2009). Workplace helping: Interactive effects of personality and momentary positive affect. *Human Performance, 22*, 321–339. https://doi.org/10.1080/08959280903120279

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease, 175*, 526–536. https://doi.org/10.1097/00005053-198709000-00004

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence, 6*, 281–294. https://doi.org/10.1007/BF02138940

Daniels, K., Hartley, R., & Travers, C. (2006). Beliefs about stressors alter stressors' impact: Evidence from two experience-sampling studies. *Human Relations, 59*, 1261–1285. https://doi.org/10.1177/0018726706069768

Daniels, K., Boocock, G., Glover, J., Hartley, R., & Holland, J. (2009). An experience sampling study of learning, affect, and the demands control support model. *Journal of Applied Psychology, 94*(4), 1003–1017. https://doi.org/10.1037/a0015517

Daniels, K., Wimalasiri, V. P., Cheyne, A. J. T., & Story, V. (2011). Linking the demands-control-support model to innovation: The moderating role of personal initiative on the generation and implementation of ideas. *Journal of Occupational and Organizational Psychology, 84*, 581–598. https://doi.org/10.1348/096317910X494269

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.

Dimotakis, N., Scott, B. A., & Koopman, J. (2011). An experience sampling investigation of workplace interactions, affective states, and employee well-being. *Journal of Organizational Behavior, 32*(4), 572–588. https://doi.org/10.1002/job.722

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior, 33*(7), 865–877. https://doi.org/10.1002/job.1803

Fullagar, C. J., & Kelloway, E. (2009). Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology, 82*, 595–615. https://doi.org/10.1348/096317908X357903

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods, 22*(4), 969–1006. https://doi.org/10.1177/1094428118802626

Goetz, T., Frenzel, A. C., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation and Emotion, 38*(3), 401–419. https://doi.org/10.1007/s11031-013-9385-y

Goetz, T., Keller, M. M., Lüdtke, O., Nett, U. E., & Lipnevich, A. A. (2020). The dynamics of real-time classroom emotions: Appraisals mediate the relation between students' perceptions of teaching and their emotions. *Journal of Educational Psychology, 112*(6), 1243–1260. https://doi.org/10.1037/edu0000415

Gruber, H., & Harteis, C. (2018). *Individual and social influences on professional learning: Supporting the acquisition and maintenance of expertise* (Vol. 24). Springer. https://doi.org/10.1007/978-3-319-97041-7

Guskey, T. R. (2014). Evaluating professional learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 1215–1235). Springer.

Hamaker, E. L. (2012). Why researchers should think "within-person". A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). Guilford Press.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (Eds.). (2006). *Experience sampling method: Measuring the quality of everyday life*. Sage.

Hersey, R. B. (1932). Rates of production and emotional state. *The Personnel Journal, 10*, 355–364.

Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality, 54*, 262–293. https://doi.org/10.1111/j.1467-6494.1986.tb00395.x

Hurlburt, R. T. (2006). *Exploring inner experience: The descriptive experience sampling method*. John Benjamins.

Ipe, M. (2003). Knowledge sharing in organizations: A conceptual framework. *Human Resource Development Review, 2*(4), 337–359. https://doi.org/10.1177/1534484303257985

Johnson, J. V., & Hall, E. M. (1988). Job strain, work place social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health, 78*(10), 1336–1342. https://doi.org/10.2105/AJPH.78.10.1336

Kaerner, T., & Warwas, J. (2015). Functional relevance of students' prior knowledge and situational uncertainty during verbal interactions in vocational classrooms: Evidence from a mixed methods study. *Empirical Research in Vocational Education and Training, 7*(11), 1–26. https://doi.org/10.1186/s40461-015-0023-7

Kaerner, T., Sembill, D., Aßmann, C., Friederichs, E., & Carstensen, C. H. (2017). Analysis of person-situation interactions in educational settings via cross-classified multilevel longitudinal modelling: Illustrated with the example of students' stress experience. *Frontline Learning Research, 5*(1), 16–42. https://doi.org/10.14786/flr.v5i1.137

Karasek, R. A., & Theorell, T. (1990). *Healthy work*. Basic Books.

Koopman, J., Lanaj, K., & Scott, B. A. (2016). Integrating the bright and dark sides of OCB: A daily investigation of the benefits and costs of helping others. *Academy of Management Journal, 59*(2), 414–435. https://doi.org/10.5465/amj.2014.0262

Lanaj, K., Kim, H. P., Koopman, J., & Matta, F. K. (2018). Daily mistrust: A resource perspective and its implications for work and home. *Personnel Psychology, 71*(4), 545–570. https://doi.org/10.1111/peps.12268

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science, 15*, 41–56.

Lim, S., Ilies, R., Koopman, J., Christoforou, P., & Arvey, R. D. (2018). Emotional mechanisms linking incivility at work to aggression and withdrawal at home: An experience-sampling study. *Journal of Management, 44*(7), 2888–2908. https://doi.org/10.1177/0149206316654544

Malloch, M., Cairns, L., Evans, K., & O'Connor, B. N. (Eds.). (2011). *The Sage handbook of workplace learning*. Sage.

Mikkonen, S., Pylväs, L., Rintala, H., Nokelainen, P., & Postareff, L. (2017). Guiding workplace learning in vocational education and training: A literature review. *Empirical Research in Vocational. Education and Training, 9*(9). https://doi.org/10.1186/s40461-017-0053-4

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*(2), 123–132. https://doi.org/10.1002/wps.20513

Napa Scollon, C. N., Prieto, C. K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being. The collected works of Ed Diener* (pp. 157–180). Springer. https://doi.org/10.1007/978-90-481-2354-4_8

North, K., & Kumta, G. (2018). *Knowledge management. Value creation through organizational learning* (2nd ed.). Springer.

Nowlis, D. P., & Cohen, A. Y. (1968). Mood-reports and the college natural setting: A day in the lives of three roommates under academic pressure. *Psychological Reports, 23*(2), 551–566. https://doi.org/10.2466/pr0.1968.23.2.551

Raudenbush, S., Bryk, A., Cheong, Y. F., du Toit, M., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Scientific Software International.

Rausch, A., Seifried, J., Wuttke, E., Koegler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training*. Retrieved from https://link.springer.com/article/10.1186/s40461-016-0035-y

Rausch, A., Koegler, K., & Seifried, J. (2019). Validation of Embedded Experience Sampling (EES) for measuring non-cognitive facets of problem-solving competence in scenario-based assessments. *Frontiers in Psychology, 10*, 1–16. https://doi.org/10.3389/fpsyg.2019.01200

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.

Reis, H. T. (2012). Why researchers should think "real-world". A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). Guilford Press.

Rintala, H., Nokelainen, P., & Pylväs, L. (2019). Informal workplace learning: Turning the workplace into a learning site. In S. McGrath, M. Mulder, J. Papier, & R. Suart (Eds.), *Handbook of vocational education and training* (pp. 1–14). Springer. https://doi.org/10.1007/978-3-319-49789-1_97-1

Sawchuk, P. H. (2009). Informal learning and work: From genealogy and definitions to contemporary methods and findings. In R. Maclean & D. Wilson (Eds.), *International handbook of education for the changing world of work* (pp. 319–331). Springer. https://doi.org/10.1007/978-1-4020-5281-1_21

Schwarz, N. (2012). Why researchers should think "real-time". A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). Guilford Press.

Seifried, J., & Klüber, C. (2006). Unterrichtserleben in schüler- und lehrerzentrierten Unterrichtsphasen [Classroom experience in student-centered and teacher-centered instructional phases]. *Unterrichtswissenschaft, 34*(1), 2–21.

Seifried, J., Brandt, S., Koegler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence – A three-step scoring procedure. *Cogent Education, 7*, Article 1719571. https://doi.org/10.1080/2331186X.2020.1719571

Sembill, D., Wolf, K. D., Wuttke, E., & Schumacher, L. (2002). Self-organized learning in vocational education – Foundation, implementation, and evaluation. In K. Beck (Ed.), *Teaching-learning processes in vocational education* (pp. 267–295). Peter Lang.

Sonnentag, S. (2015). Dynamics of well-being. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 261–293. https://doi.org/10.1146/annurev-orgpsych-032414-111347

Stone, A. (2018). Ecological momentary assessment and experience sampling. In D. Vannette & J. Krosnick (Eds.), *The Palgrave handbook of survey research*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-54395-6_64

Tynjälä, P. (2013). Toward a 3-P model of workplace learning. A literature review. *Vocations and Learning, 6*(1), 11–36. https://doi.org/10.1007/s12186-012-9091-z

Uy, M. A., Foo, M. D., & Aguinis, H. (2010). Using experience sampling methodology to advance entrepreneurship theory and research. *Organizational Research Methods, 13*, 31–54. https://doi.org/10.1177/1094428109334977

Weijs-Perrée, M., Appel-Meulenbroek, R., & Arentze, T. (2020). Analysing knowledge sharing behaviour in business centres: A mixed multinomial logit model. *Knowledge Management Research and Practice, 18*(3), 323–335. https://doi.org/10.1080/14778238.2019.1664269

Wrzus, C. (2020). Experience sampling methods. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. https://doi.org/10.1007/978-3-319-24612-3_1300

Xie, K., Heddy, B. C., & Greene, B. A. (2019a). Affordances of using mobile technology to support experience-sampling method in examining college students' engagement. *Computers & Education, 128*, 183–198. https://doi.org/10.1016/j.compedu.2018.09.020

Xie, K., Heddy, B., & Vongkulluksn, V. (2019b). Examining learning engagement in context: The opportunities and challenges of mobile technology in experience-sampling method. *Contemporary Educational Psychology, 59*, 101788. https://doi.org/10.1016/j.cedpsych.2019.101788

Jürgen Seifried (PhD) is a full professor of Economic and Business Education – Professional Teaching and Learning at the Mannheim Business School (MBS) of the University of Mannheim, Germany. His research focuses on subject didactic issues, research on competence development of teachers and trainers, simulation-based learning, learning in the workplace, and methods of competence assessment in vocational education. Juergen's research is published in international journals, and he serves as an editorial board member and reviewer for several journals.

Andreas Rausch (PhD) is a full professor of Economic and Business Education – Workplace Learning at the Mannheim Business School (MBS) of the University of Mannheim, Germany. His current research interests include workplace learning, simulation-based learning, domain-specific problem solving, competence assessment, and vocational education and training (VET). He is committed to empirical educational research with a focus on process analysis and non-cognitive facets. Andreas's research is published in international journals. He is associate editor of Vocations and Learning and serves as editorial board member and reviewer for several journals in the field.

Chapter 3
Uncovering Informal Workplace Learning by Using Diaries

Andreas Rausch, Michael Goller, and Bianca Steffen

Abstract Many of the processes and outcomes of informal workplace learning remain almost unnoticed by the learner, which makes it difficult to empirically investigate informal workplace learning using retrospective self-reports. Intensive longitudinal methods allow for a data collection in situ, that means during or close to the actual processes. One such approach is the diary method, more rarely also referred to as working journals or learning logs. This chapter provides an introduction to the diary method as a data gathering tool for investigating informal workplace learning. It provides a discussion of different forms of validity, a systematic overview of typical research questions, diary parameters such as sampling methods, recording methods, and item formats as well as reporting standards in diary studies. In the second part of this chapter, two diary studies are presented to illustrate the various forms of implementation. The first study by Rausch investigates learning from errors in the workplace with a paper-based diary. In this section, the focus is on the measurement of emotions and the lack of correlation between diary and questionnaire data of similar phenomena. The second study by Goller and Steffen investigates the informal workplace learning of nurses during a special instructional setting (student-run hospital wards) and implemented voice recording. Future directions for diary studies on workplace learning are reviewed with respect to technological developments and mixed method designs.

Keywords Diary method · Workplace learning · Process data · Research design

A. Rausch (✉)
Mannheim Business School (MBS), Economic and Business Education – Workplace Learning, University of Mannheim, Mannheim, Germany
e-mail: rausch@uni-mannheim.de

M. Goller · B. Steffen
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: michael.goller@upb.de; bianca.steffen@upb.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_3

3.1 Introduction

Learning processes at work are difficult to empirically investigate due to their often informal nature. Much workplace learning occurs unintentionally or even unconsciously as a by-product of working. In fact, professional development that results from workplace learning might often not even be perceived as learning by employees but is more often described as something that comes naturally with experience. At the same time, however, most people would agree that a large portion of work-related competence is acquired directly at work and that some areas of competence might even only be acquired through work experience. In this respect, the so-called 70% rule persists in the consulting industry and practice literature, although there are no convincing theories or empirical findings that support this myth (Clardy, 2018). Instead, legitimate research attempts to disentangle the many individual and contextual factors that influence informal learning in the workplace. However, this research often relies on retrospective data derived from questionnaires and interviews (Berings et al., 2006; Littlejohn & Margaryan, 2014; Rintala et al., 2019; Sawchuk, 2009). Interviews are useful, for instance, to investigate particular incidents which are assumed to be well remembered such as crucial errors, drastic changes and so forth. Similarly, questionnaires are particularly useful for collecting data on constructs that are assumed to be relatively stable such as personality traits, attitudes or overall job satisfaction. However, regarding varying job characteristics, even such a simple question like 'To what extent can you try new things in your daily work?' can hardly be answered validly in retrospect. In this context, it is implicitly presumed that each respondent remembers a representative amount of work situations, evaluates the opportunity to have tried something new (perhaps even without actually having done it), assigns a value to each of these situations, and calculates a time-weighted mean value (Wheeler & Reis, 1991; Schwarz, 1990). All in all, a scenario that might be highly unlikely. Moreover, if the questions refer to aspects that are considered irrelevant during the actual event, it is even more difficult to remember (Jobe, 2000). This might be the case in particular for informal learning in the workplace because people pursue working goals rather than learning goals. Still, in questionnaires respondents usually do not mind providing such averaged retrospective self-reports on varying phenomena over longer periods of time. From a research point of view, however, the validity of such measures is highly questionable. In fact, many studies that compared retrospective and real-time self-reports in different domains showed that the results can differ greatly between these data sources (Rausch, 2012; Schwarz, 2012; Tourangeau, 2000). Finally, averaged retrospective self-reports on different constructs are often used to make causal inferences. However, a correlation between self-reported scope of action and self-reported learning on the level of the individual does not imply a correlation between scope of action and learning on the level of the situation (see Rausch, 2013). It might be that one perceives autonomy on Mondays and learning on Fridays, to put it bluntly. These disadvantages of retrospective self-reports on informal workplace learning can be avoided or reduced by applying a 'daily life approach' (Reis, 2012).

Using diaries is one method of increasing popularity that attempts to reduce the aforementioned shortcomings of retrospective self-reports (Sharples & Cobb, 2015). Variations of diary methods range from unstructured weekly reports to small-step questionnaires that are completed several times a day. Stone et al. (2007) choose Ecological Momentary Assessment (EMA) as an umbrella term for different variants of real-time self-report methods. They define EMA as '... the repeated collection of real-time data on participants' momentary states in the natural environment' (Stone et al., 2007, p. 3). They list diaries, behavioral observations, self-monitoring, time budget studies, the experience sampling method, and ambulatory assessment as historical roots that arose within different disciplines. Learning or working journals, journaling in general, and learning logs are additional terms for similar approaches. Bolger and Laurenceau (2013) prefer 'intensive longitudinal methods' as a collective term and emphasize the sufficient amount of repeated measurements as a common feature of these methods. The resulting data allows the researchers to '... peek into the complexities of experiences in order to achieve a more comprehensive description of daily life' (Stone & Shiffman, 2007, p. 106).

This handbook contains three chapters on intensive longitudinal methods: the chapter by Cuyvers et al. (2022) on longitudinal multiple case study design, the chapter by Seifried and Rausch (2022) on the experience sampling method (ESM), as well as the chapter at hand on the diary method. The method suggested by Cuyvers et al. strongly differs from diaries or ESM in so far as data collection is conducted by the researchers themselves and not by the study participants. However, the latter two approaches cannot be clearly distinguished. Fisher and To (2012), for instance, distinguish ESM from diary methods in that the latter usually have only one report per day. Although this is not agreed on coherently in the literature on diary studies, ESM is indeed typically associated with a higher frequency of data collection. Similarly, Rausch (2014) suggests that diaries usually request more complex responses including free text, whereas ESM rather contains a few closed items that can be answered more quickly. Csikszentmihalyi and colleagues see the biggest difference in the fact that ESM studies are mostly signal-contigent, so that participants cannot anticipate the time of the prompt (Kubey et al., 1996). Table 3.1 suggests some typical characteristics of each method which are in line with their

Table 3.1 Typical characteristics of the diary method and the experience sampling method

	Diary method	Experience sampling method
Main focus more often on . . .	Situational context variables	Emotional states
Measurement density	Once a week to several times a day	Once a day to several times an hour
Schedules	Typically event-contingent	Also signal-contingent or interval-contingent
Main question format	Open ended questions and free text fields	Typically limited to closed questions and checkboxes
Trigger of data entry	Individual decision based on specifications given by researchers	Often technologically prompted through mobile devices

historical roots (Shiffman et al., 2008) but not to be seen as ultimate criteria for a strict separation. It offers a non-binding guide for researchers to classify their methodological approach.

In research on learning and instruction as well as respective practice, diaries and similar approaches are also used as pedagogical interventions, for instance, to trigger reflection or to document learning processes. However, these kinds of diaries are not discussed here. In this chapter, we give an overview of the key features and decisions when using the diary method as a research tool to investigate informal workplace learning. We further illustrate the diary method with two empirical studies.

3.2 Diaries as a Research Instrument

3.2.1 Validity of Diary Data on Informal Workplace Learning

'Ecological validity refers to whether a study accurately represents the typical conditions under which that effect occurs in the real world' (Reis, 2012, p. 6). Hence, intensive longitudinal methods like the diary method can '... dramatically enhance the ecological validity of research findings' (Hufford, 2007, p. 54) and help reduce the risk of overreliance on retrospective self-reports in the research on informal workplace learning (Littlejohn & Margaryan, 2014). As mentioned earlier, recall is not only often inaccurate but, in many cases, also systematically biased, that is, '... the errors made in recalling information are not just random noise; rather, they change the data in systematic ways' (Shiffman et al., 2008, p. 4). Hence, we cannot hope that somehow on average retrospective self-reports provide valid data, if only the sample size is big enough. They may provide valid data for many other purposes. However, evaluations of averages across many situations that are dynamic in nature and about characteristics that often go unnoticed even in the actual situations are very prone to bias.

Every data collection method is associated with specific measurement errors. Measurement errors can be classified under random error, which in turn is usually discussed under the label of reliability, and systematic error, which refers to validity (Bolger & Laurenceau, 2013). Reliability is usually seen as a prerequisite to ensure validity and it is approached with statistical analysis such as inter-rater comparisons, test-retest comparisons, split half and so forth (Wilson & Sharples, 2015). Validity is '... defined as whether something measures what it claims to measure' (Wilson & Sharples, 2015, p. 26). There are different types of validity that will be described in regard to issues of retrospective data collection methods in comparison to diary studies in the context of workplace learning. We hereby strongly draw on the methodological discussion provided by Wilson and Sharples (2015):

- *Construct validity* refers to the extent to which a measure only represents construct-relevant aspects and does not measure construct-irrelevant variance. For instance, retrospective self-reports can be biased due to certain situational factors at the time of data collection (e.g., some kind of mood or emotion; Brief et al., 1995). Diaries mitigate this issue as they generate multiple in-situ data points.

- *Content validity* refers to whether a measure represents all relevant parts of the construct to be measured. Diary methods can enhance content validity because the repeated measurement allows the inclusion of rarer events that often go unmentioned in retrospective surveys. Furthermore, less conscious processes of incidental learning are more likely to be captured in situ than in retrospect.
- *Face validity* relates to the degree of consensus that a measure actually captures the relevant construct. It is strongly influenced by the appearance of the measure and, in turn, has a strong impact on the acceptance of the measure by stakeholders. Since informal workplace learning occurs in everyday work processes, collecting data close to these processes has a higher face validity.
- *Internal validity* refers to the extent to which a measure shows plausible relations between hypothesised causes and effects. In a series of studies using diaries and retrospective questionnaires, Rausch (2013) investigated the relation between task characteristics and trainees' perceived learning. While based on questionnaire data, autonomy is usually highly correlated with perceived learning, there were no or even negative correlations found in the diary data. At second glance, this is not surprising because much of the learning of trainees includes social interactions with more experienced employees. However, in these situations the trainees' scope of action is often limited, which is plausible but not recognisable in retrospective questionnaire data since the latter are highly aggregated.
- *External validity* refers to the generalisability to other contexts. Usually adapting the diary method to particular contexts, may limit the generalisability of the findings. However, this makes the often-unquestioned generalisability of standard questionnaires even more debatable. The findings of the three diary studies reported by Rausch (2013) show that some correlations between task characteristics and perceived learning are relatively stable across different settings in vocational education and training while others vary.
- *Predictive validity*, as a part of criterion validity, describes 'how well the measure predicts what we eventually see happen in the real world' (Wilson & Sharples, 2015, p. 26). Since diary methods move data collection '... out of the laboratory and into the daily lives of their subjects' (Hufford, 2007, p. 54), an extrapolation of situational conditions, their subjective experience and resulting learning opportunities can be expected to be valid, given the underlying diary data is representative. However, research that uses diary data on informal learning in the workplace to predict competence development using valid performance assessment can be considered the holy grail of validation studies. No such study is known to the authors.

Summing up, there are several advantages of diary studies that enhance validity. However, besides all arguments in favour of diary studies, there are also some downsides endangering both the reliability as well as validity of a study: Asking participants to fill out a diary in their everyday working life (whether it be weekly, daily, or even on a closer schedule) might be experienced as an additional burden. This can easily lead to missing data or drop-outs. Moreover, when working under time pressure or when the work task prohibits interruption, participants cannot fill in diaries at any requested time. Particularly, in the most heated – and thus, probably most interesting – moments, the diary will be postponed in favour of current tasks. Furthermore, diaries add additional cognitive load. By design, diaries interrupt participants in their everyday work and force them to focus on perhaps unfamiliar aspects of their jobs. This cognitive shift to the diary and back to the tasks at hand might hinder participants during their workflow. All these issues can lead to less engaged participants giving sparse or no answers. Also, social desirability could be a problem, as with all self-reported data. Finally, filling in diaries might induce treatment effects. Participants might pay closer attention to phenomena or events that would otherwise have remained unnoticed (Wheeler & Reis, 1991). This could then lead to behaviour like reflection or exchange with colleagues. For instance, a study on informal workplace learning might motivate workers to engage in more information seeking as the diary explicitly asks about such behaviour. This could seriously threaten the validity of collected diary data. In sum, it is important that researchers know about potential downsides of diary studies. However, some of the detrimental effects of diaries can be mitigated if they are considered in the design of the research study (see Sect. 3.2.3).

3.2.2 Types of Research Questions

Schwartz and Stone (2007) provide a sophisticated list of seven types of research questions that can be addressed with diary data based on multi-level analysis. However, we would like to add – as a step '0' – the very basic questions of frequencies, mean values, ranges, standard deviations and so forth. Particularly in research on informal workplace learning, descriptive statistics on phenomena that have previously remained unexplored are very insightful. Table 3.2 presents the extended ideal steps of analysis provided by Schwartz and Stone (2007), enriched with exemplary research questions in the context of informal workplace learning and methodological notes.

3.2.3 Parameters of the Research Diaries

Using diaries in research on workplace learning makes it possible to address manifold research questions as illustrated in Table 3.2. In contrast to retrospective self-reports, diary research usually results in higher validity but also comes with downsides as laid out above. Increasing validity while reducing the downsides of research diaries requires a careful and elaborated design process. Table 3.3 provides the main parameters and typical options that make up a diary study (see also Rausch,

Table 3.2 Steps of analysis and exemplary research questions

Steps of analysis	Exemplary research questions
0. Location and dispersion of individual and situational variables across all subjects	How many cases of knowledge sharing does a person record on average per week? To what extent are errors perceived as an opportunity to learn at all? What are the most common working situations that are perceived to be conducive to learning?
1. Differences in situational variables between subjects	Do participants differ in their average level of perceived learning from errors? Do participants differ in their average level of stress when receiving feedback?
2. Person-level factors of differences in experience between subjects	Does feeling ashamed in an error situation vary by neuroticism? Does self-assessed competence have an influence on feeling nervous when receiving feedback?
3. Correlations between situation-level variables	Is perceived autonomy related to perceived learning on situation-level? Is the extent of asking questions related to perceived learning from feedback?
4. Stability of correlations between situation-level variables across individuals	Does the correlation between perceived autonomy and learning on the situation-level vary across participants? Does the correlation between asking questions and learning on the situation-level vary across participants?
5. Person-level factors of differences in correlations between situation-level variables	Are individual differences in the relationship of autonomy and learning associated with self-efficacy? Are individual differences in the relationship of asking questions and learning associated with prior knowledge? Do agentic individuals engage more often in behaviour conducive to learning than less agentic individuals?
6. Stability of differences in situational variables within subjects	Do interns perceive higher job demands in the first week as opposed to their last week of internship? Do new employees ask more experienced colleagues for help more often during the first month as opposed to their sixth month?
7. Temporal patterns of situational variables	Do perceived job demands follow a recurrent pattern, for instance, in the course of a week? Does perceived learning follow a recurrent pattern, for instance, in the course of a day?

2014), some of which are discussed in more detail in the following subsections. In addition to the options of the parameters we also provide comments that result mainly from our own research experience.

Table 3.3 Overview of the main parameters of a diary and the typical options

Parameter	Typical options	Comments
Phenomena of interest	*Internal*: Emotion, cognition, learning etc. *External*: Situational conditions, behaviour of others etc. *Behavioural*: particular workflows, help-seeking, etc.	Usually, a combination of objective situational characteristics and more subjective reactions is recommendable. Usually, the main research interest also determines the name of the diary. However, care must be taken so that participants are not influenced by the name of the diary (e.g., 'informal learning diary').
Recording device	*Paper-and-pencil* diaries *Web-based* diary applications for bigger screens *Apps* for mobile devices *Other* devices (e.g., voice recorders, video cameras)	The working context of the participants should be taken into account before deciding in favour of any recording device. Regarding other devices, one of the illustrating studies implemented small digital voice recorders.
Recording media	*Written* format *Voice* recordings *Image* or *video* upload	As it is evident that the recording device and recording medium must be matched to each other, the method of analysis must also be considered.
Item format	*Free text* requests: String variables Simple *check boxes*: Dichotomous variables *Standardised (Likert) scales*: Ordinal or continuous variables *Complex items* such as the circumplex item of emotional states (CES, see Fig. 3.1 in Sect. 3.3.1.2) *Multimedia items* (audio, video etc.)	Scales consisting of five or more items should be avoided; single items are often used instead (Ohly et al., 2010). Free-text responses allow us to gather contextual data. However, they take more time to fill in and depend on participants' literacy competences. It is also advised to reduce cognitive load by maintaining the order of the items.
Situational adaptability of items	*Generic items* fit all possible situations and differ only in the extent of agreement (e.g., interest, newness) *Specific items* are adapted based on previous information on the situation (e.g., after receiving feedback an apprentice is requested to indicate whether the feedback was positive or negative)	Although possible, the use of specific items is much harder to realise with paper-and-pencil diaries than with digital applications that allow researchers to programme conditions.
Sampling method	*Time-based sampling: fixed* = disclosed sampling schedule, i.e. predictable from the participants' point of view; *variable* = random from the participants' point of view *Event-based sampling*: Predefined triggers	The sampling methods can also be combined, for instance, when there is a time-based request to record the most difficult work task during the last hour.

(continued)

Table 3.3 (continued)

Parameter	Typical options	Comments
	(e.g., an error occurred) instead of a time schedule.	If the predefined events occur too often to be completely reported, participants might be requested to make their own selection, for instance, based on the representativeness of a current event.
Reference time	*Point in time*: (a) Experience exactly at the time of the entry or (b) at a particular point during the event (in case of event-sampling). *Period of time*: (c) Average experience during the last x minutes or (d) average experience since the last entry.	Study I in this chapter illustrates, for instance, the importance of the reference time when asking for emotions during an error episode.
Time delays	*Immediate recording*: Record made immediately after signalling or after the triggering event; possible option to control for time delays by means of electronic recording methods. *Delayed recording*: Time delays between trigger and record are permitted and controlled for.	Immediate recording might not be possible in work contexts in that incumbents do not have autonomy to disrupt their working activities. When using digital devices, one can calculate the time delay between the event and the entry from the log data, if the participants are to record the time of the underlying event in the diary.
Study period	*Single study period*: one continuous diary period. *Multiple study periods for change analysis*: several diary periods with longer intervals in between. *Extended single study period*: extended time span but reduced schedule (e.g., only one entry per week). *Multi-group study periods*: multiple groups with time-shifted data collection plans for each group.	If participants are requested to make several entries a day, the study period should not be longer than two or three weeks. Multiple-group designs require large samples and a randomised assignment to groups.

3.2.3.1 Item Format, Adaptability, and Recording Device

Closed items reduce participant burden because they can be answered fast. However, if the phenomena to be captured are very diverse, it could become difficult to collect enough context information to validly interpret the data. In a work task diary, for instance, it is difficult to define task categories that are meaningful, exhaustive, and disjunctive (Rausch, 2014). Therefore, it may be advisable to let the participants describe contextual characteristics in their own words using open answer formats and to categorise these descriptions afterwards using content analysis. Nowadays, natural language processing can help to categorise those descriptions.

Very short scales or even single items for each construct are advisable in diaries because the time for each diary entry is usually short, particularly if several diary entries are expected during a day. However, Ohly et al. (2010) rightly caution that the validity of shorter survey instruments may be limited. Furthermore, if the events of interest allow for a broad range of different situations (e.g., samples of work activities), it must still be ensured that all items are appropriate for each possible event (see 'generic items' in Table 3.3). Further recommendations concerning the choice of scales and items can be found in Ohly et al. (2010).

Digital recording devices enable a conditional selection of items during a diary entry (see 'specific items' in Table 3.3). For instance, when recording a customer consultation, different items can be presented to the participant than when recording a team meeting. Furthermore, digital diaries allow for more complex items such as the circumplex item of emotional states (see Study I in this chapter) or multimedia items such as audio input in the case of voice recording (see Study II in this chapter).

3.2.3.2 Sampling Method

The sampling method refers to the trigger of a diary entry. Two basic strategies can be distinguished. In *time-based sampling*, a time schedule set by the researcher triggers a diary entry. These sampling times can be fixed (e.g., 9 am, 11 am, 3 pm) or variable so that the exact times cannot be predicted by the participants. Csikszentmihalyi and colleagues used the latter procedure in most of their classical studies (Kubey et al., 1996). However, if the sampling times are predefined by others, situational circumstances at the time of the request may impede a diary entry. Furthermore, particular phenomena of interest such as errors or interactions may just be missed. Finally, the external regulation of the sampling times may also affect the participants' experience of self-determination and thus lead to decreased compliance.

In *event-based sampling*, participants are requested to make a diary entry each time a predefined event occurs. These events of interest can be internal, external or behavioural (see Table 3.3). In event-based sampling, it is crucial that all participants have the same understanding of the events of interest and recognise occurrences of these events during their daily work (for instance through an a priori workshop to train participants, see Sect. 3.2.3.4). However, this may also prime the participants towards the events of interest and lead to biases in the occurrence, perception or reporting of the events (treatment effect, see Sect. 3.2.1). Moreover, it is difficult to estimate the response rate because the true number of events is not known. Finally, if the events of interest occur too frequently, the participants are overwhelmed with recording them in the diary (Bolger & Laurenceau, 2013; Bolger et al., 2003; Fischer & To, 2012; Ohly et al., 2010; Rausch, 2014; Wheeler & Reis, 1991).

3.2.3.3 Reference Time and Time Delays

While diary methods aim at a real-time collection of data, diary data often involves some degree of retrospection (Beal & Weiss, 2003; Shiffman et al., 2008). In research on workplace learning, a respective prompt would be: 'Please think of your most difficult work task this morning and answer the following questions!' The further the data entry is from the actual event, the more likely memory bias is to occur. For example, the typical filling in of the diary at the end of the day may mean that changes in emotional state over the course of an error episode that day can no longer be accurately recalled. Furthermore, this particular time of the day may be a source of systematic bias because participants may be particularly fatigued (Shiffman et al., 2008). In addition, in event-based sampling, a time-delayed data collection is to be expected, because the event to be recorded extends over a certain period of time and is usually only recorded afterwards (Beal & Weiss, 2003). This can be an issue, for instance, when collecting data on events in the course of which strong fluctuations in emotional experience are to be expected. This is the case for error episodes or handling complex problems that take some time. Therefore, to enable data entries at an early stage of problem solving, Rausch et al. (2015) used a web-based diary that allowed participants to record problems that had not yet been solved and then complete these entries later.

3.2.3.4 Study Period and Overall Research Design

Diary studies typically require more preparation than a single cross-sectional questionnaire study or a single interview study. In a diary study, it is recommended to have a meeting or a training with the prospective participants in advance to foster compliance, to explain the sampling method of the diary, to ensure a shared understanding of the events of interest (event-based) and a shared understanding of the diary items, and to clarify any data protection concerns. Moreover, a traditional self-report questionnaire on stable characteristics is usually administered before the diary period (Ohly et al., 2010) and written consent can be obtained on this occasion. According to our experience, this personal contact strongly helps to increase compliance of the participants. Compliance in this context means '... that all questions are answered for every diary questionnaire at the time required by the study design' (Ohly et al., 2010, p. 86).

Regarding the study period, three modes can be distinguished: (1) A *single study period* is applied in most diary studies. This mode is appropriate if developmental processes in the phenomena of interest are not to be investigated. Each diary entry is then treated as a representative of this time period, independent of its concrete measurement time. Nevertheless, time influences can be investigated to a certain extent, for example by examining whether the emotional experience on Mondays differs from that on Fridays. (2) *Multiple study periods* allow for change analysis. In the simplest case, there are two diary periods with a longer interval between them in which some kind of change process is suspected to take place. (3) An *extended single study period* is applied if development processes are to be investigated. However, regarding participant burden and resulting compliance issues, a data collection over several months requires a less tight schedule for data collection with, for instance, only one or two entries per week. (4) *Multi-group study periods* allow the seamless investigation of development processes, for instance when investigating onboarding processes of a cohort of new employees. To avoid an excessively long, tightly timed diary period, the sample is divided into several groups with staggered data collection schedules. This complex design is recommended when longer-term development processes are to be recorded at a high rate. In a four-group study design, for instance, group 1 keeps the diary in week 1 and then has a three-week break; group 2 keeps the diary in week 2 and so forth.

3.2.4 Reporting Diary Research

Diary studies are by design often more complex than retrospective studies that many researchers are familiar with. It is therefore important that researchers who report a diary study do it in a way that allows other researchers (a) to fully understand what has been done and therefore to help them to interpret the reported results in terms of objectivity, reliability, and validity as well as (b) to replicate the study to check the robustness of the findings. An informative guideline on how to report dairy studies has been provided by Stone and Shiffman (2007). In the following, we briefly summarise the main points of their excellent paper and include our own experience in reporting diary research:

- *Sampling and diary procedure*. It should be clearly described what sampling method (see Sect. 3.2.3.2) including the sampling schedule (see Sect. 3.2.3.3) as well as the data collection mode (see Sect. 3.2.3.1) was chosen and how this corresponds to the research question(s) as well as the underlying theoretical framework. The readers should get a good picture of how the study was conducted and what the methodological as well as theoretical rationales were.
- *Compliance and missing data.* Due to the high demands made on participants, diary studies are prone to non-compliance resulting in delayed data recording or missing data entries. Researchers are therefore advised to transparently describe compliance rates (i.e., percentage of events that are reported based on the expected number of reported events). Especially with event-based diaries this might not always be possible. A way around this issue is to request the participants to estimate the percentage of documented problem cases out of the total number of problem cases and in addition, provide reasons for why they did not record more cases (see for example Rausch et al., 2015). Furthermore, researchers should elaborate on how diary entries that have not been recorded in full compliance (e.g., delayed entries) and missing data are treated (Grund et al., 2021).
- *Training and monitoring of study participants*. Not only are dairy studies more complex from the researcher's perspective than retrospective self-report data collection methods, but the same is true for study participants. It is therefore advised that participants are a priori trained on how to record data in compliance with the particular research design (see Sect. 3.2.3.4). This training procedure needs to be well-explained within a research report as it allows other researchers to assess the quality of the data and consequently the study findings. In addition,

any procedure to monitor and intervene in the participants' compliance needs to be described.

- *Data management and analysis*. The complexity of a diary study is also reflected in the structure of the resulting data and how data is managed as well as analysed. It is important that other researchers understand how this data was prepared for analysis (e.g., aggregation, handling of missing data) as well as what data analysis approaches are used (time-series analyses, hierarchical linear modelling, etc.) as, depending on the question, a variety of analytical methods can be considered.

3.3 Illustrating Diary Research

3.3.1 Study I: Using a Paper-Based Diary to Investigate Emotions During Error Situations

3.3.1.1 Context and Research Questions

Errors occur within goal-directed actions and become observable in a deviation of an actual state of goal achievement from an expected state of goal achievement (Rausch et al., 2017). Although errors are usually unwanted in the workplace, they are also considered an important opportunity to learn (Gartmeier et al., 2008; Rausch et al., in print; Zhao et al., 2014). However, errors – as an external performance feedback – usually evoke negative achievement emotions (Pekrun & Perry, 2014). Shame, guilt, anger, embarrassment, worry, disappointment, sadness, or fear are typical emotions after error detection (Basch & Fisher, 2000; Newton et al., 2008; Oser & Spychiger, 2005; Rausch et al., 2017; Zhao et al., 2014). Such negative emotions serve as warning signals that may trigger reflection and problem-based coping approaches that foster learning (Zhao et al., 2014; Tulis et al., 2016). However, negative emotions may also lead to disengagement, blaming others, and covering up errors (Dörner, 1996; Schwarz & Bless, 1991; Zhao et al., 2014). Negative emotions may require emotion-based coping approaches (Brown et al., 2005; Putz et al., 2013; Zhao et al., 2014). After coping, positive emotions may dominate (Basch & Fisher, 2000; Ohly & Schmitt, 2015). Furthermore, individual dispositions supposed to influence emotions and coping approaches in error situations are summarized as error orientation (Rybowiak et al., 1999). The present study (Rausch, in preparation) is in part a replication study of Rausch et al. (2017) and addresses the following research questions (RQ):

1. Does the type of error influence the perceived learning from an error?
2. How do emotional states change over the course of an error situation?
3. How do emotional states at the beginning of an error situation influence coping strategies and perceived learning?
4. How are coping strategies related to perceived learning outcomes?
5. Are there plausible correlations between aggregated diary data and questionnaire data on the person-level?

Due to limited space, only RQ 2 and 5 are illustrated in this section. Both research questions provide exemplary perspectives on validity issues that will be discussed in the conclusion.

3.3.1.2 Method

Fifty-five novices employed at a Rural District Office of a medium-sized city in Germany were approached to participate in the voluntary study. A total of 51 VET students, dual students and inexperienced middle-grade public officers participated, 32 of which were female.

The study started with a 2.5 h workshop which contained an introduction to the study and the applied diary, as well as the administration of a self-report questionnaire on biographic information, personality traits, error orientation, team psychological safety and team learning behaviour. The diary period started after the workshop and lasted for 20 workdays. After the diary period, the participants were requested to complete a short questionnaire concerning the use of the error diary.

Forty-three of the initial 51 participants took part in the diary study and altogether, they recorded $k = 352$ error situations within 20 workdays. Analysis on situation-level is based on these 352 error situations. When aggregating error situations on person-level, participants who recorded less than five error situations were excluded. Thus, analysis on person-level is based on $n = 34$ individuals.

The diary was implemented in the form of a small paper booklet with a cover sheet followed by five-page forms for 20 entries. The diary form started with several closed and free-text items on the error situation, possible causes for the error and the detection of the error. Then, emotional states were measured by using the circumplex item of emotional states (CES item, see Fig. 3.1). Different emotional states are

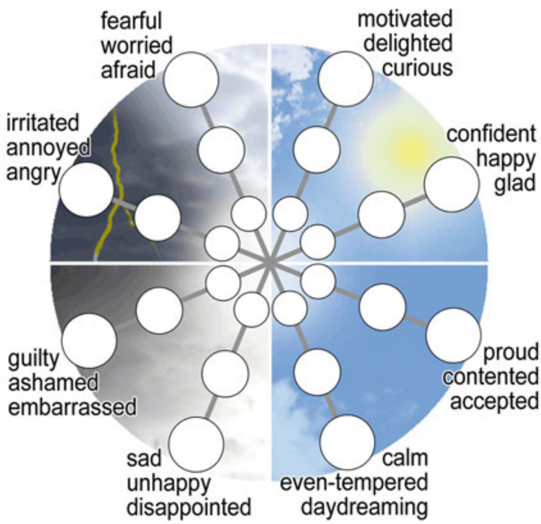

Fig. 3.1 Circumplex item of emotional states. (CES, Rausch, 2014)

arranged in a circle with states of high arousal in the upper half, states of low arousal in the lower half, positive states in the right half, and negative states in the left half. From a technological point of view, it is an image map that shows labels of emotional states on background pictures with corresponding weather conditions (e.g., a thunderstorm for negative emotional states with high arousal as shown in Fig. 3.1). Clickable circles of different sizes can be used to indicate whether and how strongly an emotional state was experienced (coded from 0 = *not at all*, if not selected to 3 = *strong*, if the outer circle was selected). Different from Rausch et al. (2017), the participants were explicitly asked to indicate how they felt directly after the detection of the error. The following items referred to coping approaches (problem-focused coping, emotion-focused coping, and self-protection). On the next page, the CES item was presented, again. This time participants were asked to indicate how they felt in later stages of the error situation. Further items referred to subjectively perceived learning from the respective error situation. All items were discussed in the workshop beforehand. Table 3.4 provides an overview of the applied diary method using the parameters introduced in Table 3.3.

Error Orientation as an individual disposition was measured by using the error orientation questionnaire (EOQ) by Rybowiak et al. (1999). The questionnaire comprises 37 items in eight scales. The items are answered on a four-point Likert scale from 1 (“does not apply at all”) to 4 (“fully applies”).

In the short questionnaire after the diary period, the participants answered questions regarding the error diary. On average, the participants estimated that they recorded 71% of the actual error situations (SD = 22.6). The participants were also asked to rate reasons why they have not recorded more error situations (answers in descending order): having forgotten to make an entry (M = 4.09; SD = 1.49), no time to make an entry (M = 3.63; SD = 1.56), making an entry was too effortful (M = 3.09; SD = 1.68); being afraid that others would consider the errors as negligible did not play a role (M = 1.45; SD = 1.23). Being asked how difficult it was to answer the diary items (from 1 = “very easy” to 6 = ”very difficult”), the participants, on average, did not find it too difficult (M = 2.66; SD = 1.11). We also asked in an open format to state which part of the diary was most difficult to complete. Half of the participants mentioned the indication of emotional states was the most difficult part.

To address RQ 2, differences between emotional states before and after coping were analysed using t-tests for dependent samples. To address RQ #5, Pearson correlations between questionnaire data and aggregated person-level diary data were calculated.

3.3.1.3 Findings

Regarding RQ 2, Table 3.5 shows that directly after the detection of an error, negative emotions prevailed while after coping, all negative emotions decreased significantly, and all positive emotions increased significantly.

Table 3.4 Summary of the diary parameters in Study I

Parameter	Implementations	Comments
Phenomena of interest	*Internal*: Emotions after error detection and after coping; cognitive evaluation of error situation and of one's own learning from the error *External*: Context information on error situation, behaviour of others in the error situation *Behavioural*: Applied coping strategies	The research interest was in the interplay of error characteristics, emotions, coping and learning. Hence, a broad array of phenomena was covered in the diary. The diary was referred to as 'work diary' to avoid overemphasising the concept of error.
Recording device	*Paper-and-pencil* diary	
Recording media	*Written* format	
Item format	*Free text* requests Simple *check boxes* *Standardised (Likert) items* *Complex items*	'Describe in a few sentences how the error situation developed.' 'Which statement applies most?' (check boxes) 'To what extent could you learn something from this mistake?' (Likert scale) Circumplex item of emotional states (CES, see Fig. 3.1)
Situational adaptability of items	*Generic items*	All items were generic items that were to be answered for all error situations.
Sampling method	*Event-based sampling: An entry should be made each time an error occurred.*	It was explained and discussed what an error situation is, in a workshop beforehand.
Reference time	*Point in time*: All items referred to the particular error situation.	For instance, emotional experience at the time of error detection.
Time delays	*Immediate recording*	The entry should be made as soon as possible after the clarification of an error situation. However, it cannot be verified whether the entries were really made in a timely manner. Furthermore, if the clarification of an error takes longer, the detection of the error has already taken place some time ago. Hence, certain delays were unavoidable.
Study period	*Single study period for* one continuous diary period.	Twenty workdays

Regarding RQ 5, Table 3.6 shows that no significant correlations between scales of the error orientation questionnaire and corresponding diary items could be found.

Table 3.5 Differences between initial and later emotional states (t-test for dependent samples)

Emotional states	Emotional states at error detection		Emotional states after coping			
	M	*SD*	*M*	*SD*	*p*	*d*
Sad/unhappy/disappointed	0.56	0.89	0.28	0.78	<.001	0.34
Guilty/ashamed/embarrassed	1.14	1.11	0.53	0.96	<.001	0.58
Irritated/annoyed/angry	0.86	1.05	0.48	0.95	<.001	0.38
Fearful/worried/afraid	0.73	1.02	0.34	0.87	<.001	0.41
Motivated/delighted/curious	0.22	0.63	0.63	1.04	<.001	−0.46
Confident/happy/glad	0.36	0.76	1.05	1.18	<.001	−0.68
Proud/contented/accepted	0.07	0.38	0.39	0.91	<.001	−0.44
Calm/even-tempered/day-dreaming	0.51	0.83	0.88	1.11	<.001	−0.37

Note: Each time the circumplex item of emotional states is presented, up to four states can be rated from 1 (“a little bit”) to 3 (“very”), all other emotional states were coded zero (“not at all”)

Table 3.6 Selected correlations between EOQ scales and diary items

EOQ scales (Cronbachs Alpha)	Example item from EOQ scale (number of items in scale)	Diary items (aggregated to person-level)	r
Learning from errors (*C.A.* = .65)	Mistakes assist me to improve my work. (4)	Perceived learning from an error.	.07
Error strain (*C.A.* = .80)	If I make a mistake at work, I `lose my cool' and become angry. (5)	Blaming oneself for the error.	.20
Covering up errors (*C.A.* = .82)	It can be useful to cover up mistakes. (6)	Ensuring that only few people get aware of my error.	.18
Thinking about errors (*C.A.* = .78)	After I have made a mistake, I think about how it came about. (5)	Finding out how and why this error could occur.	.05

Note: $n = 34$; no significant results ($\alpha = .05$)

3.3.1.4 Conclusions from Study I

The selected and preliminary results from this study on learning from errors in the workplace provide two interesting insights on the validity of diary data. First, in accordance with theoretical considerations, emotional states changed significantly in the course of an error episode. At the same time, this also demonstrates that when asking about emotional experiences, a shared understanding of the reference time is quite crucial. If one participant tends to rate his or her emotions at the time of error detection and another participant rates his or her emotions after coping with an error situation, the resulting data does not refer to the same phenomenon. In general, content validity is always questionable when participants have divergent ideas about what exactly is meant by a particular item. This underlines the importance of clear item formulations and good training of the participants.

Second, retrospective questionnaire data and situational diary data often do not show the expected correlations. From a theoretical perspective, individual dispositions are tendencies towards certain emotions or behaviours in certain situations. The general perception that mistakes help to improve one's own work should be evident in the perception of learning in concrete error situations. The latent tendency to cover up errors should become manifest in error situations in that people try to ensure that only few people get aware of their errors. However, none of these plausible correlations could be found in the data. This resembles the findings reported by Rausch (2012) and, in our opinion, this speaks against the validity of self-report questionnaires that require an accurately averaged assessment of one's experience and behaviour in particular situations over a longer time-period. Hence, in the case of investigating the less conscious everyday processes of informal workplace learning, the diary method provides the more accurate insights.

3.3.2 Study II: Using Voice-Recorded Diaries to Research Nurses' Learning at Student-Run Hospital Wards

3.3.2.1 Context and Research Questions

Student-run hospital wards (SHW) are institutionalised learning arrangements in which nursing students are fully responsible for the planning, organisation, and implementation of patient care within a real-existing hospital ward for a limited time frame during their initial nursing training (Goller et al., 2020; Hauck & Schuster, 2014; see also Grealish & Trede, 2013). Throughout this time, nursing students take over all ward duties and are only shadowed by a few already qualified nurses who may intervene in life-threatening situations or are available for pressing questions concerning patient care. SHW are accompanied by daily/weekly debriefings conducted by the training hospital and/or the nursing school aiming to maximise learning outcomes and to support students during this intensive learning phase. Within the literature, SHW are mostly thought to provide students with a powerful workplace learning setting that is able to compensate for existing deficits in the current training scheme. In fact, it has been reported that nursing students are often only entrusted with a limited set of isolated and compartmentalised auxiliary tasks during the normal workplace assignments as part of their apprenticeships (e.g., Lauber, 2017). SHW on the contrary are designed in such a way that nursing students are able to take over the full range of nursing tasks that fully qualified nurses usually engage in. Moreover, SHW should enable nursing students to more often (a) socially interact with other hospital stakeholders (like other nurses, physicians), (b) consult media resources (like handbooks, journals), as well as (c) obtain feedback and instruction from already qualified nurses. All these activities are also known to be conducive to learning but not often afforded to nursing students during normal workplace assignments.

Although nurses in Germany have been trained in this special workplace learning setting for almost 20 years, the literature lacks empirical studies investigating such arrangements that go beyond simple evaluations focussing on participants' satisfaction and perceived workload (e.g., Witte et al., 2016). Moreover, the few existing studies investigating SHW in a more elaborated way rely on retrospective data collection methods only (see Hertel & Bergjan, 2015 as well as Goller et al., 2020; Goller & Steffen, in press). In-situ evidence about what learning activities nursing students are engaging in at SHW is still missing. The aim of this study therefore was to address the following research question using diaries for data collection (see Goller & Steffen, in press for a more detailed account of the study presented here): What working situations do nursing students experience as relevant and conducive to learning during their time at the SHW?

3.3.2.2 Method

To answer the research question, two three-week SHW conducted in a German hospital over 2 consecutive years with different cohorts were investigated. The 48 participating nursing students were asked to record all kinds of learning opportunities they engaged in during their time at SHW. Each student was equipped with a small digital voice recorder to enable audio recordings. All study participants were briefed a priori about the study goals, the data collection method, as well as the meaning of a provided information card (see Fig. 3.2). They were supposed to carry the card with them at all times since it explained what kinds of situations the study addressed and also contained instructions on what information the students should record.

In total 29 of the 48 nursing students recorded learning situations ($n_{\text{Year1}} = 15$; $n_{\text{Year2}} = 14$). In the first year 107 diary entries with a length of 2:18 h ($M = 1{:}17$ min, $SD = 1{:}09$ min) and in the second year 47 entries with a length of 2:58 h ($M = 3{:}48$ min, $SD = 3{:}00$ min) were recorded. The audio material was then transcribed verbatim, anonymised, and analysed using qualitative content analysis (Kuckartz, 2014). Both deductive and inductive coding was applied. After coding all coded text passages were paraphrased and summarised. Those summaries were used to identify common themes in the recordings.

The decision in favour of open voice recordings instead of items using a written pen and paper format to collect diary data was based on a pilot study one year before the main study. Nursing work on hospital wards is usually undertaken under a tight schedule. As a result, in the final evaluation of the pilot study most students reported to not have enough time to fill in the provided paper-based diaries. Digital diaries were not possible as smartphones or handheld devices are not allowed to be used by nurses during their shift on the hospital ward. This resulted in very low response rates in the pilot study. The use of voice records was proposed and after a discourse with the nursing students from the pilot study they supported the idea with positive feedback and evaluated it to be way more practical. A detailed description of the data collection can be found in Table 3.7 using the parameters introduced in Table 3.3.

Information about voice recording

We are interested in how you are able to develop yourself as a nurse during your time on the student-run hospital ward. For this purpose, we would like to know due to what situations on the student-run hospital wards you could learn something.

For instance, learning at work can be induced by the following events:

In the case ...
- ... someone is **explaining** something to you.
- ... you can **observe** someone working.
- ... you **discussed** a work-related topic with someone.
- ... you **did** something the very **first time**.
- ... you **did** something the **first time** in a **certain way**.
- ... you **read** about some work-related matter.
- ... you **reflected** a situation within your team.

You know that you have learnt something new, in the case of ...
- ... you had a **new insight**.
- ... you can **do something new**.
- ... you can **do something better** than before.
- ... you newly **recognised** that some **things** are **interconnected**.
- ... your newly **recognised** that some **things** are **different**.
- ... you could **clear up** a personal or professional **misunderstanding**.
- ...

We ask you to **record the following information**:
- Detailled description about the experienced situation.
- What did you learn?
- Why did you learn something in your opinion?
- Do you think you could have learnt this during your normal workplace assignments? If not, why not?

We would be happy if you could record as many diary entries as possible. Please record as much information as possible.

Fig. 3.2 Data collection instruction. (Translated from Goller & Steffen, in press)

3.3.2.3 Findings

Not all voice recordings contained information that could be used for the analyses. However, the majority of the recordings made rich statements about situations that were perceived as noteworthy and conducive to learning. Without going into too much detail the nursing students reported three main reasons why they could learn during their time at the SHW:

1. *Learning opportunities*: SHW afford many opportunities to engage in work tasks that are described as being conducive to learning. That is mainly because students were able to perform a variety of significant and complex tasks they are not

Table 3.7 Summary of the diary parameters in Study II

Parameter	Implementations	Comments
General		
Phenomena of interest	*Behavioural*: (a) Task and problem solving, (b) social interaction with other hospital stakeholders, (c) consultation of media resources, (d) obtaining feedback and instruction	The main research interest was the same in both surveys. The first diary was named "Shift diary", but the student nurses referred to it as "SHW diary". That name was used for the second diary.
Study period	*Single study period for* one continuous diary period.	Data collection lasted the full three weeks of each SHW.
Pilot study		
Recording device	*Paper-and-pencil* diaries	Paper diaries were provided at the nurses' station with a box to anonymously submit them.
Recording media	*Written* format	
Item format	*Free text* requests (e.g., "How did the error occur?") Simple *check boxes (e.g.,* "Who assigned you this task?" 0 = "Somebody else", 1 = "Self-assigned") *Standardised (Likert) scales* (e.g., "I was under much time pressure today." 1 = "do not agree" to 5 = "totally agree")	The first set of questions consisted of standardised items and check boxes. In the following, free text requests and standardised items alternated for each learning situation.
Situational adaptability of items	*Generic items* (e.g., see above) *Specific items* (e.g., for an error situation: "I have addressed the error in the group." 1 = "do not agree" to 5 = "totally agree")	The first set of questions were generic. The second set of questions regarded specific learning situations and students were asked to fill them in if they applied.
Sampling method	*Fixed time-based sampling*: at the end of every shift	The nursing school offered the students additional time to fill in the diary after each shift. Unfortunately, but understandably, they instead invested that into pressing care work.
Reference time	*Period of time*: average experience during the last shift	The first set of questions asked about the last shift in general and the second set asked about specific events during that shift.
Time delays	*Delayed recording*: Time delays between trigger.	The diaries were filled out after the shift, as a result there was a time delay between the event of interest and the survey. Students hardly found time to fill the diaries out during their working hours.

(continued)

Table 3.7 (continued)

Parameter	Implementations	Comments
Main survey		
Recording device	Voice recorders	Every student nurse received a digital voice recorder to record spoken messages.
Recording media	*Voice* recordings	
Item format	*Free text* requests (e.g., open end questions)	Due to the format and practicability, only free text was requested (see Fig. 3.2).
Situational adaptability of items	*Generic items* (e. g "Why did you learn something in your opinion?")	As a lesson from the pilot study, the items were also simplified and generalised.
Sampling method	*Event-based sampling*: situations conducive to learning (e.g., explanations, discussions, reflection).	Student nurses were asked to record their diary entries directly after the events they refer to. Record data and the students themselves revealed that this was not always the case (e.g., some students used the recordings to reflect on their shift on their way home).
Reference time	*Point in time*: after the event	
Time delays	*Immediate recording*: Record made immediately after the triggering event	

usually allowed to take over and that go far beyond auxiliary tasks afforded during their regular practice phases. In addition, they could exercise agency by taking charge of their own actions including the planning and the evaluation of their performance.

2. *Social interaction*: They had to independently communicate and cooperate with peers and other medical professions. This social exchange allowed them to co-construct knowledge as well as to learn through observation.
3. *Feedback*: Students had the chance to regularly get feedback on their performance from more experienced nurses, other medical staff, and each other. This supports their reflection about their professional experiences during the SHW as well as their current competence level with regard to different nursing procedures.

In many recordings SHW were explicitly or implicitly compared with the normal workplace assignments that are part of students' apprenticeships. It was emphasised that SHW are particularly conducive to learning as the normal workplace assignments do not allow to self-responsibly engage in all kinds of relevant working tasks, to interact with other care-related or medical stakeholders, and to get constant feedback on their own working behaviour.

3.3.2.4 Conclusions from Study II

On a content level, this study mainly confirmed findings from prior research endeavours using retrospective data collection methods that investigated SHW conducted at the same hospital (Goller et al., 2022) as well as reported elsewhere in the literature (e.g., Witte et al., 2016). In this regard, the diary study mainly produced findings that triangulate other findings based on other research approaches like post-intervention interviews and questionnaires. At the same time, the diary approach allowed to document specific work situations that are perceived as relevant and conducive to learning from the nursing students' point of view. It can be assumed that these reports are more detailed and more valid than retrospective descriptions that refer to events that happened between one or three weeks ago.

At the same time, at least three critical issues should be noted: (a) Out of the 48 study participants only 29 students recorded learning situations. In addition, some participants only recorded one or two situations within two weeks. (b) Not all recordings were as detailed and rich as hoped for. (c) Not all recordings were recorded directly after the situation occurred at the workplace. In fact, a few students made explicit that they were only able to record after their shift. Based on post-study discussions with the nursing students it became clear that, similar to the written versions of the diaries tested in the pilot study, students also struggled to find time for voice recordings during their shifts. Researchers investigating professional domains that are characterised by tight work schedules and time pressure should be aware that participants–even when they are willing to take part in a study–might not find sufficient time to participate in data collection in situ.

3.4 Conclusion

For many research questions on informal workplace learning, diary data is more valid than data gained from typical retrospective measures. Furthermore, diary data allows for the analysis of situational influences. Data from retrospective self-reports may, for instance, reveal correlations between certain work characteristics such as autonomy and learning but that does not necessarily mean that autonomy in a particular situation is linked to learning in this situation. Moreover, diaries enable the analysis of change processes over time periods and of course, also the analysis of interpersonal differences (Beal & Weiss, 2003; Bolger et al., 2003 ; Fisher & To, 2012; Ohly et al., 2010; Reis & Gable, 2000; Shiffman et al., 2008; Wheeler & Reis, 1991). However, the effort of participating in a diary study may lead to reactivity (i.e., an atypical change of the natural process such as induced reflection), or to reactance (i.e., intentional dropout or even sabotage), both of which threaten the validity of a study (Rausch, 2014). These possible downsides can be mitigated by carefully designing the diary study.

In this chapter, we have discussed the phenomena of interest, recording devices, recording media, item formats, adaptability of items, sampling method, reference time, time delays, and the study period as parameters in designing a diary study. Furthermore, we illustrated the diary method by two studies on informal workplace learning. Study I by Rausch (in preparation) investigated novices' learning from errors in everyday work and implemented an extensive paper-based diary. Study II by Goller and Steffen (in press) investigated student nurses' informal learning during a special instructional setting (student-run hospital wards) and implemented voice recording. We are aware that in briefly illustrating two diary studies we do not adhere to the reporting standards that we recommended in Sect. 3.2.4. However, in the context of this chapter, the studies only serve to illustrate different contexts of application, different objectives, and different implementations of research diaries. The studies will be published in full, elsewhere.

Though diaries still provide self-reported data 'they often represent a substantial improvement over more common retrospective methods' (Reis, 2012, p. 5). Hence, further following Reis (2012), we would like to emphasize that we are not assuming that retrospective surveys are 'wrong' and diaries are 'right' but, as for every method, there are certain conditions under which one should be favoured over the other. Based on literature and our own experiences, when researching informal learning in the workplace, these conditions are: (1) The phenomena of interest or their characteristics tend to go unnoticed in the daily workflow (e.g., some types of informal learning). (2) Some of the phenomena of interest or their characteristics occur only rarely and are therefore difficult to remember in retrospect (e.g., particular tasks or feedback situations). (3) The research focuses on the relations of fluctuating situational characteristics, for instance, the effects of task characteristics on perceived learning.

Future perspectives on diary studies involve technological developments such as custom-built smartphone apps for diary research (e.g., mQuest® by cluetec, Karlsruhe, Germany). Furthermore, the combination of diary self-reports with other sources of longitudinal data such as log data, movement data or sociometric badges open up new paths for the analysis of informal workplace learning. However, more traditional approaches such as a triangulation of diary data and, for instance, interview data (e.g., Sharples & Cobb, 2015) are very promising and are rarely utilised.

References

Basch, J., & Fisher, C. (2000). Affective events-emotions matrix: A classification of work events and associated emotions. In N. Ashkanasy, C. Härtel, & W. Zerbe (Eds.), *Emotions in the workplace: Research, theory, and practice* (pp. 36–48). Quorum Books.

Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods, 6*(4), 440–464. https://doi.org/10.1177/1094428103257361

Berings, M. G. M. C., Doornbos, A. J., & Simons, P. R.-J. (2006). Methodological practices in on-the-job learning research. *Human Resource Development International, 9*(3), 333–363. https://doi.org/10.1080/13678860600893557

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Brief, A. P., Butcher, A. H., & Roberson, L. (1995). Cookies, Disposition, and Job Attitudes: The Effects of Positive Mood-Inducing Events and Negative Affectivity on Job Satisfaction in a Field Experiment. *Organizational Behavior and Human Decision Processes, 62*(1), 55–62. https://doi.org/10.1006/obhd.1995.1030

Brown, S. P., Westbrook, R. A., & Challagalla, G. (2005). Good cope, bad cope: Adaptive and maladaptive coping strategies following a critical negative work event. *Journal of Applied Psychology, 90*(4), 792–798. https://doi.org/10.1037/0021-9010.90.4.792

Clardy, A. (2018). 70-20-10 and the dominance of informal learning: A fact in search of evidence. *Human Resource Development Review, 17*(2), 153–178. https://doi.org/10.1177/1534484318759399

Cuyvers, K., Van den Bossche, P., & Donche, V. (2022). Longitudinal case study research to study self-regulation of professional learning: Combining observations and stimulated recall interviews throughout everyday work. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 579–600). Springer.

Dörner, D. (1996). The logic of failure: Recognizing and avoiding error in complex situations. Perseus.

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior, 33*(7), 865–877. https://doi.org/10.1002/job.1803

Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative knowledge: Understanding professional learning and expertise. *Vocations and Learning, 1*(2), 87–103. https://doi.org/10.1007/s12186-008-9006-1

Goller, M., & Steffen, B. (in press). Lernen auf der Schulstation [Learning during student-run hospital-wards]. In M.A. Marchwacka (Ed.), Handbuch Pflegebildung. Theoretische Ansätze, empirische Erkenntnisse und praktische Anwendungen (pp. 182–193). Hogrefe.

Goller, M., Steffen, B., & Lau, D. (2020). Schulstationen in der Pflegeausbildung: Zur Lernförderlichkeit eines komplexen Lehr-Lernarrangements [Student-run hospital wards during nursing training. On the learning potential of a complex learning arrangement]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 116*(3), 364. https://doi.org/10.25162/zbw-2020-0016

Goller, M., Steffen, B., & Lau, D. (2022). Kompetenzerwerb auf der Schulstation: Eine Mixed-Methods-Replikationsstudie [Competence development at student-run hospital wards: A mixed-methods replication study]. In U. Weyland & K. Reiber (Eds.), Professionalisierung der Gesundheitsberufe. Berufliche und hochschulische Bildung im Spiegel aktueller Forschung. Beiheft 33 der Zeitschrift für Berufs- und Wirtschaftspädagogik (pp. 21–47). Franz Steiner Verlag.

Grealish, L., & Trede, F. (2013). Student nurse led ward in aged care. In J. Higgs, D. Sheehan, J. B. Currens, W. Letts, & G. M. Jensen (Eds.), *Realising exemplary practice-based education* (pp. 93–100). Sense Publishers. https://doi.org/10.1007/978-94-6209-188-7_9

Grund, S., Lüdtke, O., & Robitzsch, A. (2021). Multiple imputation of missing data in multilevel models with the R package mdmb: A flexible sequential modeling approach. *Behavior Research Methods, 53*, 2631–2649. https://doi.org/10.3758/s13428-020-01530-0

Hauck, C., & Schuster, E. (2014). Schulstationen – Von der Definition zur Handlungsempfehlung [Student-run hospital wards. From definition to recommendation for action]. *Pädagogik der Gesundheitsberufe, 1*, 70–77.

Hertel, F., & Bergjan, M. (2015). Schulstation als 'Capstone' der Pflegeausbildung – Evaluation einer besonderen klinischen Lernumgebung [Student-run hospital wards as a 'Capstone' in nursing education. Evaluation of a special clinical learning environment]. In F. Arens (Ed.), *Praxisbegleitung in der beruflichen und akademischen Pflegeausbildung: Eine Standortbestimmung* (pp. 224–245). wvb.

Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 54–75). Oxford University Press.

Jobe, J. B. (2000). Cognitive process in self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 25–28). Lawrence Erlbaum.

Kubey, R., Larson, R., & Csikszentmihalyi, M. (1996). Experience sampling method applications to communication research questions. *Journal of Communication, 46*(2), 99–120. https://doi.org/10.1111/j.1460-2466.1996.tb01476.x

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice & using software*. SAGE.

Lauber, A. (2017). *Von Könnern lernen: Lehr-/Lernprozesse im Praxisfeld Pflege aus der Perspektive von Lehrenden und Lernenden* [Learning from skilled professionals. Learning/Teaching processes in the practical field of nursing from teachers' and learners' perspectives] Waxmann.

Littlejohn, A., & Margaryan, A. (2014). Technology-enhanced professional learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 1187–1212). Springer. https://doi.org/10.1007/978-94-017-8902-8_43

Newton, N. A., Khanna, C., & Thompson, J. (2008). Workplace failure: Mastering the last taboo. *Consulting Psychology Journal: Practice and Research, 60*(3), 227–245. https://doi.org/10.1037/1065-9293.60.3.227

Ohly, S., & Schmitt, A. (2015). What makes us enthusiastic, angry, feeling at rest or worried? Development and validation of an affective work events taxonomy using concept mapping methodology. *Journal of Business and Psychology, 30*(1), 15–35. https://doi.org/10.1007/s10869-013-9328-3

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research: An introduction and some practical recommendations. *Journal of Personnel Psychology, 9*(2), 79–93. https://doi.org/10.1027/1866-5888/a000009

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft* [Learning is painful]. Beltz

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Putz, D., Schilling, J., Kluge, A., & Stangenberg, C. (2013). Measuring organizational learning from errors: Development and validation of an integrated model and questionnaire. *Management Learning, 44*(5), 511–536. https://doi.org/10.1177/1350507612444391

Rausch, A. (2012). Prozessnahe und retrospektive Erhebungsmethoden der Arbeitsanalyse in der betrieblichen Ausbildung [Real-time and retrospective data collection methods of work analysis in in-company training]. *Empirische Pädagogik, 26*(2), 247–270.

Rausch, A. (2013). Task characteristics and learning potentials – Empirical results of three diary studies on workplace learning. *Vocations and Learning, 6*(1), 55–79. https://doi.org/10.1007/s12186-012-9086-9

Rausch, A. (2014). Using diaries in research on work and learning. In C. Harteis, A. Rausch, & J. Seifried (Eds.), *Discourses on professional learning* (Vol. 9, pp. 341–366). Springer). https://doi.org/10.1007/978-94-007-7012-6_17

Rausch, A. (in preparation). Emotions, coping, and emotions, again – An extended replication study on learning from errors in the workplace.

Rausch, A., Schley, T., & Warwas, J. (2015). Problem solving in everyday office work – A diary study on differences between apprentices and skilled employees. *International Journal of Lifelong Education, 34*, 448–467. https://doi.org/10.1080/02601370.2015.1060023

Rausch, A., Seifried, J., & Harteis, C. (2017). Emotions, coping and learning in error situations in the workplace. *Journal of Workplace Learning, 29*(5), 374–393. https://doi.org/10.1108/JWL-01-2017-0004

Reis, H. T. (2012). Why researchers should think "real-world" – A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). Guilford Press.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). Cambridge University Press.

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment, 31*(2), 226–235. https://doi.org/10.1037/pas0000662

Rybowiak, V., Garst, H., Frese, M., & Batinic, B. (1999). Error Orientation Questionnaire (EOQ): Reliability, validity, and different language equivalence. *Journal of Organizational Behavior, 20*(4), 527–547.

Sawchuk, P. H. (2009). Informal learning and work: From genealogy and definitions to contemporary methods and findings. In R. Maclean & D. Wilson (Eds.), *International handbook of education for the changing world of work* (pp. 319–331). Springer. https://doi.org/10.1007/978-1-4020-5281-1_21

Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 76–113). Oxford University Press.

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 98–119). Sage.

Schwarz, N. (2012). Why researchers should think "real-time" – A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). Guilford Press.

Schwarz, N., & Bless, H. (1991). Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning. In J. P. Forgas (Ed.), *Emotion and social judgment* (pp. 55–72). Pergamon.

Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.

Sharples, S., & Cobb, S. (2015). Methods for collecting and observing participant responses. In J. R. Wilson & S. Sharples (Eds.), *Evaluation of human work* (pp. 83–118). CRC Press. https://doi.org/10.1201/b18362

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Stone, A. A., & Shiffman, S. (2007). Capturing momentary, self-report data: A proposal for reporting guidelines. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 371–386). Oxford University Press.

Stone, A. A., Shiffman, S., & Atienza, A. (2007). In L. Nebeling (Ed.), *The science of real-time data capture: Self-reports in health research*. Oxford University Press.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–47). Lawrence Erlbaum.

Tulis, M., Steuer, G., & Dresel, M. (2016). Learning from errors: A model of individual processes. *Frontline Learning Research, 4*(4), 12–26. https://doi.org/10.14786/flr.v4i2.168

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*(3), 339–354. https://doi.org/10.1111/j.1467-6494.1991.tb00252.x

Wilson, J. R., & Sharples, S. (2015). *Evaluation of human work* (4th ed.). CRC Press.

Witte, S., Greis, G., & Darmann-Finck, I. (2016). Optimale Bedingungen für die klinische Kompetenzentwicklung auf Ausbildungsstationen/-bereichen [Optimial conditions for clinical competence development in student-run hospital wards/-section]. *Pädagogik der Gesundheitsberufe, 3*(3), 20–27.

Zhao, B., Olivera, F., & Edmondson, A. C. (2014). Learning from errors in organizations: The effects of negative emotions on motivation and cognition. In J. A. Miles (Ed.), *New directions in management and organization theory* (pp. 23–61). Cambridge Scholars Publishing.

Andreas Rausch (PhD) is a full professor of Economic and Business Education – Workplace Learning at the Mannheim Business School (MBS) of the University of Mannheim, Germany. His current research interests include workplace learning, simulation-based learning, domain-specific problem solving, competence assessment, and vocational education and training (VET). He is committed to empirical educational research with a focus on process analysis and non-cognitive facets. Andreas's research is published in international journals. He is associate editor of Vocations and Learning and serves as editorial board member and reviewer for several journals in the field.

Michael Goller (PhD) is a research associate at the Institute of Educational Science at Paderborn University. Between October 2018 and March 2020, he served as visiting professor at the University of Bamberg and between April and September 2021 as a visiting professor at the University of Jena. His current research interests comprise expertise development and workplace learning in general, including in the particular context of digitalisation, as well as learning and development of students at the boundary of institutionalised and less institutionalised learning environments such as internships or particular learning arrangements in VET.

Bianca Steffen (MA) is a research associate and PhD Candidate at the Institute of Educational Science at the University of Paderborn, Germany. In her teaching position, she is mainly responsible for student's methods training. Her current research interests comprise workplace learning and expertise development. In her PhD project, she examines intuitive decision making.

Chapter 4
Using the Vignette Technique to Increase Insight into Professional Development at Work

Veronika Anselmann and Regina H. Mulder

Abstract The Vignette Technique can be used to increase insight into professional development. Professional development can be considered to be a process with certain outcomes, such as knowledge, competences, or behaviour. A challenge, especially when using questionnaires, is to measure the direct link between the characteristics of the context, such as triggers at work, with the components of professional development, such as informal learning activities that emanate from these characteristics of the context. The opportunities of the Vignette Technique are presented in this chapter by describing the characteristics of the method with its vignettes (more specifically concerning the content, style and quality), instructions and answering mode. Furthermore, the other components of an empirical study (the design, instrument and analysis) are described, as well as characteristics of these that need to be decided upon and need to be carefully developed. Emphasized is the need for consistency between the vignettes, instructions and answering modes, as well as consistency in terms of suitable fit to other components as well as to the objective of the study. The possibilities of using the Vignette Technique and the consequences for the design and analysis of the data are illustrated with various examples of our own empirical studies using questionnaires carried out to gain insight into the different components of professional development caused by context characteristics, in different domains.

Keywords Vignette technique · Professional development · Characteristics of vignettes · Development of vignettes · Questionnaires · Context boundedness

---

V. Anselmann (✉)
Institute of Nursing Science, University of Education Schwäbisch Gmünd, Schwäbisch Gmünd, Germany
e-mail: veronika.anselmann@ph-gmuend.de

R. H. Mulder
Faculty of Human Sciences, University of Regensburg, Regensburg, Germany
e-mail: regina.mulder@ur.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_4

4.1 Introduction

To sustainably support the performance of professionals, deeper insight into professional development is needed. In addition to knowledge and performance, professionalism of working people requires them to engage in ongoing professional development (Messmann et al., 2010). Many studies have only focused on professional development as the participation in formal training (Ayeleke et al., 2019; Otoo, 2019). To address the realities of professionals' working life, professional development must be considered in a much broader way, including components that are part of workplace learning (Tynjälä, 2008) or informal learning (Kyndt et al., 2016). In doing so, different components of professional development must be distinguished. Firstly, it is important to address cognitive development processes, which can be determined by investigating the engagement in learning activities during work. Simons and Ruijters (2001, 2004) distinguished three ways in which professionals learn: (a) elaborating on their own competences, by learning from and in practice, (b) expanding theoretical knowledge and insights, by learning through research, and (c) externalizing, by contributing to the development of the profession, the team and/or organisation. Secondly, it is also essential to determine the outcomes of such processes, which can be investigated by measuring knowledge, behaviour or competences. Here, competences are defined as the integrated knowledge, skills, and attitude that a person has that are connected to a specific job (Mulder & Baumann, 2005).

Important is the context boundedness of all aspects, including the just mentioned competences, which is in empirical studies often not directly connected to actual behaviour (in a specific context). The same is true for the other components of professional development. In empirical research on professional development, studies have frequently used questionnaires making it challenging to collect data on the direct link between the characteristics of the context and the consequential professional development. The Vignette Technique (VT) offers opportunities to investigate this context boundedness, moreover with questionnaires.

Finch (1987, p. 105) introduced the VT as a research method and defined vignettes as "short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond". Sampson and Johannessen (2020, p. 57) defined vignettes in a more general way as a "written description of a (frequently fictitious) event which relates to the central topic of study". Vignettes are short descriptions of (hypothetical) persons or situations that contain information that the respondents need in order to make a judgement (Poulou, 2001). Vignettes are narrative, but they should also provide a mostly neutral description of a situation based on theory. The vignette should describe only one concrete situation or phenomenon. Vignettes can be presented as written texts, but also as photographs or films, news articles and historical sources (Törrönen, 2018). The kind of information gathered with a vignette depends on the instructions and the answering mode.

By describing a situation in a specific context, vignettes encourage respondents to find solutions and practical applications, for instance in solving the described problem (Finch, 1987; Jeffries & Maeder, 2004; Leicher & Mulder, 2018; Schoenberg & Ravdal, 2000). Sampson and Johannessen (2020) noted that vignettes, especially when they describe a "real-life" event (p. 56), can give researchers the possibility to "turn on the tap" (p. 70) and collect a significant amount of valuable data. Furthermore, vignettes can be used "to speak about issues that are sensitive, concealed, disgusting or taken for granted" (Törrönen, 2018, p. 282).

The VT has several advantages. First, a link between the context and the consequential professional development can be built in the vignette. The vignettes can describe the link between for instance the characteristics of a work task, or a broader context and then invite the respondent to mention his/her consequential professional development. Second, the direct link between triggers, such as feedback, and the professional development they cause can be part of the vignette, and can therefore be measured. Third, by using the VT it is possible to investigate the intended behaviour. Finally, all the components of professional development can be measured, including the learning behaviour in the process of development (in terms of for instance engagement in learning activities), the outcomes of development (such as knowledge) and the changes that occur over time.

To gain deeper insight into how the VT can be applied in empirical studies on professional development that use questionnaires, it is important to ask: How can the VT be used to gain insight into the relationships between the characteristics of the context and the components of professional development that emanate from that?

Thus, the goal of this chapter is to show how the VT can be used to measure the components of professional development and what characteristics of vignettes are required, as well as the instructions and the answering modes, and how the vignettes have to be developed. In this chapter, we explain VT by discussing the characteristics of vignettes and the use of the vignettes for different objectives and by addressing what kind of data can be obtained in using them. Different examples of vignettes that are suitable in relation to the objectives and the fitting instructions and answering modes used in various empirical studies are presented, followed by information on how the vignettes need to be developed.

4.2 The Vignette Technique

As previously mentioned, the VT enables researchers to focus on professional development as a process of development and on the outcomes of the process with regard to a specific setting or situation (Soydan & Stål, 1994). The VT uses a vignette, a kind of instruction and an answering mode. This section describes the characteristics of the vignettes, the kind of instructions and the answering mode. From this information the requirements for the development of these components of the VT can be derived.

4.2.1 Characteristics of Vignettes

There are different relevant characteristics of vignettes that have to be carefully considered when developing them: (1) the content of the vignette, (2) the style of the vignette and (3) the quality of the vignette.

1. In relation to the content**,** vignettes should describe a realistic situation (Hughes & Huby, 2002). Therefore, several criteria should be considered: the content should consist of authentic, realistic problems that need to be solved (Finch, 1987), that are open (incomplete) and that allow for multiple solutions and are intended to encourage independent thinking and unique responses (Jeffries & Maeder, 2004). The content must be internally consistent (Finch, 1987) and must consist of a chain of context-specific events to make it possible to infer causal influences (Miles, 1990). To be authentic, vignettes need to describe domain-specific practices and the context of a situation (Barab et al., 2000). Regarding the description of a realistic task, vignettes need to determine the complexity of the situation by describing the tasks and situations using various aspects and components that are connected (Finch, 1987). Participants should have the chance to find a solution for the described situation or develop strategies to handle it. Moreover, the vignette should be formulated in such a way that individual answers can be given, multiple ways to solve the issue can be identified or developed and a variety of methods can be used to collect rich and valid data. The content of the vignette should fit the objectives of the study, as well as the target group (e.g. the language used), in order to acquire valid data.

   Jenkins et al. (2021, p. 3) noted that when using vignettes as a research method it is necessary to ensure that they "provide an internally valid, authentic, plausible or realistic reflection of the social situation that are likely to occur in real-world settings". Vignettes have to meet different requirements in order to be able to collect comparable data. An advantage of this method is that it provides the possibility of comparing groups (Renold, 2002). Thus, because all participants respond to the same vignette, the answers are comparable, which makes the data useful, for instance, for testing hypotheses. To obtain answers from participants that can be compared, the description of the situation should be generic. Jenkins et al. (2021) argued that when participants perceive a described situation as unrealistic, they are less likely to be motivated to engage in independent thinking and interpretation. Therefore, vignettes are often based on or describe lived situations (Jenkins et al., 2021).
2. Regarding the style of the vignette, it should be a short story or narrative (50–200 words; Jeffries & Maeder, 2004) that addresses the readers' ability and style (Hughes & Huby, 2004), so that the respondents understand the vignette (Finch, 1987). When designing the vignette, the developer should start with its function, the issue it addresses and the topic it presents (Richman & Mercer, 2002). The

vignette should also be based on literature (Wilson & While, 1998) and it should be developed in collaboration with the professionals associated with the work context (Miles, 1990). The description should be written in simple words, but, when necessary, it should use key terms of the domain (Hughes & Huby, 2002).
3. When using the VT, the quality of the vignette is pivotal. It should be validated by professionals by offering them opportunities to improve the vignette (Spalding & Phillips, 2007).

4.2.2 Instructions and Answering Modes

As short stories, vignettes should consist of a story, and be consistent with the instruction and the answering mode. The kind of information gathered with a vignette depends on the instruction. For instance, when the instruction is in the form of the question, such as "What would you do?", the respondent's behaviour would be investigated. When investigating attitudes, respondents could be asked "What do you think about that?". The instruction should fit the purposes of the study. When providing clear instructions for the vignette, which fits the objective of the study, and with an appropriate answering mode, it is possible to gather both quantitative and qualitative data. While other research techniques, such as surveys with closed ended questions, ask participants to estimate their behaviour in a more general way, the VT describes a concrete situation that participants refer to in their answers (Heldbjerg & van Liempd, 2018).

The answering mode should be based on a theoretical framework, ensuring a clear definition of a construct and systematic operationalisation. The choice of the answering mode must suit the objectives of the study; it can be in the form of open end questions, which can be used to collect qualitative data, and closed ended questions, which can be used to collect quantitative data. Closed answering modes, such as Likert scaled answering modes, lead to quantitative data, which make it easier to compare the answers of the different respondents.

When combining vignettes with open answering modes, the data analysis can be time-consuming. For instance, qualitative content analysis (Mayring, 2004) or analysis of similarities (for instance by using linguistic network analysis; Fürstenau & Trojahner, 2005) could be applied. Moreover, there are several possibilities for transforming qualitative data into quantitative data. With open answering modes, the answers have to be categorised after they are collected and counted to make them usable for quantitative research. Qualitative data can through appropriate analyses be transformed into quantitative data, which then also provides opportunities to compare the answers and test hypotheses. When data are quantitative, the quantitative analyses that are required to test the hypotheses and answer the research questions can be conducted.

4.3 Use of the Vignette Technique

Based on the information presented in Sect. 4.2, it is possible to determine what aspects must be considered when developing a vignette, including the appropriate type of instructions and answering mode. In this section, several examples of studies are described to illustrate how the VT can be used in empirical research to gain insight into the aforementioned components of professional development by employing a questionnaire.

Applying the VT has consequences regarding the choices that must be made concerning the design of the study to fully utilise the opportunities that this technique provides. Furthermore, to gain systematic insight into the opportunities the VT provides, attention must be paid to the instrument and the data analyses to enable the researcher to make the right decisions. What the appropriate decisions are depends on the objectives of the study and the need to obtain valid data to answer the research questions. In Table 4.1 the different parts of an empirical study are listed, with the components that need to be part of the study using VT and the characteristics of the

Table 4.1 List of components of empirical studies using the VT

Parts of the empirical study	Components of the study	Possible characteristics of the empirical study
Research design	Design Domain/ sample	Cross-sectional, longitudinal design Various characteristics of the respondents Sample size
	Use of instrument	Options: e.g., present one, randomly distribute, participants choose one, use different versions at different measurement moments
Instrument	Vignette	**Content requirements:** Authentic, realistic, allows multiple solutions, consistent, describes domain-specific practices Call-to-action, trigger
		Style requirements: Image, written text (50–200 words) or narrative Adjusted to the readers' ability and style, use key terms of the domain
		Quality requirements: Pre-study for development and validation of vignettes Use of domain specific expertise
	Instructions	Various question(s), request Measuring professional development or specific components thereof
	Answering mode	Open answering mode or Closed answering mode (e.g. Likert scale)
Data analysis	Data Analysis	Qualitative data, Quantitative data or Mixed data Qualitative data analysis (e.g. content analysis; quantify qualitative data), Mixed methods analysis, or Quantitative data analysis (statistical analysis)

components that represent the decisions the researcher need to make in the phase of the development of the study and during conducting the study.

A description of the example studies and the relevant characteristics that are listed in Table 4.1 will be provided in the next sections.

4.3.1 Example 1: Learning from Errors

The first example is a cross-sectional study in which vignettes are used to investigate learning from errors as a component of professional development with a questionnaire (Leicher et al., 2013). In this study, geriatric nurses were able to choose one of two vignettes describing a specific situation where an error occurred. The vignettes on errors were combined with closed ended questions with a Likert scale answering mode on learning activities that occurred after having made an error. Quantitative statistical analyses were applied to the collected data. Table 4.2 shows an example of a vignette, instructions and answering mode.

4.3.2 Example 2: Team Mental Models

In this study, the focus was on the nursing teams' knowledge as one of the outcomes of professional development (Leicher & Mulder, 2018; Anselmann et al., in progress). The focus of this cross-sectional study with a questionnaire in the domain of geriatric nursing was on the knowledge of the nurses on what is needed to accomplish the work tasks of nursing teams. In this study, the vignettes described the relevant work tasks of nurses, and the participants were able to choose one of two vignettes. They could answer an open end question in their own words and describe how they accomplish the work task. The goal of this study was to determine if a team member has shared knowledge about how to accomplished the work tasks.

Table 4.2 Example of a vignette to investigate 'Learning from errors'

Vignette	'Misjudging the risk of bed sores' (Leicher et al., 2013, p. 213) While providing basic care to a bedfast patient, you recognise reddened skin on his tail bone. Considering this to not be serious, you do not initiate additional measures of prophylaxis or treatment. Providing the treatment, the next time your colleague notices skin lesions, which are a clear sign of increased risk of bed sores. In light of this, the increased risk should have been recognised at the earlier evaluation
Instruction	Please imagine the described situation and answer to the following questions with regard to the described situation. What would you do after this error occurred?
Answering mode	Closed answering mode; scale on engagement in social learning activities (5 point Likert scale; Bauer & Mulder, 2011)

Table 4.3 Example of a vignette to investigate 'Team Mental Models'

Vignette	'Giving assistance to a resident to get dressed' (Leicher & Mulder, 2018, p. 16) After basic care in the morning, it is your task to give help to a resident at the age of 72 with getting dressed. The resident uses a walker and is restricted in her orientation regarding time and place. She is able to recognise the nurses and is aware of her own situation. Regarding the time structure and orientation in the retirement home she is restricted in her cognitive abilities. Because of her chronicle pulmonary disease and a heart failure she complains about laboured breathing under exertion. She does not need oxygen supply and is also able to do all functional self-care and maintenance tasks by her own under the guidance of an elder care nurse
Instruction	What work steps would you take to handle the described situation? What work steps are most important?
Answering mode	Open answering mode

Qualitative content analysis and linguistic network analysis (cf. Fürstenau & Trojahner, 2005) were applied to analyse the data. By using qualitative content analysis, it was possible to identify the knowledge components of the individual team members. We used linguistic network analysis to determine to what extend the knowledge of a member of a team is similar to that of the other team members, and determine from that what the team member model is. In Table 4.3 the vignette used to investigate team mental models is described.

4.3.3 Example 3: Complexity of Work Tasks and Informal Learning Activities

In the research on the effects of characteristics of the work task, we chose to include the complexity of the task in different vignettes. In this cross-sectional study (Hirschmann & Mulder, 2018) in the IT-sector, about half of the respondents randomly received the questionnaire with one vignette and the rest of the respondents received the other version.

We made a direct connection between a characteristic of the context, in this case work tasks and more specifically their complexity, and the learning activities that emanate from them (Hirschmann & Mulder, 2018) by developing vignettes containing a more and a less complex realistic situation at work. Complexity of work tasks depend on the number of elements and their interrelations. Elements can be sub-tasks, aims, and possible solutions. Complexity is thus operationalised as the number of elements (knots) and the relations between these elements. The vignette with low task complexity had 7 knots and 6 relations between the constructs, whereas the vignette with high task complexity consisted of 12 knots and 16 relations. The Table 4.4 contains the vignette with high complexity.

The answers were given on the items representing learning activities. A framework of possible informal learning activities was used where cognitive and physical,

Table 4.4 Example of a vignette to investigate 'Complexity of Work Tasks and Informal Learning Activities'

Vignette	'Programming the web shop' (Hirschmann & Mulder, 2018, p. 52) In the following, an authentic, but hypothetical example of your daily work is presented The producer of tools "Toolmaster" would like to extend his enterprise. The product range comprises different tools of the renowned producer as well as possible equipment. You take employment to program the web shop. The following features should be implemented: • Presentation of the goods, with an appealing layout • Consumer basket • Suggestions of additional products • Payment, with the use of a third-party supplier (note: Necessary for accepting credit cards, etc.) • Database for addresses, also for international customers, with a plausibility check • Recognition of customers and a list of recommendations on previous purchases • Integration with stock system, including inventory audit and direct enter on stock • Possibility to contact
Instruction	Please, read the vignette thoroughly. Try to imagine the situation During accomplishing the work task, you detect a process that could not be integrated into the software concept. What would you do?
Answering mode	Closed answering mode; 6 point Likert scale answering mode

as well as individual and social learning activities are distinguished. This framework formed the basis of the items on informal learning activities that the respondents could select, such as analysing the situation, searching the internet, discussing with colleagues and reflecting together with colleagues on feedback. For data analyses various suitable quantitative analyses were conducted.

4.3.4 Example 4: Uncertainty and Proactive Work Behaviour

Another more complex phenomenon to investigate is the respondents' feelings about the work context in relation to the behaviour it causes. In Jacob et al. (2019), the focus was on the relationships between the perceived context, in particular on uncertainty, and its effects on components of proactive work behaviour, namely innovative work behaviour and job crafting. In this cross-sectional study with questionnaires, in the finance sector, the objective was to determine the behaviour caused by state, effect and response uncertainty (Millikin, 1987), which in this study, we investigated specifically in relation to work. In a pre-study, which consisted of document analysis and interviews, vignettes were developed containing a hypothetical authentic situation consisting of all three types of uncertainty. The participants needed to choose one out of three different vignettes on the basis of what topic fits

Table 4.5 Example of a vignette to investigate 'Uncertainty and Proactive Work Behaviour'

Vignette	'FinTech and its competition in the finance sector' (Jacob et al., 2019) Imagine you face the following work situation: The speed of technological development makes it is unclear what the next major IT breakthrough might be. Google, Facebook, apple and other global internet platforms are already investing in the FinTech sector (robo-advisory platforms, crowdfunding or credit and factoring). It is not clear how technological development will impact current ways of working. You and three other colleagues are given an assignment to analyse FinTech trends and to report on the findings. It is clear that these findings will have consequences on your organisation
Instruction	"Related to the work situation described above, please state how frequently you carry out the following activities within your work context What do you do in this situation? I. . ."
Answering mode	Closed answering mode, on a 6 point Likert scale from 1 = "never" to 6 = "always"

best, that is, which was most realistic in relation to their actual work. All the vignettes contained a clear call-to-action requiring the participants to answer items on innovative work behaviour and job crafting. In Table 4.5 the vignette, the instruction and the answering mode is described.

The vignette contains a clear call-to-action as well as all three types of uncertainty. The activities they could select are aspects of innovative work behaviour and job crafting. After providing the answers, the participants were asked to describe how they experienced this case, especially what kind of uncertainty they experienced in the vignette, and how strong that was perceived (the instruction was "in this situation, I am unsure about. . .", followed by items representing aspects of the three forms of uncertainty on which the answers needed to be selected at a 6 point Likert scale from 1 = 'never true for me' to 6 = 'completely true for me'). The collected quantitative data were analysed with appropriate statistical analyses that were also suitable to the objectives of the study.

As previously noted, this was a cross-sectional study. To obtain more insight into processes and changes, vignettes can also be used in longitudinal studies. For example in a longitudinal study to measure teachers' professionalism in terms of competences. In longitudinal studies, the demands to obtain valid data require answers that are not caused by recognition of the content of the vignette, which leads to the need to use different vignettes over time. Therefore, we developed numerous vignettes for every component of the professionalism of teachers in relation to the different roles of teachers (such as the role of developer) and in relation to the different aspects of their work (such as dealing with diversity) (Mulder et al., 2017). In all three measurement moments, another vignette measuring the same competences can be presented to every teacher participating in the study. This procedure of using different versions of vignettes makes it possible to independently measure the same construct at three different measurement moments.

To summarise, in this section we provided examples of the different vignettes that were used in various empirical studies in different domains. The examples illustrate the different characteristics of vignettes and their use in questionnaires. We provided examples showing how vignettes can be used in cross-sectional and longitudinal studies in order to measure the components of professional development, for example engagement in learning activities, or to measure the outcomes, such as knowledge or competences. Thus, we illustrated how vignettes can fit the objectives of studies and how they can be different in terms of the level of inquiry and analysis (individual or team) and in terms of the aspects of the context of interest (domain, job).

4.4 Development of Vignettes

To consider the components of the vignettes, the content, style and quality, and the choices that need to be made concerning these components as well as all other components of the study (see Table 4.1), there are certain requirements in relation to the development of the vignettes. Therefore, two rules for developing good vignettes were applied: we developed vignettes with taking all the components of the study into account and realising the three aspects of vignettes, and we developed them in collaboration with experts (Mulder, 2015). In all our studies in which vignettes were used, they were developed or validated in a pre-study. When designing vignettes, two different approaches were used. In some studies, experts provided content (written or orally) for the vignettes. For instance, interviews were conducted with experts or representatives of the target group to collect examples of "real-life" situations. Based on the descriptions, we developed first drafts of the vignettes. Then, the drafts were validated with a next round of interviews with experts, and if required, the drafts were revised based on the results. In other studies, we started to develop vignettes based on document analysis and we validated them with experts who gave oral or written feedback, which was used to optimize the vignettes.

The following example of the development process illustrates how high-quality vignettes can be realised. An adequate way to achieve this is to conduct a pre-study with semi-structured interviews with experts. The example concerns the pre-study on learning from errors where the goal was to identify error cases (Leicher et al., 2013). Vignettes were developed in an interview study with three nursing experts. The experts had to meet specific criteria to participate in the study: they needed to have extensive experience in their job and to have a supervisory function. In this way, we ensured that they could share insights on all parts of a nursing organisation. In the interviews, the experts were asked to describe typical error situations. They reported examples for errors in nursing. We analysed the data with qualitative content analysis using a deductive strategy to evaluate the types of errors. The error examples that were most frequently mentioned, and with which rich information about dealing with errors could be collected, were selected to develop the error vignettes. As a credibility check on how the error cases could potentially impact the results, we asked the participants to rate the authenticity and severity of the error cases and the degree to which they could identify with them. This was an indicator for the quality of the vignette.

4.5 Discussion and Conclusions

To gain deeper insight into professional development and to foster professional development in practice, good measurement is required. It is especially challenging to measure the direct relationships between the characteristics of the context (as for instance work tasks and interactions at work) and professional development, such as the process of development (in terms of actual, expected and/or intended behaviour) and the outcomes (in terms of behaviour, attitude, competences, performance and knowledge). This challenge is taken up in this chapter by exploring the opportunities afforded by using the VT (Finch, 1987) in questionnaires.

The VT consists of a vignette, instructions, and an answering mode. There are requirements that need to be considered when developing and using vignettes, the instructions and the answering mode. Firstly, these three components have to be consistent and coherent. Secondly, VT must fit the objectives of the study. Furthermore, the design of the study should fit. Depending on whether the study uses a cross-sectional or longitudinal design by using questionnaires, researchers must decide how many vignettes they need to develop and if the vignettes are similar or differ in respect to the relevant characteristics, such as the complexity of the work task. To achieve authentic and realistic vignettes, it is essential to obtain information from experts in the domain being studied. Therefore, the development of vignettes needs a pre-study, that can consist of interviews with the appropriate experts. In this development phase suitable decisions need to be made in relation to the content, the style and the quality of the vignettes. In relation to the research design, researchers have to decide between several options, such as letting participants choose vignettes, randomly distribute them in the sample, and use different vignettes at different measurement moments. We have provided different examples of such options. Furthermore, researchers must decide what kind of data they need to collect to be able to answer the research questions and meet the objectives of the study and what consequences that choice has on the instructions and the answering mode. Depending on the answering mode, qualitative and/or quantitative data can be collected. The kind of analysis depends on the kind of data that are collected. The usual analysis techniques that are needed to answer quantitative research questions can be applied to the quantitative data that are collected, or to qualitative data that is quantified. Moreover, with applying for instance content analysis, qualitative data can be used to answer the qualitative research questions.

As shown in the examples of use of VT in various empirical studies, this technique offers researchers opportunities to investigate the links between different aspects of the context and different components of professional development, such as the process of professional development itself (e.g., engagement in learning activities), an occurrence (such as errors or feedback) that can trigger learning activities, a characteristic of work (e.g., the complexity of the work task), the work climate or an even broader aspect, such as the perceived uncertainty. We also think the VT provides opportunities to delve even deeper into the processes of professional development, by for instance investigating the interaction between colleagues.

Interaction can for example be investigated by analysing the threads of the interaction including the reactions of different people within a team (Iiskala et al., 2015). For instance, this could be realised with the VT by including in the instruction the task to describe different steps used for communication in a team meeting.

This chapter provides insight into the opportunities the VT delivers, especially for the use in questionnaires. However, there are other instruments for questionnaires, such as the Critical Incident Technique (CIT; Flanagan, 1954) that may provide even better opportunities to investigate the direct relationships between the context and the aspects of professional development, since with the CIT actual situations experienced by the respondents are described, including their actual behaviour (Mulder, 2015). It can be assumed that the data collected with the VT are less valid, because they were collected in relation to a realistic, but hypothetical, situation. However, the advantage of using vignettes is that also sensitive topics (e.g. emotions) can be investigated, since vignettes provide a certain distance because they do not directly refer to a specific sensitive occurrence a respondent has experienced. Furthermore, data collection with the CIT is more time-consuming for the respondents, since they have to write down the details of the critical incident, which hinders the researcher's ability to acquire valid and rich data. Moreover, the VT data that is collected from the different respondents is comparable, because their answers refer to the same situation. This is more difficult to realise with the data collected by the CIT.

Last but not least, vignettes can be used as a tool to gain insight into (informal) learning at work, and they can be used as a diagnostic instrument in formal learning situations (training) as a tool to foster learning by using them as trigger learning processes (Leicher & Mulder, 2018) in a similar way as for instance problems or cases in problem based learning (Schmidt & Moust, 2000). This demonstrates that the VT has the potential to be more than just a technique that can be used to increase insight into professional development and its relationships with the characteristics of the context.

References

Anselmann, V., Fürstenau, & Mulder, R. H. (in progress). *Team mental models of nursing teams*.

Ayeleke, R. O., North, N. H., Dunham, A., & Wallis, K. A. (2019). Impact of training and professional development on health management and leadership competence: A mixed methods systematic review. *Journal of Health Organization and Management, 28*(33), 354–379. https://doi.org/10.1108/JHOM-11-2018-0338

Barab, S. A., Squire, K. D., & Dueber, W. (2000). A co-evolutionary model for supporting the emergence of authenticity. *Educational Technology Research and Development, 48*(2), 37–62. https://doi.org/10.1007/BF02313400

Bauer, J., & Mulder, R. H. (2011). In search of a good method for measuring learning from errors at work. In M. Van Woerkom & R. F. Poell (Eds.), *Workplace learning: Concepts, measurement, and application* (pp. 111–127). Routledge. https://doi.org/10.4324/9780203850084-15

Finch, J. (1987). The vignette technique in survey research. *Sociology, 21*(1), 105–114. https://doi.org/10.1177/0038038587021001008

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358. https://doi.org/10.1037/h0061470

Fürstenau, B., & Trojahner, I. (2005). Prototypische Netzwerke als Ergebnis struktureller Inhaltsanalysen [Prototypical networks as a result of structural content analysis]. In P. Gonon, R. Huisinga, F. Klauser, & R. Nickolaus (Eds.), *Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der DGfE. Kompetenz, Kognition und neue Konzepte der beruflichen Bildung* (pp. 191–202). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-86895-4_13.

Heldbjerg, G., & Van Liempd, D. (2018). Vignettes in critical theory investigations. In L. Young & P. V. Freytag (Eds.), *Collaborative research design: Working with business for meaningful findings* (pp. 313–340). Springer.

Hirschmann, K., & Mulder, R. H. (2018). Effects of complexity of work tasks on informal learning at work in the IT domain. In G. Messmann, M. Segers, & F. Dochy (Eds.), *Informal learning at work: Triggers, antecedents, and consequences* (pp. 40–62). Routledge Taylor & Francis Group. https://doi.org/10.4324/9781315441962-3

Hughes, R., & Huby, M. (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing, 37*(4), 382–386. https://doi.org/10.1046/j.1365-2648.2002.02100.x

Hughes, R., & Huby, M. (2004). The construction and interpretation of vignettes in social research. *Social Work & Social Science Review, 11*(1), 36–51. https://doi.org/10.1921/swssr.v11i1.428

Iiskala, T., Volet, S., Lehtinen, E., & Vauras, M. (2015). Socially shared metacognitive regulation in asynchronous CSCL in science: Functions, evolution and participation. *Frontline Learning Research, 3*(1), 78–111. https://doi.org/10.14786/flr.v3i1.159

Jacob, L., Chaudhari, V., Mulder, R. H., & Littlejohn, A. (2019, August). *The impact of uncertainty on proactive work behaviour: The role of motivation.* Paper presented at the 18th EARLI Conference, Aachen.

Jeffries, C., & Maeder, D. (2004). Using vignettes to build and assess teachers' understanding of instructional strategies. *The Professional Educator, 27*(1 & 2), 17–28.

Jenkins, N., Ritchie, L., & Quinn, S. (2021). From reflection to diffraction: Exploring the use of vignettes within post-humanist and multi-species research. *Qualitative Research, 21*(6), 975–989. https://doi.org/10.1177/1468794120920258

Kyndt, E., Vermeire, E., & Cabus, S. (2016). Informal workplace learning among nurses: Organisational learning conditions and personal characteristics that predict learning outcomes. *Journal of Workplace Learning, 28*(7), 435–450. https://doi.org/10.1108/JWL-06-2015-0052

Leicher, V., & Mulder, R. H. (2018). Development of vignettes for learning and professional development. *Gerontology & Geriatrics Education, 39*(4), 464–480. https://doi.org/10.1080/02701960.2016.1247065

Leicher, V., Mulder, R. H., & Bauer, J. (2013). Learning from errors at work: A replication study in elder care nursing. *Vocations and Learning, 6*(2), 207–220. https://doi.org/10.1007/s12186-012-9090-0

Mayring, P. (2004). Qualitative content analysis. *A Companion to Qualitative Research, 1*(2), 159–176.

Messmann, G., Mulder, R. H., & Gruber, H. (2010). Relations between vocational teachers' characteristics of professionalism and their innovative work behaviour. *Empirical Research in Vocational Education and Training, 2*(1), 21–40. https://doi.org/10.1007/BF03546487

Miles, M. B. (1990). New methods for qualitative data collection and analysis: Vignettes and pre-structured cases. *International Journal of Qualitative Studies in Education, 3*(1), 37–51. https://doi.org/10.1080/0951839900030104

Millikin, F. J. (1987). Three types of perceived uncertainty about the environment. State, effect and response uncertainty. *Academic of Management Review, 12*(1), 133–143.

Mulder, R. H. (2015). Using critical incidents and vignette technique in HRD research to investigate learning activities and behaviour at work. In M. N. K. Saunders & P. Tosey (Eds.), *Handbook of research methods on human resource development* (pp. 258–272). Edward Elgar Publishing. https://doi.org/10.4337/9781781009246

Mulder, R. H., & Baumann, S. (2005). Competencies of in-company trainers. In H. Gruber, C. Harteis, R. H. Mulder, & M. Rehrl (Eds.), *Bridging individual, organisational, and cultural aspects of professional learning* (pp. 105–109). Roderer.

Mulder, R. H., Sauer, S., & Kempka, F. (2017). FALKO-PA: Ein Instrument aus flexibel einsetzbaren Vignetten zur Erfassung pädagogischer Kompetenzen. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff, & R. H. Mulder (Eds.), *Falko: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (pp. 337–380). Waxmann Verlag.

Otoo, F. N. K. (2019). Human resource development (HRD) practices and banking industry effectiveness: The mediating role of employee competencies. *European Journal of Training and Development, 43*(1–2), 188–210. https://doi.org/10.1108/EJTD-07-2018-0068

Poulou, M. (2001). The role of vignettes in the research of emotional and behavioural difficulties. *Emotional and Behavioural Difficulties, 6*(1), 50–62. https://doi.org/10.1080/13632750100507655

Renold, E. (2002). Using vignettes in qualitative research. *Building Research Capacity, 3*, 3–5.

Richman, J., & Mercer, D. (2002). The vignette revisited: Evil and the forensic nurse. *Nurse Researcher, 9*(4), 70–82. https://doi.org/10.7748/nr2002.07.9.4.70.c6199

Sampson, H., & Johannessen, I. A. (2020). Turning on the tap: The benefits of using 'real-life' vignettes in qualitative research interviews. *Qualitative Research, 20*(1), 56–72. https://doi.org/10.1177/1468794118816618

Schmidt, H. G., & Moust, J. H. (2000). Factors affecting small-group tutorial learning: A review of research. In D. H. Evensen & C. E. Hmelo-Silver (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 19–52). Routledge.

Schoenberg, N., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology, 3*(1), 63–74. https://doi.org/10.1080/136455700294932

Simons, P. R. J., & Ruijters, M. C. (2001). Work-related learning: Elaborate, extend, and externalise. In L. Nieuwenhuis & W.-J. Nijhof (Eds.), *The dynamics of VET and HRD systems* (pp. 101–114). Twente University Press (TUP).

Simons, P., & Ruijters, M. C. (2004). Learning professionals: Towards an integrated model. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (Eds.), *Innovation and change in professional education: Vol. 2. Professional learning: Gaps and transitions on the way from novice to expert* (Vol. 2, pp. 207–229). Springer. https://doi.org/10.1007/1-4020-2094-5_11

Soydan, H., & Stål, R. (1994). How to use the vignette technique in cross-cultural social work research. *Scandinavian Journal of Social Welfare, 3*(2), 75–80.

Spalding, N. J., & Phillips, T. (2007). Exploring the use of vignettes: From validity to trustworthiness. *Qualitative Health Research, 17*(7), 945–962. https://doi.org/10.1177/1049732307306187

Törrönen, J. (2018). Using vignettes in qualitative interviews as clues, microcosms or provokers. *Qualitative Research Journal, 18*(3), 276–286. https://doi.org/10.1111/j.1468-2397.1994.tb00060.x

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154.

Wilson, J., & While, A. E. (1998). Methodological issues surrounding the use of vignettes in qualitative research. *Journal of Interprofessional Care, 12*(1), 79–86. https://doi.org/10.3109/13561829809014090

Veronika Anselmann (PhD) is an Assistant Professor in Nursing Education at the Institute of Nursing Science at the University of Education Schwäbisch Gmünd, Germany. In her research she is interested in informal and formal learning of nurses, learning in teams and nurses´ professional identity.

Regina H. Mulder (PhD) is Full Professor of Pedagogy/Educational Science at the Faculty of Human Sciences at the University of Regensburg, Germany. Her research interests include a variety of topics in vocational education and training (VET) and on learning in organizations, such as relations between technology and professional development, leadership, evaluation of training, innovative work behaviour, informal learning at work, feedback, and team learning.

Chapter 5
Integrating Self-Reports and Electrodermal Activity (EDA) Measurement in Studying Emotions in Professional Learning

Susanna Paloniemi, Markku Penttonen, Anneli Eteläpelto, Päivi Hökkä, and Katja Vähäsantanen

Abstract Studies on emotions in learning have been mostly conducted through self-reporting questionnaires and interviews conducted after the learning situation, which seldom focus on professional workplace learning. Meanwhile, workplace and professional learning research has been introduced with emerging methods to capture the learning processes at multiple levels. Previous research on emotions has suggested that self-reports should be supplemented with psychophysiological data. Challenged by this, this chapter aims to present and discuss the integration of self-reported and psychophysiological data in studying emotions and professional learning. The empirical research data are obtained from five Finnish university teachers participating in a laboratory research comprising self-reported data (through the Emotion Circle application and a Stimulated Recall Interview) and psychophysiological data collected from the subjects' electrodermal activity. The process data on the quality, strength and duration of emotions are synchronised with those on psychophysiological, individual-level peaks of emotional arousal. The findings are discussed in terms of the possibilities and limitations of using complementary data in researching emotions in professional learning.

S. Paloniemi (✉) · A. Eteläpelto
Department of Education, University of Jyväskylä, Jyväskylä, Finland
e-mail: susanna.paloniemi@jyu.fi; anneli.etelapelto@jyu.fi

M. Penttonen
Department of Psychology, University of Jyväskylä, Jyväskylä, Finland
e-mail: markku.penttonen@jyu.fi

P. Hökkä
Faculty of Education and Culture, University of Tampere, Tampere, Finland
e-mail: paivi.hokka@tuni.fi

K. Vähäsantanen
Finnish Institute for Educational Research, Department of Education, University of Jyväskylä, Jyväskylä, Finland
e-mail: katja.vahasantanen@jyu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_5

Keywords Emotions · Professional learning · Self-reports · Multimodal methods · Electrodermal activity (EDA)

5.1 Introduction

Amid the rapidly changing and digitalised working life, continuous professional learning is necessary for individuals as well as for work communities and organisations (Goller & Paloniemi, 2017; Harteis, 2018). Accordingly, professional learning has been extensively explored, theorised, and supported in workplace contexts (Billett et al., 2014; Tynjälä, 2013). In these contexts, professional learning is understood in various ways, such as the development of expertise and competence, organisational renewal and the (trans)formation of work practices. Recently, professional learning has also been conceptualised as the (re-)negotiation of identity (Eteläpelto et al., 2014; Vähäsantanen et al., 2017a). Such professional identity learning particularly occurs through collaboration within social interactions (e.g. Lave & Wenger, 1991), as well as experimentation and reflection regarding one's professional mission and practices (e.g. Vähäsantanen et al., 2017b; Zwart et al., 2015). Recent research has further shown that emotions play a central role particularly in group-based identity learning (Vähäsantanen et al., 2020). Despite convincing findings on professional learning, there is a lack of research on the role of emotions in professional learning. So far, only a few studies have explored emotions and professional learning among professionals (e.g. Rausch et al., 2017).

Topically, emotions are understood as situational and dynamic responses to a personally meaningful situation, event, or person (Gooty et al., 2010). They are expected to be elicited by a specific target or cause and include physiological reactions and action sequences (Barsade & Gibson, 2007). There has long been discussions on how emotions trigger or inhibit human behaviour (Hommel et al., 2017) and how their emergence is closely related to the autonomic nervous system (ANS) activity (Levenson, 2014; Mauss et al., 2005). Accordingly, emotions are understood as multicomponential phenomena, including experiential, psychophysiological and behavioural components (Gendolla, 2017; Kreibig, 2010; Mauss & Robinson, 2009).

However, the findings obtained so far on emotions in professional learning settings are mostly based on questionnaires or interview data collected after the learning situations. Thus, such self-reports have neither captured the multicomponential nature of emotions nor have they been gained to capture the process data of emotions in learning. Consequently, it has been suggested that self-reports should be combined with other methods that measure the psychophysiological and behavioural components of emotions (Azevedo et al., 2016). However, in such multimethod measurements, one must address the questions of complementarity, interchangeability, validity and reliability (Eteläpelto et al., 2018). In addition, there are challenges regarding the differing time windows of the various modalities. The questions then arise as to how to overcome the methodological challenges of multicomponential measurement and how to apply methodologically advanced tools in rigorous ways to investigate emotions in professional learning.

In this chapter, we focus on researching emotions in the professional learning process and the integration of self-reported and psychophysiological research data within the field of professional learning and development. We start by briefly introducing emotions and learning, and then applying the self-reports and psychophysiological measurement of electrodermal activity (EDA) in emotion research. The empirical data utilised and illustrated later in this chapter are derived from a study exploring emotions in a professional learning process by integrating two types of self-reported data Emotion Circle (EC) and Stimulated Recall Interview (SRI) and EDA measurement data. In our study, professional learning is mainly understood as professional identity learning (Eteläpelto et al., 2014). Such learning covers processes in which professionals become aware of, reflect on and process their professional commitments, values and careers, as well as their strengths and weaknesses at work (Vähäsantanen et al., 2017b). It also encompasses the notion of an individual as an active and reflective participant—that is, someone who is responsible for learning and constructing change at a personal level within a given context. The integration of self-reported and EDA measurement data is illustrated through a closer look at the empirical data collected in a within-subject design.

5.2 Emotions Related to Learning

In terms of student learning, research has shown the central role of emotions in motivational processes, self-efficacy and active engagement, each of which plays a salient role in productive learning (Pekrun et al., 2006). Recently, there has been a growing interest in studying collaborative learning processes, with a focus on emotional or affective aspects alongside cognitive and motivational ones (Noroozi et al., 2020). Although research investigating the role of emotions in collaborative learning (Pijeira-Díaz, 2019; Pietarinen, 2021), self-regulated learning (Dindar et al., 2020; Järvenoja et al., 2018, 2020) and learning outcomes (Parong & Mayer, 2021) has increased, it has solely been implemented in educational contexts.

Yet, the research on emotions and professional learning or learning at work is still largely missing. Only a few studies have addressed the interplay of emotions and learning at work. A recent literature review on emotions and learning at work (Hökkä et al., 2020) has utilised the existing empirical evidence on the topic and showed that the relation between emotions and learning is understood in various ways within the field. Within the studies reviewed, emotions are mainly defined as emotional experiences and responses, and learning at work mainly refers to learning through participatory practices. Most of these studies have focused on the active role of emotions in supporting and/or hindering learning at work, where some have a viewpoint of emotions at work being influenced by learning. As only a few studies have investigated the interplay of emotions and learning at work, the role of emotions in professional learning and learning at work needs to be better understood (see also Benozzo & Colley, 2012).

So far, there exists some evidence on the role of emotions in professional identity learning. Meijers (2002) emphasised that a safe environment is essential for teachers' identity learning. This is not an environment that protects individuals from pain and uncertainty, as these emotions are also important starting points for the whole identity process. Rather, a safe environment allows an individual to handle and process such emotions in productive and creative ways and share them so that one's burden becomes lighter. Similarly, Winkler (2018) showed the complex role of emotions in professional identity work. First, emotions (e.g. self-doubt, fear, confusion, and dissatisfaction) can work as triggers for identity work. Second, identity work can be seen as an emotional endeavour that can include many emotions, such as fear, anxiety, unhappiness, uncertainty, shame and frustration. Third, emotions can be seen as outcomes of identity work; that is, successful and unsuccessful identity work can bring about emotions such as happiness, relief, frustration, shame and worry. Vähäsantanen et al. (2020) also showed that professional identity learning is a rich emotional endeavour during which emotions both support (e.g. inspiration) and hinder (e.g. fear) identity learning processes, and that emotions (e.g. compassion and shame) emerge as outcomes of the identity learning process.

In the research conducted on adult learning, positive emotions have been found to broaden the scope of perception, whereas negative emotions of anxiety and fear have been found to be related to narrowing of the perception and curiosity necessary for active and agentic learning (Fredrickson & Branigan, 2002; Perry, 2006; Storbeck & Maswood, 2015; Sung & Yih, 2015; Vähäsantanen et al., 2017b). In team-based learning, social and self-conscious emotions, such as compassion, love, shame, anxiety and anger, have been found to influence how team members see each other and how they perceive the team's future (Homan et al., 2015). Similarly, emotions have been found to act as precursors for active teamwork and collaboration (Mäkikangas et al., 2017; Watzek et al., 2019). Negative emotions and emotion-focused coping have been found to be provoked by errors at work (Rausch et al., 2017). However, the relation between negative emotions and learning is not straightforward. There is evidence that negative emotions can also foster learning (Hökkä et al., 2020). Consequently, the complex relation between pleasant and unpleasant emotions and learning needs to be further explored (Hökkä et al., 2020; Pietarinen, 2021; Vähäsantanen et al., 2020).

Despite the growing evidence on the meaning of emotions in learning settings, there is still no shared theoretical understanding of emotions. The discrete approach to emotions refers to the universally shared basic emotions that correspond to specific facial expressions, such as anger or joy (Damasio, 1999; Ekman, 2016). The dimensional approach understands emotions through the pleasant–unpleasant and low–high intensity dimensions (e.g. Ekman, 2016; Harmon-Jones et al., 2016). Emotions are then characterised in terms of valence, which refers to the subjective feeling of pleasantness or unpleasantness, and arousal, which refers to the subjective state of feeling activated or deactivated (Barrett & Russell, 1999). Mauss and Robinson (2009) concluded that the dimensional perspective has gained more support. In addition, emotions can be understood as culturally coded social entities, such as safety, pride or embarrassment. In this way of thinking, emotions are seen as mixed and socially produced categories that have a weight of tradition and the everyday experience behind them (Russell, 2003). Despite the variation in ways of understanding emotions, scholars have broadly agreed that emotions can be characterised as situational and intense, involving certain physiological responses to an event, entity, or person (Gooty et al., 2010; Mauss & Robinson, 2009; Zelenski & Larsen, 2000).

5.3 EDA Measurements in Emotion and Learning Research

An agreement on the multicomponential nature of emotions implies a multimethod measurement of emotions (e.g. Eteläpelto et al., 2018). Recently, there has been a growing interest in exploring multimodal data in learning research. So far, however, only a few empirical studies have focused on emotions and professional learning that utilise multimodal data. To date, self-reports (i.e. questionnaires and interviews) have been the most widely used methods in researching emotions, particularly those in learning settings (Noroozi et al., 2020). As an economic data collection device, there are many advantages of utilising self-reports: they enable a nuanced description of personally meaningful emotions, producing differentiated and detailed descriptions of emotions, thoughts and bodily sensations (Pekrun, 2016) in certain contexts and situations. However, self-reports are limited to conscious emotional responses and are most often collected retrospectively. In addition, there are both individual and cultural differences in the ways emotions are expressed, or even in the awareness of one's emotions (Azevedo et al., 2016). This implies that self-reporting methods should be supplemented with other types of data that capture, for example, the physiological component of emotions.

Another aspect that is important in considering self-report methods is that the data are usually collected retrospectively after the learning situation. This is often the case if, for example, interviews or questionnaires are used. Due to the time and nature of the learning situation, it is seldom the case that self-reports on emotions can easily be done during the learning situation. In the last few years, applications for online self-reporting on emotions have been introduced and experimented (Eteläpelto et al., 2018; Lehtonen et al., 2020). The study described in this chapter aims to develop an online assessment EC tool of emotions, especially for professional learning situations (see also Seifried & Rausch, 2022, Chap. 2, and Rausch et al., 2022, 2015). The first version described here paved the way for the development of the mobile online version of the EC (www.emotioncircle.fi).

It has long been recognised that the autonomic nervous system (ANS) plays a central role in emotions. As Levenson (2014) noted, 'when it comes to emotions, all roads lead to ANS'. ANS functions through two opposite but interacting regulation systems: the sympathetic and the parasympathetic nervous system. For example, when experiencing fear or anxiety, the sympathetic nervous system produces the 'fight or flight' response, whereas the parasympathetic nervous system is responsible for calming our body and mind ('rest and digest'). From the multicomponential perspective of researching emotions, the main value of ANS measurement is that it creates visible reactions and experiences that are neither spoken nor observable (Karvonen, 2017). Although there is an agreement on the multicomponential nature of emotions, there is no agreement on the relations among the components of subjective experience, ANS and behavioural changes. These relations are often addressed in terms of coherence, referring to the coordination, or association, of a person's experiential, behavioural and physiological responses as an emotion unfolds over time (Mauss et al., 2005).

Several (psycho)physiological indicators have been used in research in various disciplines, including heart rate (HR), heart rate variability (HRV), cortisol measurements and electroencephalography (EEG) (see also Kärner & Sembill, 2022, and Silvennoinen et al., 2022). In this chapter, we focus on one of the most often used psychophysiological measurements in emotion research: EDA. With a history of more than a century, EDA is a term used for variations in electrical conductance of the skin, including phasic changes that result from sympathetic activity (Boucsein, 2012; Kreibig, 2010). The notion of sympathetic arousal is present whenever EDA is used as a psychophysiological indicator. The skin is innervated by the sympathetic nervous system, which is the controller of the so-called 'fight-or-flight' response. EDA represents a valid indicator of emotional arousal (Kreibig, 2010). It refers to changes in skin conductance (SC) and can be measured with non-intrusive techniques providing information on emotional arousal, increased cognitive workload and task engagement. The higher the conductance level rises, the more elevated an individual's emotional arousal becomes. Compared to, for example, cardiovascular measurements, EDA is the only one considered to be a pure marker of sympathetic activation (Pijeira-Díaz, 2019). The eccrine sweat glands, uniquely innervated by the sympathetic nervous system, are the primary determinants of EDA (Boucsein, 2012). Their highest density occurs in the palmar and plantar (or sole) regions. Thus, EDA can be easily measured with a properly spaced pair of electrodes in the palm and sole areas. However, as the placing of electrodes on such regions interferes with daily activities, alternative locations, such as the wrist, are also used. The most recently advanced wearable devices for EDA measurement are wristbands (Dindar et al., 2020; Järvenoja et al., 2018; Pijeira-Díaz, 2019) and smart rings (Lehtonen et al., 2020). Furthermore, measuring arousal is theoretically linked to the circumplex model (Russell, 2003) of researching emotions.

Although there are challenges with data collection and analysis, technological solutions have already provided advanced methods for both EDA data collection and analysis. So far, they have not yet been widely used in learning research, combining various research methods and data. In their recent review on using multimodal data in learning research, Noroozi et al. (2020) found that the most widely used method combinations were interviews, surveys and observations. In contrast, the least-used methods were biometric and objective measures, with no studies utilising EDA. This is attributed to the finding that in learning research, cognitive and motivational aspects are overwhelmingly represented, not the emotional dimension. Furthermore, none of the studies included in the review were conducted in workplace learning or in vocational education contexts. Consequently, the triangulation of subjective (e.g. self-reports) and objective (e.g. arousal measurement) data is largely missing within the field of researching emotions in the learning of adults, especially in workplace and professional learning research.

So far, self-reported and EDA data have been used together in studies implemented in educational and schooling contexts (e.g. Dindar et al., 2020; Järvelä et al., 2021; Järvenoja et al., 2018; Villanueva et al., 2018), as well as virtual learning environments (e.g. Parong & Mayer, 2021). One of the few studies combining self-reports in the working life context was conducted by Lehtonen et al. (2020), who collected electrodermal activity data through smart rings and self-reported data on emotions and learning through the Learning Tracker application and interviews. The results showed that the participants reported more positive than negative emotions at work, the most frequent emotions being contentment and frustration. Concerning arousal, the level was higher during the morning hours of work and decreased during the afternoon hours. However, the study did not investigate the relations among arousal, emotions and learning. As the data were analysed independently, the question of integration and supplementary understanding remains open.

In the study described in this chapter, we aim to investigate the integration and complementarity of subjective experiences (self-reports) and psychophysiological (ANS) data. Accordingly, we understand emotions as multicomponential, dimensional and personally meaningful situational responses to an event or person/s that are elicited by a particular target or cause (see also Eteläpelto et al., 2018). The empirical study presented next illustrates the integration of self-reported interview data with EDA measurements in studying emotions in professional identity learning.

5.4 Integrating Self-Reports and EDA: An Empirical Study on Emotions in Professional Learning

The empirical study described here originates from a Finnish research project, *The Role of Emotions in Agentic Learning at Work*. The project aimed to develop innovative tools for the online assessment of emotions, together with utilising complementary research methodology within the field. In this chapter, we utilise a study conducted in the early phases of the project, when we started piloting the research methodology and the online emotion assessment tool. This was done by continuing the research collaboration with university teachers who had participated in our research project on professional identity and agency a few years earlier (Vähäsantanen et al., 2017b, 2020). As part of this collaboration, the teachers participated in an identity coaching programme comprising six workshops conducted in small groups. The workshops were video-recorded, and we could use this material in the project, now focusing on the emotional aspects of professional learning.

The present study has two aims. First, to examine the types of emotions that emerge while participants assess professional learning situations. The following research question is addressed: What types of emotions do the participants report while watching personally meaningful professional learning episodes of an identity coaching programme? The second aim is to integrate multiple methods and data in researching emotions connected to professional learning processes. In doing so, we focus on answering the following question: What types of information do self-reports and EDA provide for understanding emotions in professional learning?

5.4.1 Research Design and Data Collection

As suggested by the researchers investigating concurrent responses and addressing the coherence of multiple sub-systems of emotions (Kreibig, 2010; Levenson, 2014; Mauss & Robinson, 2009), a within-subject design was chosen for the study. The study reported here aimed to capture emotions simultaneously on two levels: (1) subjective experience (self-reports) and (2) ANS (psychophysiological measurement). Our purpose was to investigate concurrent measures applicable to changing emotions, especially those occurring within learning processes. In the best case, this would provide us insights into the subjective experience of emotions—observed concurrently with ANS functioning—as well as indications on how the measures can provide information on emotions in learning.

The data were collected in a laboratory setting where five university teachers watched selected video recording episodes (Fig. 5.1) from an identity coaching programme in which they had previously participated (Vähäsantanen et al., 2017b). The selection of video episodes was based on the participants' own assessments of the most important and meaningful learning situations during the identity coaching programme. All episodes were derived from the last workshop meeting, in which the participants worked via using an arts-based professional body method. This method combines individual and collective processes of identity work. As an individual task, the participants created their professional bodies using various

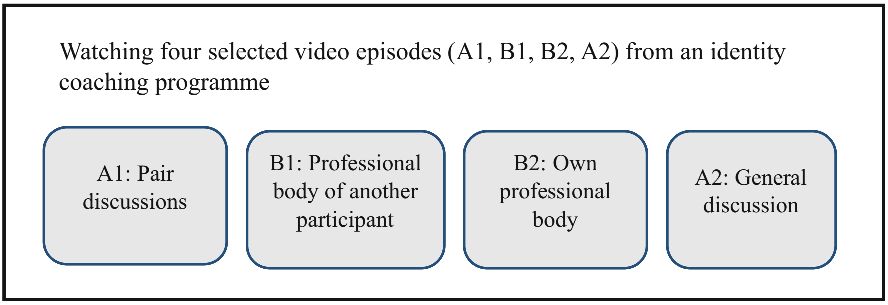

Fig. 5.1 Description of four selected video episodes

materials, such as clippings and drawings (Vähäsantanen et al., 2020). While personalising their bodies, the participants were guided to address different themes, such as professional history, core commitments and future dreams. In the last workshop of the programme, professional bodies served as material for collective processes. For example, each participant talked about one's own body, and the others responded to them by applying drama methods. The selected video episodes were organised and presented to the participants according to the principles of reverse counterbalancing design (ABBA; Goodwin, 2008). Accordingly, the first (A1) and last (A2) video episodes were selected in terms of representing neutral episodes. The second (B1) and third (B2) episodes represented the most meaningful learning situations for each participant.

The teachers participated in the study individually, and the procedure comprised consecutive Sessions 1 and 2. In Session 1, the teachers watched four video episodes, and subjective emotion reports were elicited using an online application EC (Eteläpelto et al., 2018), especially developed for the self-assessment of emotions concurrently with watching videos. The EC contained 12 written emotion words with a colourful graphic interface (Fig. 5.2) and was developed in line with the circumplex model of emotions (Russell, 2005) as well as the data-driven analysis

Fig. 5.2 Display of EC

of identity coaching participants' assessments of the emotions during the programme. In Session 1, the participants were asked to evaluate the nature and quality of their emotions through the EC while watching the video episodes. During this session, the participants freely used the EC according to their own choice by clicking on the emotion words. For the purposes of the study reported here, only the frequencies of emotions reported were included, while the level of a certain emotion was not accounted for.

Immediately after the first session, in Session 2, the participants' emotion assessments (as well as their connections to learning) were elaborated in an emotion-driven stimulated recall interview (SRI) (Kagan et al., 1963). Here, the video episodes, together with one's EC assessments, were shown to the participants. Furthermore, the participants were encouraged to describe, explain and elaborate the emotion assessments they had made through the EC and to freely share their thoughts and feelings with the researcher. The video was stopped while the participants were talking, to give them space and time to reflect on their emotion assessments and the learning situation at hand. The SRI data were obtained through video recording of the whole laboratory setting.

Throughout Sessions 1 and 2, psychophysiological data were collected. The ANS activity was measured using the QuickAmp amplifier and data acquisition system (Brain Products, Gilching, Germany; www.brainproducts.com). EDA was measured through two skin conductance (SC) electrodes, which were placed on the participants' non-dominant palms to avoid as much measurement error as possible (Karvonen, 2017).

5.4.2 Data Analysis

The analysis started by processing the different data sources separately and was conducted in line with the research questions. When investigating the types of emotions that are reported when watching and assessing professional learning situations, the self-reporting data collected through the EC were utilised. In doing so, we calculated the frequency of the emotion words that the participants reported when watching the video episodes (i.e. during Session 1).

The total number of emotions reported through the online EC was 357 (Table 5.1). Both pleasant and unpleasant emotions emerged in the participants' EC self-reports. The majority ($n = 294$) of reported emotions were pleasant, with joy being the most commonly reported emotion ($n = 78$), followed by surprise ($n = 50$) and excitement ($n = 47$). In addition, courage and compassion, as pleasant emotions, were assessed to characterise the professional identity learning occurring in the identity coaching programme. The most common unpleasant emotion reported was frustration ($n = 17$), followed by sorrow ($n = 15$) and anxiety ($n = 9$).

The participants differed from each other in terms of what emotions they reported, how many different emotions they reported and how many times they clicked on the emotion word while watching the video episodes. The individual variation in the

Table 5.1 Total number of emotions reported through the EC

Emotion words	Jane	Heather	Carol	Elsa	Lena	Total
Excitement	13	5	19	1	10	47
Surprise	7	8	20	15		50
Joy	11	20	29	10	8	78
Courage	8	3	15	2	5	33
Compassion	2	13	18	13	2	48
Safety	3	6	13	2	14	38
Sadness	1		6	8		15
Frustration	6	10	1			17
Shame	3	3		2		8
Fear	5			1		6
Anxiety	5	1		3		9
Irritation		6		2		8
Total	63	75	121	59	39	357

number of emotions reported ranged from 39 to 121. Out of the 12 available emotions in the EC, some participants reported a broader range of emotions than others (range 5–11). These differences were attributable to the personal differences in experiencing, acknowledging and expressing emotions. Alternatively, the episodes they watched might not have caused emotional responses. Clearly, the self-reported data through the EC alone were not sufficient for investigating emotions and professional learning.

Concerning the subject-specific EDA data, we focused on the skin conductance response (SCR) peaks during Session 1. These peaks are strongly associated with emotional responses to experimental conditions (Järvenoja et al., 2018; Manolov & Onghena, 2017). EDA was analysed with the Ledalab programme (version 3.4.6) written in Matlab (Benedeck & Kaernback, 2010; www.ledalab.de). Before the analysis, the sampling rate was reduced to 10 Hz, which was high enough to represent rapid changes in skin conductance (SC) related to SNS activation. The rapid components of SC were extracted as skin conductance responses (SCRs) and written in a Microsoft Excel file. The SCRs were normalised by computing the average and standard deviation of the sessions and calculating the z-scores. Values above 2.0 were considered statistically significant at the $p < .05$ level. Given that 5% of the values exhibited this property on two tailed normal distribution, they can be considered to represent statistically significant SNS activation.

The EDA levels and high peaks indicated emotional arousal for each participant. All five participating teachers had EDA peaks with an agreed-upon level of z-score > 2. In addition, the number of peaks varied across the participants, as did the z-score levels. However, these peaks alone gave us little information, especially in terms of emotions and learning being the ultimate focus of the study. Physiological reactions remain as reactions unless they are accompanied by individual emotional interpretations of the situation (Järvenoja et al., 2018). Thus, we matched the

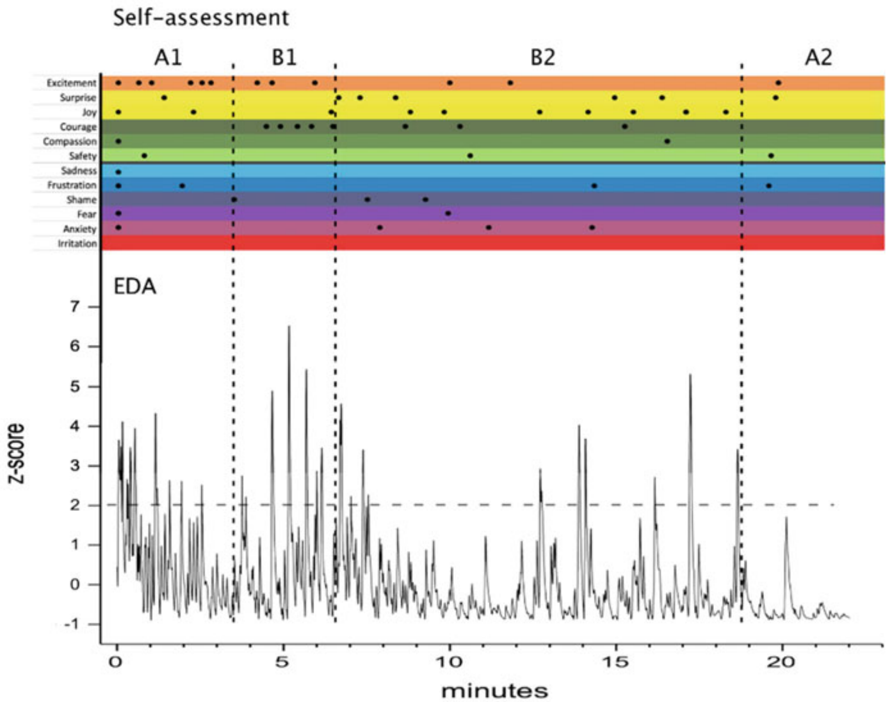

Fig. 5.3 Example of a visual combination of the self-report of emotions through EC and EDA data

two datasets of EC reports and EDA together for a visual analysis. In this way, we could identify the moments in which both emotional arousal and self-report on emotions were present.

A within-subject analysis applying visual analysis was used to scrutinise the possible connections between the EDA peaks and self-reported emotions through the EC. The basic assumption was that high-level EDA indicates emotional arousal, whereas low-level EDA relates to passive action (e.g. Kreibig, 2010). Thus, we looked for the possible connections between EDA level peaks (z-score $> 2 = p < .05$) and self-reported emotions. An example of a subject-specific combination of EDA at the bottom and the EC data on the top is presented in Fig. 5.3, where the horizontal line illustrates the emotional arousal level significant in the data, the colourful lines above represent the 12 emotions in EC and the dots represent the participant's assessments of these emotions.

5.4.3 Integrating EDA and Self-Reports from EC and SRI

To broaden our understanding and utilise multimodal data in examining emotions, we proceeded first with matching SRI self-report data with the EC and EDA data.

SRI data, together with the video transcriptions, were used to capture the subjective interpretations of the moments at hand and the related emotion assessments. The moments with high EDA peaks were examined more deeply. In doing so, the available self-reported data (i.e. transcripted video episodes, emotions reported through EC and transcripted SRI data) were examined simultaneously. We focused on the descriptions, reflections and assessments that the participant reported concerning the original learning situation (i.e. the video episodes from an identity coaching programme), the research situation (i.e. watching the video episodes) and in terms of the emotions reported through EC.

In focusing on the relatedness of the EDA peaks (i.e. objective data) and self-reported emotions (i.e. subjective data), we went through the data to see whether only one or both of them were present. Consequently, to understand connectedness, especially in terms of professional identity learning, we focused on closely analysing the moments in which both were present. The participants described, explained and reflected on their emotion assessments in the emotion-driven SRIs. The emotion-driven SRI data showed, in detail, how the reported emotions emerged and were related to professional learning. Joy and surprise emerged in connection to learning, such as the contents of the learning situation and discussing one's professional identity. In addition, joy was reported in the happy moments within the group. Excitement was reported by the participants in connection to one's own membership and participation in the group (collective sharing and creation) and to the actions of other members and the coach. Safety (as well as compassion) was reported in connection to other group members and to the importance of a psychologically safe learning environment in identity work. Concerning the unpleasant emotions reported, frustration and irritation (accompanied by shame) were connected to perceiving oneself professionally, as well as to one's own behaviour in the group. Furthermore, these unpleasant emotions emerged in connection to a slow pace of working, boring contents or behaviour of other group members. Eventually, the emotion-driven SRI data facilitated the addition of a new layer to the data analysis and increased our understanding of which and how emotions are related to the professional learning that occurs in the context of identity coaching programmes.

For a closer look at the integration of self-reports and EDA, we took Carol as an example. Figure 5.4 illustrates the EDA data at the bottom and the EC data at the top. The z-score value of 2 is marked with a horizontal line to illustrate the emotional arousal level significant in the data. Individual peaks larger than 2 are marked with red circles. The higher the EDA level, the higher the arousal. The above colourful lines represent the 12 emotions available in EC, and the dots represent single assessments of the emotions expressed by the participants (i.e. they clicked the emotion word in EC). In Carol's case, the total duration of watching the four video episodes (A1, B1, B2 and A2) was slightly less than 17 min. There are 10 circulated EDA peak moments in Fig. 5.3. Based on the self-report datasets of EC and SRI, these moments included the following emotion assessments through EC, and based on the SRI data, the emotions were related to different events and/or persons:

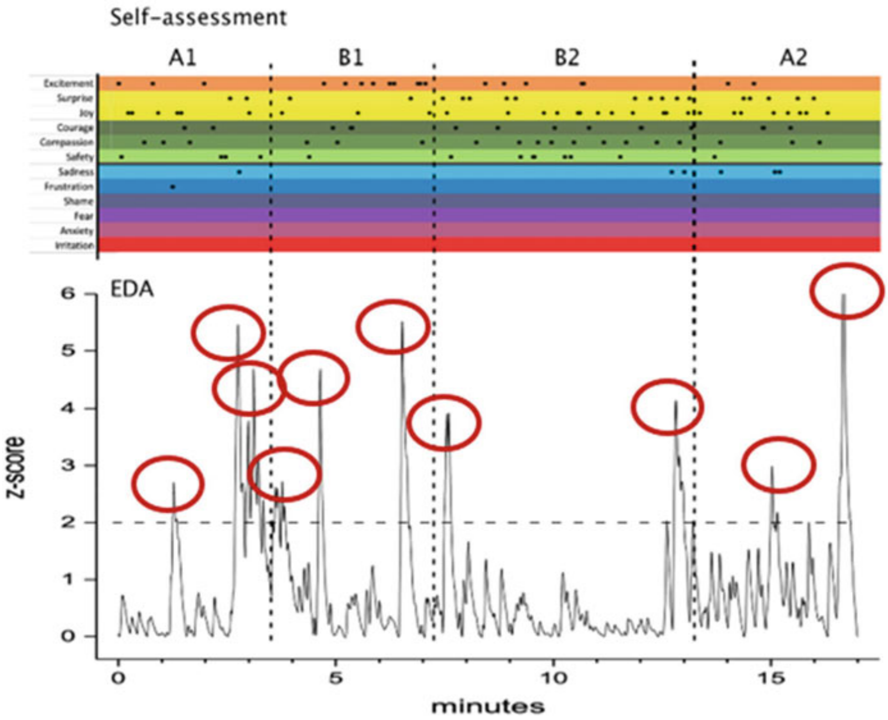

Fig. 5.4 Carol's self-reported emotions through EC and EDA peaks

Peak 1: Sadness towards one group member.
Peak 2: Sadness towards one group member.
Peak 3: No emotion reported—laughs aloud.
Peak 4: Surprised when seeing the video contents after a long time.
Peak 5: Excitement related to learning within the group.
Peak 6: Excitement related to professional identity work.
Peak 7: Joy, safety and compassion towards other group members.
Peak 8: Compassion towards one group member.
Peak 9: Sadness towards one group member.
Peak 10: Sadness towards one group member.

The above 10 moments of high EDA peaks reveal that the interpretation of EC reports and EDA is not a straightforward process. Not all peaks in the EDA data are related to self-reports on emotions. In Carol's case, peak 3 is an example in which there is no self-reported emotion at all, but the EDA level rises due to laughing aloud (which, as such, represents emotional behaviour). In addition, even though there is a connection between self-reported emotion and the EDA peak, there is not necessarily a relation with professional learning. For example, Carol reported sadness when watching the video episodes and said that it was related to seeing certain people and thinking of the unexpected changes in their personal lives after the coaching programme ended. The high EDA level in relation to sadness can also be attributed to the physical behaviour (crying and wiping the tears). In addition, Carol reported that sadness as an unpleasant emotion was related to the notion that the unique and meaningful moments with other participants and the identity coaching programme are already in the past and, thus, out of reach. This clearly illustrates that one emotion can emerge in different situations and be related to much different contents and events. This is also the case with excitement in Carol's case: excitement was found to have several positive associations with the professional learning process.

Carol's most often-reported emotions in addition to excitement were joy, surprise and compassion (Table 5.1). She explained in the SRI that joy and excitement were related to the ways of working in the identity coaching programme (e.g. drama methods) and to the learning processes and outcomes these methods produced. Furthermore, joy was related to the possibilities of sharing experiences with other group members and to be heard in shared encounters. In terms of social membership within the group, she reported compassion and safety as meaningful emotions supporting professional identity learning.

In general, the study findings revealed the emergence of both pleasant and unpleasant emotions in professional learning (see also Vähäsantanen et al., 2020). Emotions with positive valence activated and elicited professional identity learning. Furthermore, positive mixed social emotions were meaningful for the participants' learning in a group. Theoretical and educational significance emerged from a deeper understanding of complementarity and the different relations between the self-reported data on the quality of emotions and the psychophysiological data on the intensity of emotions. However, the questions of how to overcome the methodological challenges of multicomponential measurements and how to apply methodologically advanced tools in rigorous ways remain to unanswered (see also Al-Machot et al., 2019; Dindar et al., 2020; Järvenoja et al., 2018; Pijeira-Díaz, 2019). Thus, more research needs to be conducted to determine an optimal integration of various methods, as well as in terms of utilising multimethod approaches for understanding emotions in authentic professional learning contexts.

5.5 Discussion

Emotional aspects and the meaning of emotions in learning processes have gained increasing attention recently. Most of the studies have been conducted in educational and classroom contexts and have focused on the emotions of students. Thus far, the role of emotions in professional and workplace learning has remained largely underexplored. The vast development of research methods in emotion research has slightly entered the field of professional learning research as well. One developmental path is to use multimodal or multicomponential methods and data integration. Taken together, it is both the lack of knowledge and research on emotions in learning processes and the opportunities offered by advanced and innovative research methods that set new challenges and possibilities for researching professional and workplace learning processes in the future.

In this chapter, we discussed the role of emotions in professional identity learning. The most often reported emotions in such a learning context were found to be pleasant. Similar to a previous study (Winkler, 2018), in this study as well, excitement and surprise were found to be important activating emotions in learning. In addition, the role of a safe environment in professional identity learning was critical (Meijers, 2002). Our findings further indicated that social, mixed emotions (such as compassion, safety and courage) are the most important for active and open membership and participation within a group (see also Vähäsantanen et al., 2020). Overall, the professional identity learning process is not without emotions—in contrast, it is an emotional endeavour in which the phases of a learning process are intertwined with various emotions.

For the development of researching emotions and professional learning, we argue for a need to apply a multimethod approach. The study illustrated here provides support for integrating self-reporting and especially electrodermal activity data in researching emotions in professional learning. High EDA levels are related to both pleasant and unpleasant self-reported emotions in the context of professional identity learning. As such, arousal does not indicate a specific emotion but needs to be connected to a subjective description and report on the emotion in a specific learning context and situation. Thus, using self-reported and psychophysiological data within the subjects provides complementary information on the valence and arousal of emotions.

Our study confirmed the previous notion that electrodermal activity is a valid measurement of arousal and is applicable to researching emotions (e.g. Karvonen, 2017; Pijeira-Díaz, 2019). The data are fairly easy to collect and implement, and the interpretation of EDA levels alone is quite straightforward. However, EDA is also an indicator of physiological arousal, meaning that the subjects' movements should be minimised when measuring EDA. This was also evident in Carol's case, where physical movement in relation to joy (laughing) and sadness (crying) was present. Such behavioural-level reactions can result in EDA peaks and explain why, for example, sadness was related to high EDA levels in our study. Despite the vast development of mobile measurement technologies, the validity of EDA data (Milstein & Gordon, 2020) is still a challenge in implementing research under natural settings (e.g. in everyday learning situations at work). Furthermore, the skin conductance (SC) level is individual and dependent on many subjective (e.g. age, sex, medication and physical exercise) and contextual factors (e.g. temperature and humidity). In our study, even though conducted in a controlled laboratory setting, there were individual differences in the EDA levels between the participants. As we adopted a within-subject design in the study, this was not a problem. However, when implementing a between-subject study utilising EDA, these are important aspects to be considered (Karvonen, 2017). One of the challenges in investigating emotions and professional learning in the future is how to move on to research designs focusing on a between-subject study.

The study described in this chapter had a methodological aim of integrating self-reported data and psychophysiological measurement data in researching the role of emotions in professional learning situations. The implementation of the study and the findings supported the notion of utilising complementary data in researching emotions in professional learning processes. Based on the study results, we argue that self-reported data on the quality of emotions and psychophysiological data on emotional arousal provide data for a more comprehensive understanding of the role of emotions in professional learning processes. Furthermore, the emotion circle (EC) application developed and used in this study offered a promising online tool for concurrent self-reporting of emotions during the learning situation. The stimulated recall interview (SRI) method was crucial in determining the most meaningful learning moments and the specific quality of reported emotions during these moments. Self-reported data on the valence of emotions and psychophysiological data on emotional arousal can provide complementary data for understanding the role of emotions during professional learning processes. The practices, descriptions and reflections of learning situations and processes need to be carefully collected and documented. Without a clear relation to when and what is happening in terms of learning, the complementarity of the data remains scarce. In our study, it was the SRI interview that enabled this contextual and subjective interpretation.

However, the integration of multicomponential data is not without challenges. One basic challenge is the synchronisation of data. The research design with several data collection devices needs to be carefully set in terms of time frames. In our laboratory study, this was enabled and cross-checked through video recordings and sound marks at certain points. Thus, the self-reported and EDA data could be synchronised for the purposes of the analysis. Another aspect that needs special attention is the direction of the analysis. There are several choices to be made and questions to be answered concerning the analysis: Where should one start and with what data? Should the datasets be first analysed separately and then combined? Which of the datasets is the one directing the interpretation of the combined data? When using complementary multicomponential data, the relation between research questions and the data used becomes critical. Hence, the ultimate question is how the different data types complement each other in answering the research question set. In addition, as using and analysing multiple data types is time-consuming, there is a need to identify valid and economic ways for research and practical purposes. Consequently, more research on how to determine an optimal combination and integration of various data collecting and learning analytics methods is required.

5.5.1 Future Steps

The study described in this chapter was conducted in a laboratory setting in the early phases of the research project. Through this study, we have been able to further develop the EC application and narrow down our focus on methodological combinations in a way that meets the multicomponential understanding. In developing the EC application, we have asked about the most meaningful and relevant emotions for and in learning to be included in such a self-reporting tool. Together with the usability approach, having less cognitive stress when using the application has been an important aspect in developing the EC further (Eteläpelto et al., 2018). Consequently, an online application with an iconic design including the six most important emotions in terms of learning was launched in 2019. The current version of the EC application is more user-friendly and easier to use in different learning contexts. Our next aim is to focus further on the usability and content-specific development of the EC in specific work domains in everyday learning situations. This raises the question of how to include descriptive self-reported data on the application. For research purposes in particular, there is also a need to develop an application that can be more easily synchronised with other types of research data. Thus, the development of an online self-reporting application of emotions at work offers one layer of data for the integration of various data collection and analytics methods in researching emotions during professional learning processes.

New insights and elaborations for researching emotions and professional learning in real working-life situations are still missing. Novel technologies and applications will open up new research opportunities to be utilised in studying emotions in working-life learning settings. Various intelligent sociotechnical systems have rapidly emerged in emotion recognition, for example, in medical-patient monitoring, and emotion-aware intelligent systems (Kyamakya et al., 2021). The development of wearable sensors, typically used for sports and health activity tracking, has in their part paved the way for the emergence of an unobtrusive measurement of physiological responses within the research field (Pijeira-Díaz, 2019). With more sophisticated and advanced technical sensors and online tools, there are numerous opportunities to collect multimodal and componential data in everyday learning situations in working life. For studying emotions, EDA has been proved to be a valid measurement of ANS; thus, it offers a promising option in conducting multimodal research on emotions and learning in professional contexts. Furthermore, such data can also be used for development purposes (with feedback to the participants) at work. When working towards this goal, we require more research-based knowledge on the possibilities and best practices of integrating different data types in investigating emotions in workplace learning. Currently, we are facing a huge variety of opportunities that are waiting for exploration and implementation in real-life settings. This, we believe, will be one path towards a more coherent understanding of professional learning processes and the research methodology applied in learning sciences.

Acknowledgements Special thanks go to Hanna Liljapelto who designed and employed the visual datasets combining EC and EDA data. The authors are grateful to Tuomas Eteläpelto, Vesa Lappalainen, Lauri Viljanto and Petri Kinnunen for technical expertise and assistance. The research has been funded by the Academy of Finland [grant number 288925].

References

Al-Machot, F., Elmachot, A., Ali, M., Al-Machot, E., & Kyamakya, K. (2019). A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors. *Sensors, 19*(7), 1659. https://doi.org/10.3390/s19071659

Azevedo, R., Taub, M., Mudrick, N., Farnsworth, J., & Martin, S. A. (2016). Interdisciplinary research methods used to investigate emotions with advanced learning technologies. In M. Zembylas & P. Schutz (Eds.), *Methodological advances in research on emotion and education* (pp. 231–243). Springer.

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect. Controversies and emerging consensus. *Current Directions in Psychological Science, 8*, 10–14. https://doi.org/10.1111/1467-8721.00003

Barsade, D., & Gibson, S. G. (2007). Why does affect matter in organizations? *Academy of Management Perspectives, 21*, 36–59. https://doi.org/10.5465/amp.2007.24286163

Benedeck, M., & Kaernback, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*, 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Benozzo, A., & Colley, H. (2012). Emotion and learning in the workplace: Critical perspectives. *Journal of Workplace Learning, 24*(5), 304–316. https://doi.org/10.1108/13665621211239903

Billett, S., Harteis, C., & Gruber, H. (Eds.). (2014). *International handbook of research in professional and practice-based learning* (Springer international handbooks of education) (Vol. 1). Springer.

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer.

Damasio, A. R. (1999). *The feeling of happiness: Body and emotion in the making of consciousness*. Harcourt.

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies, 25*, 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science, 11*(1), 31–34. https://doi.org/10.1177/1745691615596992

Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2014). Identity and agency in professional learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (Vol. 2, pp. 645–672). Springer.

Eteläpelto, A., Kykyri, V.-L., Penttonen, M., Hökkä, P., Paloniemi, S., Vähäsantanen, K., Eteläpelto, T., & Lappalainen, V. (2018). A multi-componential methodology for exploring emotions in learning. *Frontline Learning Research, 6*(3), 6–36. https://doi.org/10.14786/flr.v6i3.379

Fredrickson, B. L., & Branigan, C. (2002). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 16*, 313–332. https://doi.org/10.1080/02699930441000238

Gendolla, G. H. E. (2017). Do emotions influence action? Of course, they are hypo-phenomena of motivation. *Emotion Review, 9*(4), 348–350. https://doi.org/10.1177/1754073916673211

Goller, M., & Paloniemi, S. (Eds.). (2017). *Agency at work: An agentic perspective on professional learning and development*. Springer.

Goodwin, C. J. (2008). *Research in psychology: Methods and design* (5th ed.). Wiley.

Gooty, J., Connelly, S., Griffith, J., & Gupta, A. (2010). Leadership, affect and emotions: A state of the science review. *The Leadership Quarterly, 21*, 979–1004. https://doi.org/10.1016/j.leaqua.2010.10.005

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PLoS One, 11*(8), 1–25. https://doi.org/10.1371/journal.pone.0159915

Harteis, C. (Ed.). (2018). *The impact of digitalization in the workplace*. Springer.

Hökkä, P., Vähäsantanen, K., & Paloniemi, S. (2020). Emotions in learning at work: A literature review. *Vocations and Learning, 13*(1), 1–25. https://doi.org/10.1007/s12186-019-09226-z

Homan, A. C., Van Kleef, G. A., & Sanchez-Burks, J. (2015). Team members' emotional displays as indicators of team functioning. *Cognition and Emotion, 30*(1), 134–149. https://doi.org/10.1080/02699931.2015.1039494

Hommel, B., Moors, A., Sander, D., & Deonna, J. (2017). Emotion meets action: Towards an integration of research and theory. *Emotion Review, 9*(4), 295–298. https://doi.org/10.1177/1754073916689379

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction, 72* , 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvenoja, H., Järvelä, S., Törmänen, T., Näykki, P., Malmberg, J., Kurki, K., Mykkänen, A., & Isohätälä, J. (2018). Capturing motivation and emotion regulation during a learning process. *Frontline Learning Research, 6*(3), 85–104.

Järvenoja, H., Malmberg, J., Törmänen, T., Mänty, K., Haataja, E., Ahola, S., & Järvelä, S. (2020). A collaborative learning design for promoting and analyzing adaptive motivation and emotion regulation in the science classroom. *Frontiers in Education, 5*, 111. https://doi.org/10.3389/feduc.2020.00111

Kagan, N., Krathwohl, D. R., & Miller, R. (1963). Stimulated recall in therapy using video tape: A case study. *Journal of Counseling Psychology, 10*, 237–243. https://doi.org/10.1037/h0045497

Kärner, T., & Sembill, D. (2022). Multimodal stress assessment in working and learning contexts: Using physiological, observational and experience-based data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 111–136). Springer.

Karvonen, A. (2017). *Sympathetic nervous system synchrony between participants of couple therapy* (Jyväskylä studies in education, psychology and social research 599). University of Jyväskylä. https://jyx.jyu.fi/bitstream/handle/123456789/55925/Karvonen_Anu_screen.pdf?sequence=1&isAllowed=y

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010Get

Kyamakya, K., Al-Machot, F., Haj Mosa, A., Bouchachia, H., Chedjou, J. C., & Bagula, A. (2021). Emotion and stress recognition related sensors and machine learning technologies. *Sensors, 21*, 2273. https://doi.org/10.3390/s21072273

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lehtonen, E., Puhakka, I., Pylväs, L., & Nokelainen, P. (2020). *Asiantuntijuuden kehittyminen, tunteet ja työpaikka oppimisympäristönä* [Emotions and development of expertise in the workplace] (The Finnish Work Environment Fund report no 118083). https://oma.tsr.fi/api/projects/20a33852-4761-4a08-9831-dece09ce66e8/attachment/baae2065-0f81-4a58-91d8-f19b9a1f332e [In Finnish].

Levenson, R. W. (2014). Autonomic nervous system and emotion. *Emotion Review, 6*(2), 100–112. https://doi.org/10.1177/1754073913512003

Mäkikangas, A., Bakker, A. B., & Schaufeli, W. B. (2017). Antecedents of daily team job crafting. *European Journal of Work and Organizational Psychology, 26*(3), 421–433. https://doi.org/10.1080/1359432X.2017.1289920

Manolov, R., & Onghena, P. (2017). Analyzing data from single-case alternating treatment designs. *Psychological Methods, 16*(March), 1–25. https://doi.org/10.1037/met0000133

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*, 209–237. https://doi.org/10.1080/02699930802204677

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behaviour and physiology. *Emotion, 5*, 175–190. https://doi.org/10.1037/1528-3542.5.2.175

Meijers, F. (2002). Career learning in a changing world: The role of emotions. *International Journal for the Advancement of Counselling, 24*(3), 149–167. https://doi.org/10.1023/A:1022970404517

Milstein, N., & Gordon, I. (2020). Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience, 14*, 148. https://doi.org/10.3389/fnbeh.2020.00148

Noroozi, O., Pijeira-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies, 25*, 5499–5547. https://doi.org/10.1007/s10639-020-10229-w

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*, 226–241.

Pekrun, R. (2016). Using self-report to assess emotions in education. In M. Zembylas & P. A. Schultz (Eds.), *Methodological advances in research on emotion and education*. Springer International Publishing.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*(3), 583–597. https://doi.org/10.1037/0022-0663.98.3.583

Perry, D. P. (2006). Fear and learning. Trauma-related factors in the adult education process. *New Directions for Adult and Continuing Education, 110*, 21–27. https://doi.org/10.1002/ace.215

Pietarinen, T. (2021). *Affect in collaborative and virtual inquiry learning* (Insights into small student groups and teachers in the classroom). Annales Universitatis Turkuensis. University of Turku. https://www.utupub.fi/bitstream/handle/10024/152012/Annales%20B%20540%20Pietarinen%20DISS.pdf?sequence=1&isAllowed=y

Pijeira-Díaz, H. J. (2019). *Electrodermal activity and sympathetic arousal during collaborative learning* (Acta Universitatis Ouluensis E184). University of Oulu. http://jultika.oulu.fi/files/isbn9789526222196.pdf

Rausch, A., Schley, T., & Warwas, J. (2015). Problem solving in everyday office work – A diary study on differences between experts and novices. *International Journal of Lifelong Education, 34*(4), 448–467. https://doi.org/10.1080/02601370.2015.1060023

Rausch, A., Seifried, J., & Harteis, C. (2017). Emotions, coping and learning in error situations in the workplace. *Journal of Workplace Learning, 29*(5), 374–393. https://doi.org/10.1108/JWL-01-2017-0004

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2005). Emotion in human consciousness is built on core affect. *Journal of Consciousness Studies, 12*(8–10), 26–42.

Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.

Silvennoinen, M., Parviainen, T., Malinen, A., Karjalainen, S., Manu, M., & Vesisenaho, M. (2022). Combining physiological and experiential measures to study adult learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 137–164). Springer.

Storbeck, J., & Maswood, R. (2015). Happiness increases verbal and spatial working memory capacity where sadness does not: Emotion, working memory and executive control. *Cognition and Emotion, 30*, 925–938. https://doi.org/10.1080/02699931.2015.1034091

Sung, B., & Yih, J. (2015). Does interest broaden or narrow attentional scope? *Cognition and Emotion, 30*(8), 1485–1494. https://doi.org/10.1080/02699931.2015.1071241

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*(1), 11–36. https://doi.org/10.1007/s12186-012-9091-z

Vähäsantanen, K., Paloniemi, S., Hökkä, P., & Eteläpelto, A. (2017a). Agentic perspective on fostering work-related learning. *Studies in Continuing Education, 39*(3), 251–267. http://www.tandfonline.com/doi/full/10.1080/0158037X.2017.1310097

Vähäsantanen, K., Hökkä, P., Paloniemi, S., Herranen, S., & Eteläpelto, A. (2017b). Professional learning and agency in an identity coaching programme. *Professional Development in Education, 43*(4), 514–536. https://doi.org/10.1080/19415257.2016.1231131

Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2020). University teachers' professional identity work and emotions in the context of an arts-based identity coaching program. In L. McKay, G. Barton, S. Garvis, & V. Sappa (Eds.), *Arts-based research, resilience and well-being across the lifespan* (pp. 233–256). Palgrave.

Villanueva, I., Campbell, B. D., Raikes, A. D., Jones, S. H., & Putney, L. G. (2018). A multimodal exploration of engineering students emotions and electrodermal activity in design activities. *Journal of Engineering Education, 107*(3), 414–441. https://doi.org/10.1002/jee.20225

Watzek, V., Anselmann, V., & Mulder, R. H. (2019). Team learning and emotions during teamwork: A qualitative study. *Research Papers in Education, 34*(6), 767–789. https://doi.org/10.1080/02671522.2019.1568525

Winkler, I. (2018). Identity work and emotions: A review. *International Journal of Management Reviews, 20*(1), 120–133. https://doi.org/10.1111/ijmr.12119

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality, 34*(2), 178–197. https://doi.org/10.1006/jrpe.1999.2275

Zwart, R. C., Korthagen, F. A. J., & Attema-Noordewiever, S. (2015). A strength-based approach to teacher development. *Professional Development in Education, 41*(3), 579–566. https://doi.org/10.1080/19415257.2014.919341

Susanna Paloniemi (PhD) is a senior lecturer in the Department of Education, University of Jyväskylä, Finland. Her current research and teaching interests include professional agency, workplace learning, and emotions in work organisations. She has utilised various research methods in her research, and thus holds an interest in methodological elaborations within the field. Her recent publications include articles in Frontline Learning Research, Studies in Continuing Education, and Professional Development in Education. She acts as an associate editor of Vocations and Learning journal.

Markku Penttonen (PhD) works as a senior researcher in the Department of Psychology, University of Jyväskylä, Finland. He is an Adjunct Professor in the field of cognitive neuroscience. His current research focuses on the role of neuronal network oscillations in learning and memory. He also studies bodily and autonomic nervous system synchronization between clients and therapists in couple therapy. His recent publications include articles in Frontiers in Neural Circuits, Journal of Neurophysiology, Psychophysiology, Frontiers in Psychology and Family Process and Psychotherapy.

Anneli Eteläpelto (PhD) is a Professor (emerita) in the Department of Education, University of Jyväskylä, Finland. She has widely published in academic collected editions and journals, such as Learning and Instruction, Educational Research Review, Scandinavian Journal of Education, Professional Development in Education, Journal of Teacher Education, Teaching and Teacher Education, Teaching and Teachers: Theory and Practice, Frontline Learning Research. She has led several research projects funded by the Academy of Finland, of which the most recent one explored the role of emotions in agentic learning at work (REAL). Prof. Eteläpelto has participated in international assessment panels, such as of University of Helsinki and Finnish institute for health and welfare.

Päivi Hökkä (PhD) is a professor of continuous learning (tenure track) at the department of Education and Culture at Tampere University, Finland. She has published widely in the field of workplace and adult learning. Currently she works as a co-PI of the international consortium Teaching that Matters for Migrant Students/TEAMS (2020-2023) and co-PI of Emotional Agency in Enhancing Strategic Implementation project (2021-2023).

Katja Vähäsantanen (PhD) is a senior researcher at the Finnish Institute for Educational Research, University of Jyväskylä, Finland. She is an Adjunct Professor in the field of working life research and development, University of Tampere, Finland. Her current research interests include lifelong guidance as well as professional identity, agency, learning, and emotions in work organisations. She is also interested in the potential of different methodological approaches, notably the narrative approach and its applicability to research on learning in working life. Her recent publications include articles in Vocations and Learning, Journal of Workplace Learning, Frontline Learning Research, Studies in Continuing Education, and Professional Development in Education.

Chapter 6
Multimodal Stress Assessment in Working and Learning Contexts Using Physiological, Observational and Experience-Based Data

Tobias Kärner and Detlef Sembill

Abstract This chapter focuses on multimodal stress assessment in working and learning contexts, with special consideration of cortisol measures and cardiovascular markers. First, stress-inducing and stress-buffering potentials of working and learning at the workplace are discussed. Second, selected research examples are reported to illustrate the application of psychophysiological methods and the integration of multimodal stress data. Three examples illustrate different application and measurement scenarios, measurement frequencies and data calculations, as well as different interrelations between physiological, psychological (experience-based self-assessments) and observation-based measurements. The chapter concludes by specifying an integrative framework for multimodal assessment and evaluation of learning and working environments, with recommendations for research practice.

Keywords Multimodal stress assessment · Heart rate variability · Cortisol · Experience sampling · Video-based analysis

Abbreviations

ANS	autonomic nervous system
AUC_G	area under the curve with respect to the ground
AUC_I	area under the curve with respect to increase
bpm	beats per minute

The authors would like to thank Susanna Paloniemi, Julia K. Weiß and Matthias W. Bottling for their valuable comments that helped to improve the chapter.

T. Kärner (✉)
Economic and Business Education, University of Hohenheim, Stuttgart, Germany
e-mail: tobias.kaerner@uni-hohenheim.de

D. Sembill
Economic and Business Education, University of Bamberg, Bamberg, Germany
e-mail: detlef.sembill@uni-bamberg.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_6

CG	control group
HF	high frequency range
HPA axis	hypothalamic-pituitary-adrenal axis
HR	heart rate
HRV	heart rate variability
Hz	hertz
LF	low frequency range
ms	milliseconds
nm	nanomolar
NN	normal-to-normal
PNS	parasympathetic nervous system
RMSSD	root mean square of the sum of the squares of differences between adjacent NN intervals
SNS	sympathetic nervous system
TG	treatment group
VAR(1)	first-order vector autoregressive model

6.1 Topic and Structure of the Chapter

In general, stress is not a one- but a multi-dimensional phenomenon, as somatic and psychological processes and context conditions constitute a complex system of relations (Adam, 2006; Kärner et al., 2017; Kemeny, 2003). Stress manifests itself on the physiological level and also on the level of subjective experience. Person-related characteristics can promote or inhibit the successful management of stress (e.g. Kärner, 2017; Kärner et al., 2018; Raemdonck et al., 2014). Likewise, workplace conditions can influence whether stress results or is prevented (e.g. Panari et al., 2010; Sembill, 2015). From a research methodological perspective, these circumstances make a multimodal assessment of stress necessary, whereby physiological, observational/context-related and experience-based data are considered in the analyses.

In that regard, this chapter is dedicated to stress in working and learning contexts, with special consideration of psychophysiological methods for stress assessment. In the second section, stress-inducing and stress-buffering potentials of working and learning in the workplace are described. The third section is dedicated to the description of physiological stress responses. In particular, it considers the functioning of the hypothalamic-pituitary-adrenal (HPA) axis, which triggers the release of the cortisol hormone, before having a closer look at the structure of the autonomic nervous system (ANS), which triggers cardiovascular functions that are measurable via heart rate (HR) and heart rate variability (HRV) markers. In the fourth section, three selected research examples are reported to illustrate the application of psychophysiological methods for measuring stress in the context of working and learning. First, a laboratory experiment aimed at identifying the effects of work-related problem solving on cortisol measures and cardiovascular markers is described. Second, a single-subject study investigating time-lagged relationships between the self-reported demands on teachers resulting from daily classroom work and cardiovascular markers is reported. Third, a study assessing interactions between baseline cortisol levels of trainee industrial clerks, observed classroom demands and the trainee industrial clerks' stress experience is described. These three examples illustrate different application and measurement scenarios. The selected examples include both an experiment (Example 1) and two field studies (Examples 2 and 3). With regard to the respective measurement frequencies and data calculations, the examples involve both repeated measurements and aggregated measured values: the secretion of cortisol and cardiovascular stress markers at different points in time (Examples 1 and 2) and the overall secretion of cortisol over a specific time period (Example 3). Furthermore, different interrelations between physiological, psychological (experience-based self-assessments) and context-based measurements are considered in the examples: interrelations between physiological stress markers and experience data resulting from self-assessments (Example 2) and interrelations between physiological markers, observational data and experience data (Example 3). The chapter concludes with an integrative framework for multimodal assessment and evaluation of learning and working environments (Sect. 6.5) and with recommendations for research practice (Sect. 6.6).

6.2 Stress-Inducing and Stress-Buffering Potentials of Working and Learning in the Workplace

When considering working and learning in the workplace, different relationships to stress-related processes can be identified: both the effects of stress on learning and performance but also the buffering effects of learning-related activities and outcomes with regard to the resulting stress reactions.

Regarding the effects of stress on learning and performance, generally short-term and mild stress increases cognitive abilities whereas severe and prolonged stress can disrupt or block corresponding memory functions and thus hinder learning processes (Sapolsky, 1996). As psychophysiological research shows, too much stress can impair both long-term memory performance (e.g. De Quervain et al., 2000; Sauro et al., 2003) and working memory performance (e.g. Oei et al., 2006; Schoofs et al., 2008), which includes the entire information processing system (encoding, consolidation, retrieval) (Vogel & Schwabe, 2016). Impairments in cognitive performance under stress are due, among other things, to dysregulation of the HPA axis and an increased basal cortisol level (Lupien et al., 1994; see Sect. 6.3.1). On a psychological level, it can be assumed that orientation, planning and control processes are increasingly impaired at a high stress level. Work-related demands strain or overstrain mental capacity and tie up task-related attention. Furthermore, cognitive resources are needed to maintain concentration and regulate negative affectivity, as the latter can in turn negatively influence task-related attention and information processing (Bartholdt & Schütz, 2010). In that regard, the relationship between stress and learning performance can be described by its non-linear relationship (so-called Yerkes-Dodson law). It can be assumed that different performance optima apply to different individuals, marking the activation level at which individual learning and memory performance is at its maximum (Gluck et al., 2010; Sauro et al., 2003). In addition to the different levels of activation, the type of stress is important and whether the stress is associated with hindrances or challenges (Stevenson & Harper, 2006). In this case, stress associated with hindrances is linked to lower learning performance whereas stress associated with challenges is associated with higher learning performance (LePine et al., 2004).

Regarding learning-related activities and outcomes in the workplace, it is noted that they can buffer stress by contributing to a reduction of uncertainty and an improvement in well-being. In this context, Panari et al. (2010) showed that the demand/strain relationship was stronger when the opportunity for workplace learning and development was low within the organization. Raemdonck et al. (2014) found in their study that job demands and a self-directed learning orientation positively influence learning at work. In addition to learning-related buffering effects with regard to the stress experienced, it has been shown that acquired skills and knowledge or professional expertise can also serve as a stress buffer (e.g. Kärner, 2017). For example, Nikolova et al. (2014) report buffering effects of high professional knowledge and skills on the relationship between job-related demands and emotional exhaustion. Considering the differences between experts and novices, it can be stated that high pressure prevents novices from competently performing complex skills because attention is directed to task-irrelevant aspects and the lack of automaticity can lead to an increase in procedural errors. Expertise in this context is characterised, among other things, by the extent to which a person has internalised strategies for coping with arousal-provoking situations that enable them to minimise stress in such situations and focus on the task at hand. In that case, continuous and repeated confrontation with arousal-provoking situations causes the operational mode in action-regulation processes to shift from an intentional mode to an automated mode. The automated mode hence bears stress-buffering potential (Nitsch, 1982; Tenenbaum et al., 2008).

Arousal-provoking situations in professional domains can have various forms. In many occupations, workers are confronted with demands that are work-specific problem-solving issues (Candy & Matthews, 1998; Rausch et al., 2015; Tynjälä, 2013). In general, a person is confronted with a problem if he or she has a goal but does not immediately know how to resolve it (Newell & Simon, 1972). In many problem situations at the workplace, the given state, goal state or operators are not clearly specified (cf. Mayer & Wittrock, 2006). Furthermore, specific work characteristics act as barriers to moving from an actual state to the desired goal state, with

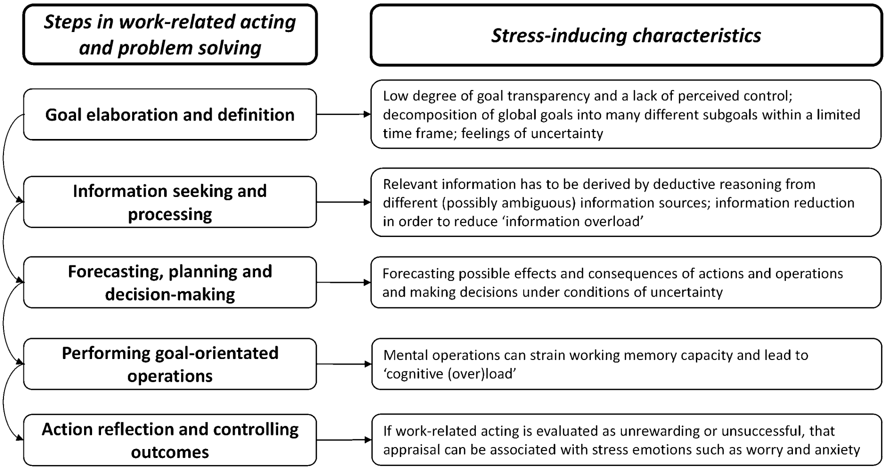

Fig. 6.1 Stress-inducing potentials of work-related acting and problem solving

goal-orientated actions being necessary to overcome such barriers (Frensch & Funke, 1995; Funke, 1991; Hacker, 2003). In this regard, Sembill (1992) stresses the necessity to understand learners' emotional states as a constitutive element in the acquisition and changes of knowledge, as well as in problem solving. Dörner and Wearing (1995) argue that the process of action regulation within problem solving can be subdivided into various partial activities that bear the stress-inducing potentials characterised in Fig. 6.1 (for a detailed description, see: Kärner, 2017; Kärner et al., 2018).

In addition to individual coping characteristics and supportive learning and working conditions in the workplace, social structures in the form of dyadic relationships and social (professional and private) networks are of crucial importance in the context of stress and coping, for instance, with demands resulting from work-related collaborative problem-solving activities (e.g. Von Davier & Halpin, 2013). Depending on their quality and characteristics, social relationships at the workplace can contribute to coping with job-related demands but can also become an additional stress factor. Therefore, in connection with stress and coping processes in the workplace, corresponding processes must be considered within a system of complex social interdependencies. Hence, successful coping with work-related stress may not only be a product of the individual but rather one of the work-related networks an individual is integrated into (Kärner et al., 2021d).

As illustrated, working and learning in the workplace can bear both stress-inducing and stress-buffering potentials. In the following, physiological stress reactions are examined in more detail.

6.3 Physiological Stress Responses

6.3.1 Hypothalamic-Pituitary-Adrenal Axis and Cortisol Response to Workplace Stress

From a somatic point of view, Selye (1973, p. 692) defines stress as 'the nonspecific response of the body to any demand made upon it'. There are two primary somatic systems involved in physiological stress reactions: the HPA axis and the ANS (Sect. 6.3.2). The HPA axis triggers the release of the cortisol hormone (e.g. Michaud et al., 2008). Due to intermediate hormonal regulation mechanisms, the cortisol concentration in response to a stressor peaks about 15–20 min after acute stress induction (Dickerson & Kemeny, 2004). Moreover, cortisol levels follow a natural diurnal rhythm, with a peak after getting up in the morning (the so-called cortisol awakening response) and a decrease continuously thereafter during the day (Guyton, 1986; Rensing et al., 2006; see Fig. 6.2a).

Takahashi et al. (2005, p. 353) conclude that individuals 'with chronic high levels of cortisol may have blunted acute neuroendocrine response to experimental social stress, possibly due to saturated HPA reactivity'. Extant research, including meta-analyses, shows that work-related stress increases cortisol levels (Michaud et al., 2008) or may result in incomplete recovery after performing a demanding task (Sluiter et al., 2000). Concerning work-related stress, Caplan et al. (1979) found that the level of office workload affects the circadian rhythm of cortisol. Thus, high-workload employees did not show the expected decrease in cortisol concentration over the day, which in turn indicates a chronic elevation in the level of HPA activation. Lundberg and Hellström (2002) found that the overall workload, assessed by the amount of overtime worked, was positively associated with the morning

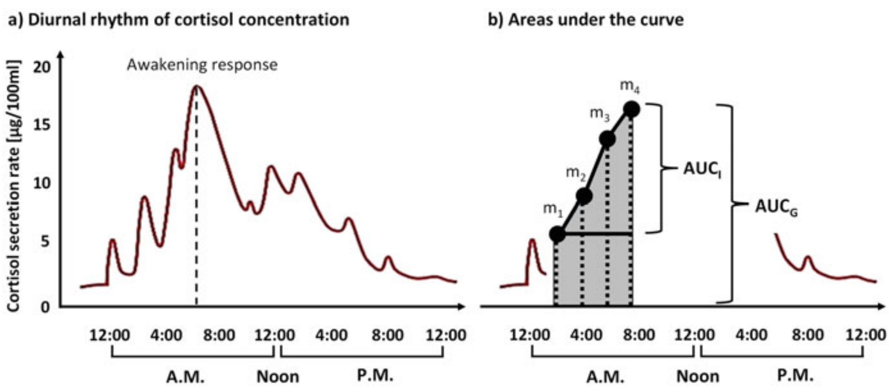

Fig. 6.2 (**a**) Diurnal rhythm of cortisol concentration and (**b**) areas under the curve. Own illustrations (referring to Fekedulegn et al., 2007; Guyton, 1986; Pruessner et al., 2003), with $m_{1...4}$ denoting individual measurements

salivary cortisol level. Thus, work-related overloads can attenuate the natural diurnal decrease in cortisol (Stawski et al., 2013).

In endocrinological studies, a frequently used method for assessing the overall secretion of cortisol over a specific time period (e.g. over 1 day or over a period of specific stimulus exposition) is the 'area under the curve with respect to the ground' (AUC_G) (Pruessner et al., 2003; see Fig. 6.2b). For calculating the individual areas under the curves, a common method is the trapezoid formula for AUC_G (with identical periods of time in between taking the measurements) following Pruessner et al. (2003) (Eq. 6.1):

$$\mathrm{AUC}_G = \sum_{i=1}^{n-1} \frac{\left(m_{(i+1)} + m_i\right)}{2} \tag{6.1}$$

with:

- m_i denoting the individual measurements at time points *1* to *i* and
- *n* denoting the total amount of measures.

AUC_G indicates the total area under the curve of all cortisol measures over a specified time period and acts as an indicator for the 'total hormone output' (Pruessner et al., 2003, p. 928). There is theoretical and empirical evidence that AUC_G is related to chronic stress, but current research shows that high chronic stress levels are not always associated with elevated baseline cortisol. Differences found in empirical relations between AUC_G and chronic stress may result from methodological, theoretical and analytical issues (Saxbe, 2008).

Another common measure is the 'area under the curve with respect to increase' (AUC_I), which 'is calculated with reference to the baseline measurement and it ignores the distance from zero for all measurements and emphasises the changes over time' (Fekedulegn et al., 2007, p. 651) (Eq. 6.2; Pruessner et al., 2003):

$$\mathrm{AUC}_I = \left(\sum_{i=1}^{n-1} \frac{\left(m_{(i+1)} + m_i\right)}{2}\right) - (n-1) \cdot m_1 \tag{6.2}$$

with:

- m_i denoting the individual measurements at time points *1* to *i*,
- *n* denoting the total amount of measures and
- m_1 denoting the first measurement.

On comparing AUC_G and AUC_I, Fekedulegn et al. (2007, p. 651) point out that 'AUC_G is assumed to be a measure more related to total hormonal output, whereas AUC_I is a parameter that emphasises the changes over time and is more related to sensitivity of the system'.

6.3.2 Autonomic Nervous System and Cardiovascular Stress Markers

In addition to the HPA axis, the ANS is also involved in the physiological stress reaction (Kemeny, 2003). The situational adaptability of the ANS is indicated by the antagonistic relationship between the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS enables the organism to adapt rapidly to demanding situations whereas the PNS controls the physiological resting functions. The dynamic interplay between the SNS and PNS, and thus the adaptability of the ANS according to changing situational demands, can be measured by changes in HR and HRV, both of which are indicators of the cardiac system's activity (Kemeny, 2003; Malik et al., 1996). HR (measured in beats per minute: bpm) indicates the contractions of the heart per minute. As a time-domain measure of HRV, the RMSSD (measured in milliseconds: ms) is operationalised as the root mean square of the sum of the squares of differences between adjacent NN (normal-to-normal) intervals. The RMSSD represents an index for the reactivity of the PNS (Thomas et al., 2019). As a frequency domain measure of HRV, the LF/HF ratio indicates power in the low frequency range (LF, 0.04–0.15 Hz, measured in ms^2) in relation to the power in the high frequency range (HF, 0.15–0.4 Hz, measured in ms^2). LF primarily reflects the activity of the SNS whereas HF primarily reflects the activity of the PNS (Malik et al., 1996).

Reviewed literature and existing meta-studies (Castaldo et al., 2015; Järvelin-Pasanen et al., 2018; Kim et al., 2018) mainly indicate increases of the HR and LF/HF ratio and decreases of the RMSSD under conditions of work-related stress (summarised in Table 6.1).

Table 6.1 The HR, RMSSD and LF/HF ratio as indicators of work-related stress

ANS indicators	Measures and correlates	Reference literature
HR ↑	Job stress questionnaire Simulated emergency case	Clays et al. (2011) Kaegi et al. (1999)
RMSSD ↓	Effort–reward imbalance questionnaire Occupational stress questionnaire Job content questionnaire Physical–mental task Effort–reward imbalance questionnaire	Garza et al. (2015) Lindholm et al. (2009) Loerbroks et al. (2010) Taelman et al. (2011) Uusitalo et al. (2011)
LF/HF ratio ↑	Job stress questionnaire Effort–reward imbalance questionnaire	Clays et al. (2011) Garza et al. (2015)
	Computer work task	Hjortskov et al. (2004)

Note: Overview adapted from Castaldo et al. (2015), Järvelin-Pasanen et al. (2018) and Kim et al. (2018)

6.4 Example Studies

From a research methodological perspective, there are different approaches to studying work-related stress in more detail. For example, there are experimental studies on cardiac regulation effects for particular office tasks (Taelman et al., 2011), field studies using ambulatory electrocardiography recordings (Clays et al., 2011), diary studies on relationships between problem-solving activities in everyday office work and stress-related experience (Rausch et al., 2015) or meta-analytic studies on effects of workplace stressors on health outcomes (Goh et al., 2015). Furthermore, there are studies that use both somatic stress responses and self-reports of stress experience (Marchand et al., 2016) or only consider one or the other. To provide insights into possible study scenarios, three selected studies are presented below to illustrate different application and measurement scenarios, measurement frequencies and data calculations, as well as different relationships between physiological, psychological (experience-based self-assessments) and observation-based measurements (see Table 6.2).

Table 6.2 Characteristics of the example studies[a]

Example	Study design	Measures	Relationships investigated between measures
Study 1	Experimental study	Only physiological data (HR, RMSSD, cortisol) reported, situational demands in the form of a standardised problem-solving task kept constant	Only analyses with physiological data reported
Study 2	Field study	Physiological data (HR, RMSSD), experience data/self-assessments (self-reported demands)	Time-lagged relationships between experience data and physiological data
Study 3	Field study	Physiological data (baseline cortisol), experience data (self-reported stress), observational data (‘objective’ classroom demands)	Interactions between physiological data and observational data concerning experience data

[a]It is important to note that the respective publications contain additional information that, for reasons of limited space, will not be reported here

6.4.1 *Example Study 1: Effects of Work-Related Problem Solving on Cortisol Response and Cardiovascular Stress Markers*

As described in Sect. 6.2, on the one hand, work-related stress can affect learning and performance; on the other hand, learning-related activities and outcomes in the workplace can buffer work-related stress. As the acquisition of competence and expertise can be inhibited by stress on the one hand but acquired competences can also have a stress-compensating effect on the other hand, special attention must be paid to the conditions of stress induction in the context of work-related problem-solving activities. In that regard, this example describes the findings of an experimental study that investigated associations between work-related problem solving and both ANS and HPA axis activity (for a detailed description of the study and findings, see Kärner et al., 2018).

6.4.1.1 Sample and Method

The sample recruited for the experiment consisted of 20 male and 38 female students from a faculty of social sciences and economics with an average age of 24.8 years (SD = 3.5). To identify the effects of complex problem solving on physiological stress responses, participants were confronted with a computer-based problem scenario from the business administration domain (Rausch et al., 2016; Seifried et al., 2016). The scenario required participants to select an optimal supplier by calculating acquisition prices and performing a value analysis. In order to demonstrate that stress responses were related to problem-solving activities rather than to using the software, a treatment–control group design was chosen: the treatment group (TG, $n = 41$) worked on the problem scenario and was asked to solve the complex problem; the control group (CG, $n = 17$) was instructed to inspect the computer-based scenario and check the software's usability but were not required to solve the problem. Further information on the study procedure and measurements can be found in Fig. 6.3a.

6.4.1.2 Measures

Cardiovascular reactivity was assessed in terms of HR and HRV was operationalised as the RMSSD. Participants' cardiovascular reactivity was measured continuously during the task via a chest belt and storage devices from Medeia® that recorded the data wirelessly on an integrated memory chip. The internal sample frequency was 1 kHz and the HR was transmitted to Qhrv Assessment software (Medeia®, Santa Barbara, CA, USA)[1] in order to calculate the RMSSD indicator.

[1] https://www.medeia.com/

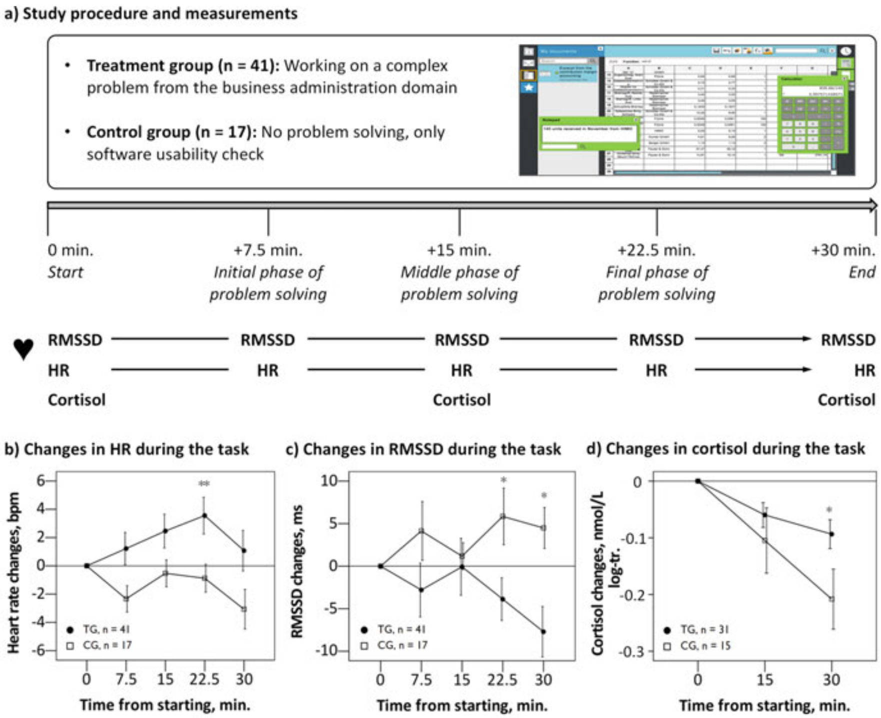

Fig. 6.3 (**a**) Study procedure and measurements, plus changes in HR (**b**), RMSSD (**c**) and cortisol concentration (**d**) during the task. Error bars represent ±1SE; $^{**}p < 0.01$. (Own illustration referring to Kärner et al., 2018)

For assessing the cortisol response, saliva was collected via unstimulated passive drool into polypropylene microtubes (Sarstedt®, Nürmbrecht, Germany) using a shortened straw.[2] Saliva samples were stored at −20 °C immediately after collection. Before analysis, the samples were thawed at room temperature, vortexed and centrifuged twice at 2500 *g* for 15 min. Cortisol concentration was quantified using cortisol saliva immunoassay kits from IBL® (Hamburg, Germany). The supernatant from each probe was transferred in duplicate to a precoated microwell plate on the day of the assay and the optical density of the probes was measured at 450 nm and 620 nm (reference) using a 96-well ELISA (enzyme-linked immunosorbent assay) reader (Thermo Fisher®, Vantaa, Finland). Cortisol concentration was calculated from the optical density using Scan It® software (version 3.0, Thermo Scientific®). Intra-assay coefficients of variance were below 5% and inter-assay coefficients were below 10%. Because repeated cortisol measures (such as other endocrine time-series data) often lack normality and homoscedasticity due to characteristics of the biochemical analysis techniques and nonlinear dynamics of natural

[2]Biochemical analysis for measuring the cortisol concentration was carried out by the co-author (N. Minkley, Bochum) of the article.

determinants of endocrine processes (cf. Miller & Plessow, 2013), the cortisol data were log-transformed to ensure an approximately normal distribution.

Cardiovascular activity was measured continuously during the work-related problem-solving task and also during a recovery period after finishing the task. HR and HRV were aggregated to the following intervals by arithmetic averages: 0 min, start of the experiment; +7.5 min, initial phase of problem solving; +15 min, middle phase of problem solving; +22.5 min, final phase of problem solving; +30 min, end of the experiment (see Fig. 6.3a). Saliva samples were collected at the start, 15 min into the scenario and at the end of the scenario. Taking into account individual changes from the starting values, we subtracted the start values from the subsequent values to obtain baseline-corrected data for HR, RMSSD and cortisol concentration (cf. Roberts et al., 2004). The experiment was conducted between 10:00 a.m. and 12:00 noon to control for diurnal variation in cortisol levels. Participants were instructed to refrain from food or drink an hour before the experiment in order to avoid possible contamination of saliva.

6.4.1.3 Findings

Mean differences between groups for each point of measurement were tested via *t*-tests. The results show that at 22.5 min ($t[50.98] = 2.72$, $p = 0.009$) the TG participants showed significantly higher HR than the CG participants (Fig. 6.3b). Furthermore, TG participants showed significantly lower RMSSD values at 22.5 min ($t[54] = -2.11$, $p = 0.040$) and 30 min ($t[54] = -2.38$, $p = 0.021$) than CG participants (Fig. 6.3c). At the end of the problem-solving task at 30 min ($t[44] = 2.21$, $p = 0.032$), TG participants showed significantly higher cortisol values than CG participants (Fig. 6.3d).

6.4.2 Example Study 2: Time-Lagged Effects of Teachers' Experienced Classroom Demands on Autonomic Stress Reactions

The second example study focuses on teachers' workplace stress. According to Kyriacou and Sutcliffe (1978, p. 2), 'teacher stress may be defined as a response of negative affect [. . .] by a teacher usually accompanied by potentially pathogenic physiological and biochemical changes (such as increased heart rate or release of adrenocorticotrophic hormones into the bloodstream) resulting from aspects of the teacher's job [. . .]' In that regard, in this example the time-lagged relationships between teachers' experienced classroom demands and autonomic stress reactions will be examined (for a detailed description of the study and findings, see Kärner & Höning, 2021).

6.4.2.1 Sample and Method

A single-case, short-term longitudinal study was conducted and the participant was a 31-year-old female teacher who had worked since 2014 at a German high school. The participant was observed over a period of 3 weeks during their regular lessons and a total of 66 observation points with a mean time lag ($t - 1 \rightarrow t$) of about 26 min were collected.

6.4.2.2 Measures

For the assessment of teachers' experienced classroom demands, a smartphone-based experience sampling application (Movisens GmbH, Karlsruhe, Germany)[3] was used and the participant was asked to rate classroom demands at randomly chosen points in time. One item referred to the assessment of achievement-related diversity in class and the perceived responsibility to fulfil the teacher's role to respond to students' different needs (cf. Van Dick, 1999) (item: 'I cannot respond to students' different needs') and another item referred to the disquietude in class due to disturbing noise (cf. Meder et al., 2008) (item: 'There is disquietude in class'). The two items were rated on a five-point Likert scale from 1 ('I strongly disagree') to 5 ('I strongly agree').

The participant's autonomic reactivity (HR, RMSSD) was measured continually during the observation period via a psychophysiological ambulatory measurement system for the assessment of electrocardiography signals (Movisens GmbH, Karlsruhe, Germany). For statistical analysis, the assessed electrocardiography measures were aggregated via arithmetic means (2 min before the experience sampling signal, 2 min after the experience sampling signal and the minute including the experience sampling signal) in order to synchronise them with the self-ratings for experienced classroom demands.

6.4.2.3 Findings

To examine time-lagged relationships between teachers' experienced classroom demands and autonomic stress reactions, an $N = 1$ time-series analysis with a bivariate cross-lagged model for continuous dependent variables was used, also referred to as a first-order vector autoregressive VAR(1) model (Muthén & Muthén, 1998–2017). The two models, which will be described below, show almost significant ($p < 0.10$) time-lagged effects of teachers' perceived classroom demands on autonomic stress reactions.

Figure 6.4a illustrates the relationship between 'I cannot respond to students' different needs' and the HR. Here, an almost significant ($p = 0.066$) positive effect

[3] https://www.movisens.com/de/

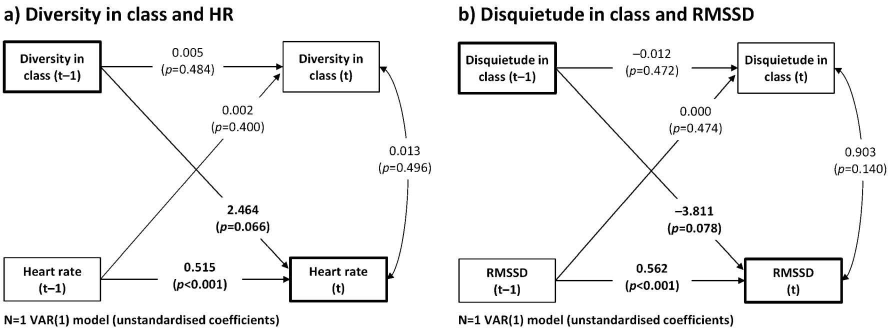

Fig. 6.4 Teachers' experienced classroom demands and autonomic reactivity. (Own illustration referring to Kärner & Höning, 2021)

of the experienced demand of achievement-related diversity in class (at time *t*–1) on the HR at time *t* was found. Thus, the more demanding the diversity in class at *t*–1 evaluated by the participant, the higher her HR at time *t*.

Figure 6.4b illustrates the relationship between 'There is disquietude in class' and the RMSSD. Here, an almost significant ($p = 0.078$) negative effect of the experienced demand of disquietude in class due to disturbing noise (at time *t*–1) on the HRV indicator RMSSD at time *t* was found. Thus, the more demanding the disquietude in class at *t*–1 evaluated by the participant, the lower her HRV (RMSSD) at time *t*. The time-lagged effects found remain stable even when controlling for the autocorrelations of the two cardiovascular indicators.

6.4.3 Example Study 3: Interactions Between Baseline Cortisol Levels of Trainee Industrial Clerks and Classroom Demands Concerning Stress Experience

Examples 1 and 2 above illustrate relationships between situational factors and physiological outcome measures, whereas Example 3 here illustrates the relationship between persons, situational characteristics and the stress outcome.

The acquisition of problem-solving competence on the learners' side is considered an essential outcome of professional teaching–learning processes. Particularly in the context of self-organised, problem-oriented learning situations, the ability to solve complex problems is promoted by overcoming barriers, thus supporting individual competence development (Dochy et al., 2003; Sembill et al., 2002; Vygotskii, 1978). The individual learning processes are by no means free of experienced stress, which results from an interaction of individual characteristics and characteristics of the learning environment (Achtenhagen, 1978). If a potential source of stress is evaluated as a threat and the associated perceived stress exceeds the person-related and/or environment-related resources, or if these are assessed as deficient in terms of appropriate coping, then strain turns into stress. Furthermore, individual physiological vulnerability factors, such as increased cortisol baseline levels, can contribute to situational demands being experienced as stressful (Sembill, 2012). In that regard, this third example study addresses the question of the extent to which increased cortisol baseline levels as a physiological vulnerability factor of learners influence the relationship between 'objective' classroom demands and individual stress experience (for a detailed description of the study and results, see Kärner et al., 2017).

6.4.3.1 Sample and Method

Within a short-term longitudinal study, 53 trainee industrial clerks (18 males, 35 females; mean age 19.5 years, SD = 4.8) from two school classes (28 students in class A; 25 students in class B) in a public German vocational training school were investigated during nine school lessons on the subject of 'economic business processes' (Kärner, 2015).

Before investigating the lessons in school via video-based analysis of classroom demands, the baseline cortisol concentration was measured. During the learning processes in school, the trainee industrial clerks' stress experience was measured in 10-minute time intervals via the continuous-state sampling method (Sembill et al., 2008) using a mobile device (Palm Tungsten E2®) (see upper part of Fig. 6.5).

6.4.3.2 Measures

Baseline cortisol concentration was measured using biochemical salivary analysis (Immumed Ltd., Institute for applied Immunology, Munich, Germany). The trainee industrial clerks received small plastic tubes and were instructed to provide saliva samples during a relaxed weekend day without school or work at the following times: 8:00 a.m., 12:00 noon, 4:00 p.m. and 8:00 p.m. Areas under the curves with respect to the ground (AUC_G; see Fig. 6.2, Eq. 6.1 and the upper part of Fig. 6.5) were calculated for assessing baseline cortisol secretion as an indicator for chronic elevation of HPA activation.

Observed classroom demands were operationalised by the amount of student-centred learning and by the quality of cognitive challenge during education. When assessing the amount of student-centred learning, time intervals of 15 s each were coded on the basis of the videos. When assessing the quality of cognitive challenge during education, time intervals of 1 min each were coded and we used a four-point Likert-type scale based on Bloom's taxonomy to assess the complexity of learning content and tasks to be worked on (increasing levels of difficulty: 0 = 'apply already known content', 1 = 'analyze contexts', 2 = 'synthesize complex ideas', 3 = 'evaluate complex problems'; cf. Bloom et al., 1956). The video coding was realised with Videograph®software.[4] The single

[4] http://www.dervideograph.de/

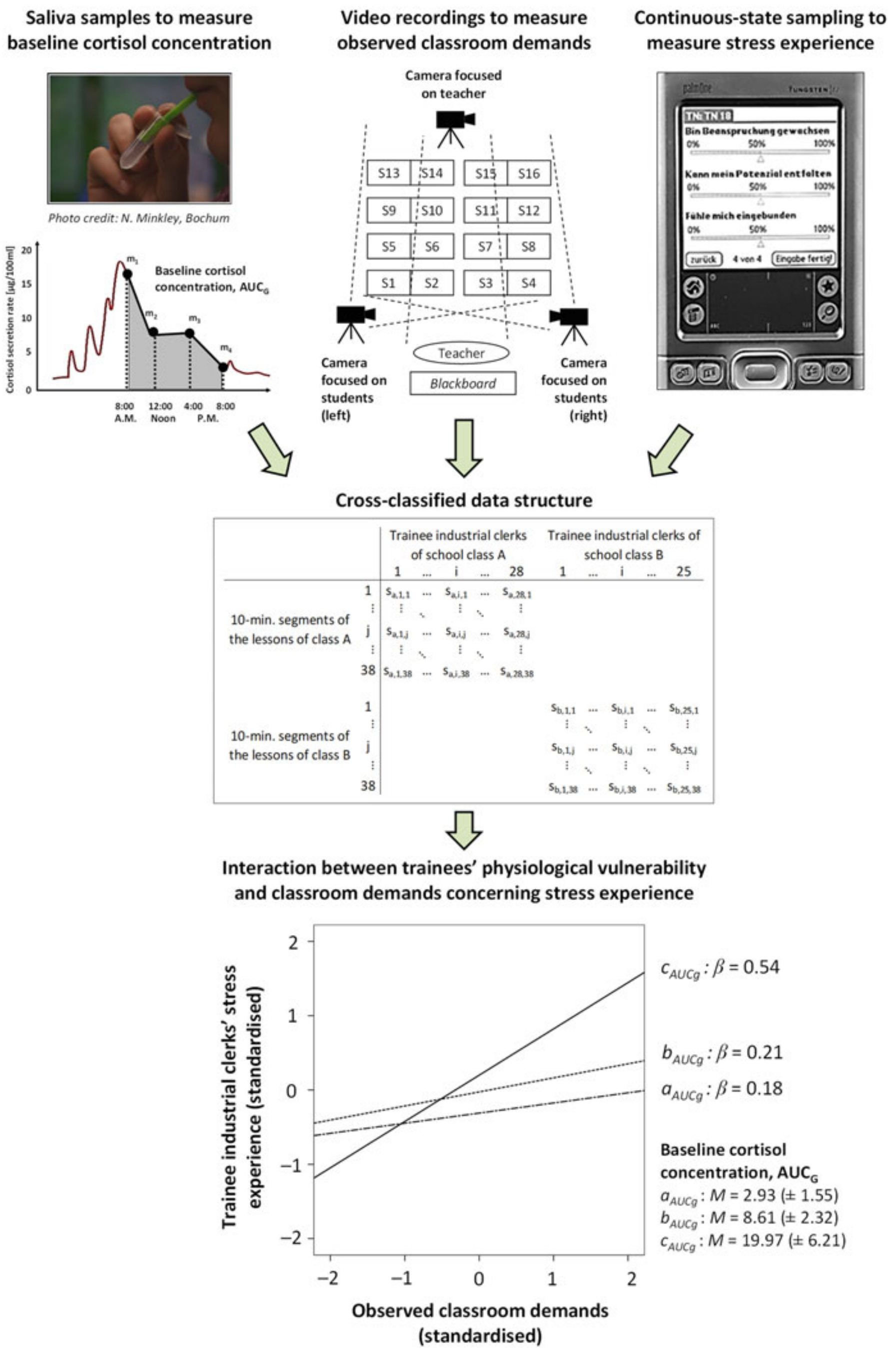

Fig. 6.5 Modelling of trainee industrial clerks' stress experience. (Own illustration referring to Kärner et al., 2017)

coded intervals were aggregated to 10-min intervals, synchronising information on context conditions and the person-related experience sampling data. Measures of the amount of student-centred learning and the cognitive challenge during education were aggregated via factor scores as estimated values of the factor 'classroom demands'. Low scale values indicate mainly teacher-centred, question-developing instruction with low content complexity; high scale values indicate mainly student-centred instruction in the form of partner or group work, in which tasks or problems with high content complexity were worked on. There was a maximum of 38 measurement points for observed classroom demands per school class (see 10-min segments of the lessons of class A and class B in the middle part of Fig. 6.5).

Trainee industrial clerks' stress experience was measured via two items ('I'm under time pressure' and 'I'm under pressure to succeed'). Trainee industrial clerks had the opportunity to rate their actual experience on a visual analogue mood scale from 0 ('I fully disagree') to 100 ('I fully agree'). The two items were aggregated via factor scores as estimated values of the factor 'stress experience'. There was a maximum of 38 measurement points for stress experience per student.

6.4.3.3 Findings

Using cross-classified multilevel longitudinal modelling (see the illustration of the cross-classified data structure in the middle part of Fig. 6.5), interactions between trainee industrial clerks (characterised by baseline cortisol concentrations) and situational conditions (characterised by observed classroom demands) were assessed. The results show that baseline cortisol concentration affects the relation between observed classroom demands and trainees' stress experience: trainees with above-average AUC_G values ('high physiological vulnerability') show a stronger increase in stress experience with increasing classroom demands than students who have below-average AUC_G values ('low physiological vulnerability'). The bottom part of Fig. 6.5 illustrates the group-specific linear growth trajectories of different AUC_G groups classified via three percentiles according to their AUC_G values.

In the three examples reported, different methods and approaches of stress assessment in the context of learning and working were illustrated: in the first example, only physiological stress indicators were taken into account; in the second example, physiological stress indicators were combined with experienced situational stress factors; and in the third example, the triangulation of physiological, observational and experience-based data was performed. Let us now go back to the theoretical-conceptual level and systematise possible modelling of person- and situation-related variables and their interactions.

6.5 Framework for Multimodal Assessment and Evaluation of Learning and Working Environments

An essential advantage of physiological methods can be located in the objectivity of measurement that is relatively independent of self-reporting biases (cf. Dawson et al., 2011). This seems to be important, especially when investigating implicit aspects of action processes that cannot be directly verbalised by the individual or observed via video-based analyses for example. However, one has to consider that physiological measures are not meaningful per se in the context of working and learning but have to be interpreted in their interplay with psychological parameters (e.g. traits, experiences), work- and learning-related behaviours and with particular situational stimuli in order to obtain a more 'holistic picture' of the interaction of individuals with their working and learning environment (Kärner, 2017; Kärner et al., 2017; Sembill & Kärner, 2018).

Taking into account that person-related situational behavioural, somatic and psychological states (e.g. varying cortisol levels, HR or RMSSD values and stress experience) are 'nested' within persons (characterised by various traits in the sense of relatively time-stable characteristics relevant in a specific working context, such as vocational expertise or coping skills) and the continuously changing situational conditions of the working context (e.g. varying demands of the tasks that persons have to fulfil or social support at the workplace), a deeper knowledge about the complex and even non-linear and/or time-lagged interrelations between traits, states and continuous changing situational context conditions seems to be essential for a more holistic understanding of working and learning. Thus, from a methodological point of view, the application of cross-classified data structures and corresponding statistical multilevel models (e.g. Heck et al., 2010; Nezlek, 2007) seems to be the means of choice. Here, cross-classification considers that the multiple state-measures are not only nested within persons but also belong to observation units of the working context corresponding to the time of measurement (Goldstein, 1994; Hill & Goldstein, 1998; for applications, see: Kärner & Kögler, 2016; Kärner et al., 2017). In that regard, Fig. 6.6 illustrates a possible framework for multimodal

		Working and learning context *1*			…	**Working and learning context** *k*			***Context***
		Persons (1…n) of context *1*				**Persons** (1…n) of context *k*			***Person***
		1	…	*n*		1	…	*n*	
Changing **situational conditions** (1…t) of the context *1*	1	State$_{1,1,1}$	…	State$_{1,n,1}$					***Situation***
	⋮	⋮	⋱	⋮					
	t	State$_{1,1,t}$	…	State$_{1,n,t}$					
…					⋱				
Changing **situational conditions** (1…t) of the context *k*	1					State$_{k,1,1}$	…	State$_{k,n,1}$	
	⋮					⋮	⋱	⋮	
	t					State$_{k,1,t}$	…	State$_{k,n,t}$	

Fig. 6.6 Framework for multimodal assessment and evaluation of learning and working environments. (Own illustration referring to Sembill & Kärner, 2018)

assessment and evaluation of learning and working environments with person-related states (indexed with *k,n,t*; e.g. behaviours, psychological states and physiological states) nested within persons (1...*n*) working/learning in a specific context (1...*k*; e.g. different departments, companies or industries) and exposed to continuously changing situational context conditions (1...*t*; e.g. varying levels of work-related autonomy or demands, or hindrances or challenges of working conditions).

The above-mentioned aspects and the framework at hand imply a broader understanding of heterogeneity that considers both physiological and psychological relatively time-stable characteristics and time-variant states, as well as relatively time-stable and time-variant environmental characteristics and conditions. In that regard, at least the following three sources of heterogeneity can be identified that have to be considered when analysing and evaluating learning and working environments:

- Inter-individual variability between individuals within a learning or working context (e.g. different levels of expertise, emotional-motivational dispositions, physiological dispositions);
- Intra-individual variability within individuals over time and across different situations (e.g. temporally varying rates of progression in knowledge and skill acquisition, temporally varying hormonal responses);
- Variability in the interactions between individual characteristics and states and situational characteristics of the learning or work context (e.g. situational anxiety as a function of workplace conditions and individual-specific introversion) (cf. Kärner et al., 2021b, c).

Taken together, future research could thus focus on a 'multi-dimensional mapping' of work-related context conditions and on an analysis of corresponding effects on person-related outcomes (e.g. job satisfaction, learning success, well-being, stress). Such research could finally result in empirically validated 'multi-dimensional data cubes' that characterise the antecedents and effects of humanely working and learning environments (cf. Kärner & Höning, 2021; Kärner et al., 2017; Sembill & Kärner, 2020). This also seems particularly relevant for working and learning at the workplace, because what is ultimately expressed as performances and working and learning culture, and for which internal dispositions, competencies and mental processes are hypothetically attributed, is dependent on physiological processes. What is opened up socially, technically and pedagogically as an opportunity also has a feedback effect on physiological processes through these interactions – up to and including a possible change in the genetic material (so-called epigenesis). With regard to a holistic view of learning and working, it is therefore indispensable to consider not only the social, group-interactive and individual levels but also the physiobiological level, because physiobiological processes are the basis for all other processes and structures (Kärner et al., 2021a; Sembill & Kärner 2018; for an in-depth discussion of the relationship between mental and somatic processes, see: Beck, 1994).

6.6 Concluding Recommendations for Research Practice

This chapter has attempted to show that psychophysiological methods are generally well suited to studying stress in the context of working and learning in the workplace. Nevertheless, due to the limited extent of the chapter, the explanations had to remain superficial on many points. For the integration of psychophysiological or, in particular, multimodal stress assessment methods in empirical studies on working and learning at the workplace, we therefore recommend considering the following aspects in connection with study design and implementation as well as data analysis and interpretation of results:

- As mentioned above, physiological markers are often not meaningful per se in the context of research on working and learning, so it would first be necessary to find out which learning-related or psychological and/or physiological theory can serve as the basis for deriving corresponding hypotheses.
- With reference to corresponding theories, it would also be necessary to figure out which specific physiological indicators are related to which specific psychological and/or learning- and work-related variables (measured via tests or self-reports). In this case, the assumed relationships would have to be specified by means of corresponding hypotheses.
- Although the use of different modes of measurement is generally to be evaluated positively with regard to the avoidance of a common-method bias (cf. Podsakoff et al., 2003), it should be examined to what extent physiological indicators can validly reflect corresponding psychological constructs (measured via tests or self-reports) and/or situational measures.
- If the characteristics of a learning and working environment are included in the hypotheses and empirical analyses by means of observation-based methods, it would be necessary to define precisely which interrelationships are expected between the external (context-/stimulus-related) characteristics and psychological and physiological processes and responses.
- In educational research, main effects are often not meaningful per se due to complex interaction structures of variables and there is therefore a risk of over-interpreting corresponding main effects. Thus, possible interaction effects between the variables of interest should be theoretically founded and empirically explored (cf. Kärner et al., 2017). Hence, it is useful to make pre-assumptions about possible interactions between personal characteristics, states and continuously changing context conditions, as well as about possible temporal dynamics between time-varying indicators of psychological and physical processes and environmental conditions.
- Especially for physiological processes and markers, such as the cortisol measures and cardiovascular markers described in this chapter, it is evident that these often follow less linear and more non-linear principles (e.g. Friederichs et al., 2020; Stroud et al., 2004), which must be taken into account in corresponding evaluations.

References

Achtenhagen, F. (1978). Beanspruchung von Schülern. Methodisch-didaktische Aspekte [Students' stress. Methodical and didactical aspects]. In *BMBW Werkstattberichte, Bundesminister für Bildung und Wissenschaft*. Aumüller Druck KG.

Adam, E. K. (2006). Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. *Psychoneuroendocrinology, 31*(5), 664–679. https://doi.org/10.1016/j.psyneuen.2006.01.010

Bartholdt, L., & Schütz, A. (2010). *Stress im Arbeitskontext. Ursachen, Bewältigung und Prävention*. Beltz.

Beck, K. (1994). Das Leib-Seele-Problem und die Erziehungswissenschaft. Ein Orientierungsversuch. In G. Pollak & H. Heid (Eds.), *Von der Erziehungswissenschaft zur Pädagogik?* (pp. 227–267). Deutscher Studien Verlag.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.

Candy, P., & Matthews, J. (1998). Fusing learning and work: Changing conceptions of workplace learning. In D. Boud (Ed.), *Current issues and new agendas in workplace learning* (pp. 12–29). National Centre for Vocational Education Research.

Caplan, R. D., Cobb, S., & French, J. R. P., Jr. (1979). White collar workload and cortisol: Disruption of a circadian rhythm by job stress? *Journal of Psychosomatic Research, 23*(3), 181–192. https://doi.org/10.1016/0022-3999(79)90003-5

Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., & Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control, 18*, 370–377. https://doi.org/10.1016/j.bspc.2015.02.012

Clays, E., De Bacquer, D., Crasset, V., Kittel, F., De Smet, P., Kornitzer, M., Karasek, R., & De Backer, G. (2011). The perception of work stressors is related to reduced parasympathetic activity. *International Archives of Occupational and Environmental Health, 84*(2), 185–191. https://doi.org/10.1007/s00420-010-0537-z

Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011). The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics, 4*(2), 111–116. https://doi.org/10.1037/a0022619

De Quervain, D. J.-F., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience, 3*(4), 313–314. https://doi.org/10.1038/73873

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*(5), 533–568. https://doi.org/10.1016/S0959-4752(02)00025-7

Dörner, D., & Wearing, A. J. (1995). Complex problem solving: Toward a (computer-simulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving. The European perspective* (pp. 65–99). Lawrence Erlbaum.

Fekedulegn, D. B., Andrew, M. E., Burchfiel, C. M., Violanti, J. M., Hartley, T. A., Charles, L. E., & Miller, D. B. (2007). Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosomatic Medicine, 69*(7), 651–659. https://doi.org/10.1097/PSY.0b013e31814c405c

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 3–25). Lawrence Erlbaum.

Friederichs, E., Kärner, T., Friederichs, K. M., & Lehmeier, S. (2020). Too much of a good thing? Curvilinear relationships between self-calming and autonomic responses in paediatric patients. *Medical Hypotheses, 137*(109536), 1–11. https://doi.org/10.1016/j.mehy.2019.109536

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Erlbaum.

Garza, J. L., Cavallari, J. M., Eijckelhof, B. H., Huysmans, M. A., Thamsuwan, O., Johnson, P. W., & Dennerlein, J. T. (2015). Office workers with high effort-reward imbalance and overcommitment have greater decreases in heart rate variability over a 2-h working period. *International Archives of Occupational and Environmental Health, 88*(5), 565–575. https://doi.org/10.1007/s00420-014-0983-0

Gluck, M. A., Mercado, E., & Myers, C. E. (2010). *Lernen und Gedächtnis. Vom Gehirn zum Verhalten*. Spektrum.

Goh, J., Pfeffer, J., & Zenios, S. A. (2015). The relationship between workplace stressors and mortality and health costs in the United States. *Management Science, 62*(2), 608–628. https://doi.org/10.1287/mnsc.2014.2115

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research, 22*(3), 364–375. https://doi.org/10.1177/0049124194022003005

Guyton, A. C. (1986). *Textbook of medical physiology* (7th ed.). Saunders.

Hacker, W. (2003). Action regulation theory: A practical tool for the design of modern work processes? *European Journal of Work and Organizational Psychology, 12*(2), 105–130. https://doi.org/10.1080/13594320344000075

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. Routledge.

Hill, P. W., & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Statistics, 23*(2), 117–128. https://doi.org/10.3102/10769986023002117

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology, 92*(1–2), 84–89. https://doi.org/10.1007/s00421-004-1055-z

Järvelin-Pasanen, S., Sinikallio, S., & Tarvainen, M. P. (2018). Heart rate variability and occupational stress: Systematic review. *Industrial Health, 56*(6), 500–511. https://doi.org/10.2486/indhealth.2017-0190

Kaegi, D. M., Halamek, L. P., Van Hare, G. F., Howard, S. K., & Dubin, A. M. (1999). Effect of mental stress on heart rate variability: Validation of simulated operating and delivery room training modules. *Pediatric Research, 45*(4), 77A. https://doi.org/10.1203/00006450-199904020-00463

Kärner, T. (2015). *Erwartungswidrige Minderleistung und Belastung im kaufmännischen Unterricht. Analyse pädagogischer, psychologischer und physiologischer Aspekte*. Lang.

Kärner, T. (2017). A mixed-methods study of physiological reactivity to domain-specific problem solving: Methodological perspectives for process-accompanying research in VET. *Empirical Research in Vocational Education and Training, 9*(10). https://doi.org/10.1186/s40461-017-0054-3

Kärner, T., & Höning, J. (2021). Teachers' experienced classroom demands and autonomic stress reactions: Results of a pilot study and implications for process-oriented research in vocational education and training. *Empirical Research in Vocational Education and Training, 13*(8). https://doi.org/10.1186/s40461-021-00113-3

Kärner, T., & Kögler, K. (2016). Emotional states during learning situations and students' selfregulation: Process-oriented analysis of person-situation interactions in the vocational classroom. *Empirical Research in Vocational Education and Training, 8*(12). https://doi.org/10.1186/s40461-016-0038-8

Kärner, T., Sembill, D., Aßmann, C., Friederichs, E., & Carstensen, C. H. (2017). Analysis of person-situation interactions in educational settings via cross-classified multilevel longitudinal modelling: Illustrated with the example of students' stress experience. *Frontline Learning Research, 5*(1), 16–42. https://doi.org/10.14786/flr.v5i1.137

Kärner, T., Minkley, N., Rausch, A., Schley, T., & Sembill, D. (2018). Stress and resources in vocational problem solving. *Vocations and Learning, 11*(3), 365–398. https://doi.org/10.1007/s12186-017-9193-8

Kärner, T., Bottling, M., Friederichs, E., & Sembill, D. (2021a). Between adaptation and resistance: A study on resilience competencies, stress, and well-being in German VET teachers. *Frontiers in Psychology, 12*(619912), 1–16. https://doi.org/10.3389/fpsyg.2021.619912

Kärner, T., Keller, T., Schneider, A., Albaner, D., & Schumann, S. (2021b). Ein Rahmenmodell zur Gestaltung technologisch unterstützter adaptiver Lehr- und Lernprozesse [A framework for designing technologically supported adaptive teaching and learning processes]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 117*(3), 351–371. https://doi.org/10.25162/zbw-2021-0016

Kärner, T., Warwas, J., & Schumann, S. (2021c). A learning analytics approach to address heterogeneity in the classroom: The teachers' diagnostic support system. *Technology, Knowledge and Learning, 26*(1), 31–52. https://doi.org/10.1007/s10758-020-09448-4

Kärner, T., Weiß, J. K., & Heinrichs, K. (2021d). A social perspective on resilience: Social support and dyadic coping in teacher training. *Empirical Research in Vocational Education and Training, 13*(24), 1–26. https://doi.org/10.1186/s40461-021-00126-y

Kemeny, M. E. (2003). The psychobiology of stress. *Current Directions in Psychological Science, 12*(4), 124–129. https://doi.org/10.1111/1467-8721.01246

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation, 15*(3), 235–245. https://doi.org/10.30773/2Fpi.2017.08.17

Kyriacou, C., & Sutcliffe, J. (1978). A model of teacher stress. *Educational Studies, 4*(1), 1–6. https://doi.org/10.1080/0305569780040101

LePine, J. A., LePine, M. A., & Jackson, C. L. (2004). Challenge and hindrance stress: Relationships with exhaustion, motivation to learn, and learning performance. *Journal of Applied Psychology, 89*(5), 883–891. https://doi.org/10.1037/0021-9010.89.5.883

Lindholm, H., Sinisalo, J., Ahlberg, J., Jahkola, A., Partinen, M., Hublin, C., & Savolainen, A. (2009). High job control enhances vagal recovery in media work. *Occupational Medicine, 59*(8), 570–573. https://doi.org/10.1093/occmed/kqp141

Loerbroks, A., Schilling, O., Haxsen, V., Jarczok, M. N., Thayer, J. F., & Fischer, J. E. (2010). The fruits of one's labor: Effort-reward imbalance but not job strain is related to heart rate variability across the day in 35-44-year-old workers. *Journal of Psychosomatic Research, 69*(2), 151–159. https://doi.org/10.1016/j.jpsychores.2010.03.004

Lundberg, U., & Hellström, B. (2002). Workload and morning salivary cortisol in women. *Work and Stress, 16*(4), 356–363. https://doi.org/10.1080/0267837021000064427

Lupien, S., Roch Lecours, A., Lussier, I., Schwartz, G., Nair, N. P. V., & Meaney, M. J. (1994). Basal cortisol levels and cognitive deficits in human aging. *Journal of Neuroscience, 14*(5), 2893–2903. https://doi.org/10.1523/JNEUROSCI.14-05-02893.1994

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., Schwartz, P. J., & [Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology]. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*, 354–381. https://doi.org/10.1161/01.CIR.93.5.1043

Marchand, A., Juster, R.-P., Durand, P., & Lupien, S. J. (2016). Work stress models and diurnal cortisol variations: The SALVEO study. *Journal of Occupational Health Psychology, 21*(2), 182–193. https://doi.org/10.1037/a0039674

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 287–303). Routledge.

Meder, L., Dorsemagen, C., & Krause, A. (2008). Observational stress analysis at school: Classroom teaching as an example of interaction work. *Psychology of Everyday Activity, 1*(1), 23–32.
Michaud, K., Matheson, K., Kelly, O., & Anisman, H. (2008). Impact of stressors in a natural context on release of cortisol in healthy adult humans: A meta-analysis. *Stress, 11*(3), 177–197. https://doi.org/10.1080/10253890701727874
Miller, R., & Plessow, F. (2013). Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations. *Psychoneuroendocrinology, 38*(6), 941–946. https://doi.org/10.1016/j.psyneuen.2012.09.013
Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality, 21*(6), 789–810. https://doi.org/10.1002/per.640
Nikolova, I., Van Ruysseveldt, J., De Witte, H., & Syroit, J. (2014). Wellbeing in times of task restructuring: The buffering potential of workplace learning. *Work and Stress, 28*(3), 217–235. https://doi.org/10.1080/02678373.2014.92960
Nitsch, J. R. (1982). Handlungspsychologische Ansätze im Sport. In A. Thomas (Ed.), *Sportpsychologie. Ein Handbuch in Schlüsselbegriffen* (pp. 26–41). Urban and Schwarzenberg.
Oei, N. Y. L., Everaerd, W. T. A. M., Elzinga, B. M., Van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress, 9*(3), 133–141. https://doi.org/10.1080/10253890600965773
Panari, C., Guglielmi, D., Simbula, S., & Depolo, M. (2010). Can an opportunity to learn at work reduce stress? A revisitation of the job demand-control model. *Journal of Workplace Learning, 22*(3), 166–179. https://doi.org/10.1108/13665621011028611
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*(7), 916–931. https://doi.org/10.1016/S0306-4530(02)00108-7
Raemdonck, I., Gijbels, D., & Van Groen, W. (2014). The influence of job characteristics and self-directed learning orientation on workplace learning. *International Journal of Training and Development, 18*(3), 188–203. https://doi.org/10.1111/ijtd.12028
Rausch, A., Schley, T., & Warwas, J. (2015). Problem solving in everyday office work: A diary study on differences between experts and novices. *International Journal of Lifelong Education, 34*(4), 448–467. https://doi.org/10.1080/02601370.2015.1060023
Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training, 8*(9). https://doi.org/10.1186/s40461-016-0035-y
Rensing, L., Koch, M., Rippe, B., & Rippe, V. (2006). *Mensch im Stress: Psyche, Körper, Moleküle*. Elsevier.
Roberts, A. D., Wessely, S., Chalder, T., Papadopoulos, A., & Cleare, A. J. (2004). Salivary cortisol response to awakening in chronic fatigue syndrome. *British Journal of Psychiatry, 184*(2), 136–141. https://doi.org/10.1192/bjp.184.2.136
Sapolsky, R. M. (1996). Why stress is bad for your brain. *Science, 273*(5276), 749–750. https://doi.org/10.1126/science.273.5276.749
Sauro, M. D., Jorgensen, R. S., & Pedlow, C. T. (2003). Stress, glucocorticoids, and memory: A meta-analytic review. *Stress, 6*(4), 235–245. https://doi.org/10.1080/10253890310001616482
Saxbe, D. E. (2008). A field (researcher's) guide to cortisol: Tracking HPA axis functioning in everyday life. *Health Psychology Review, 2*(2), 163–190. https://doi.org/10.1080/17437190802530812

Schoofs, D., Preuß, D., & Wolf, O. T. (2008). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology, 33*(5), 643–653. https://doi.org/10.1016/j.psyneuen.2008.02.004

Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., Schley, T., Siegfried, C., Egloffstein, M., Küster, J., Wuttke, E., Sembill, D., Martens, T., & Wolf, K. D. (2016). Problemlösekompetenz angehender Industriekaufleute – Konzeption des Messinstruments und ausgewählte empirische Befunde (DomPL-IK). In K. Beck, M. Landenberger, & F. Oser (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* [Technology-based measurement of competencies in vocational education and training. Findings of the BMBF funding initiative ASCOT] (pp. 119–138). W. Bertelsmann Verlag.

Selye, H. (1973). The evolution of the stress concept. The originator of the concept traces its development from the discovery in 1936 of the alarm reaction to modern therapeutic applications of syntoxic and catatoxic hormones. *American Scientist, 61*(6), 692–699.

Sembill, D. (1992). *Problemlösefähigkeit, Handlungskompetenz und Emotionale Befindlichkeit: Zielgrößen Forschenden Lernens* [Problem-solving ability, action competence and emotional state: Target variables of research-based learning]. Hogrefe.

Sembill, D. (2012). Was bedeutet die Hirnforschung für die Schul- und Lernkultur? In J. Warwas, P. Harder, & D. Sembill (Eds.), *Kultur der Schule – Schule der Kultur(en)?* (pp. 85–112). Schneider Hohengehren.

Sembill, D. (2015). Berufliche Bildung in einer Lerngesellschaft. In A. Weber, L. Peschkes, & W. E. L. De Boer (Eds.), *Return to Work – Arbeit für alle: Grundlagen der beruflichen Reintegration* (pp. 83–93). Gentner.

Sembill, D., & Kärner, T. (2018). Bewertung und Ausbalancierung – Heuristiken für onto- und soziogenetische Schichtungsmodellierungen in der Bildungsforschung. In M. Huber & S. Krause (Eds.), *Bildung und Emotion* (pp. 169–194). Springer VS. https://doi.org/10.1007/978-3-658-18589-3_10

Sembill, D., & Kärner, T. (2020). Emotionen sind Macht. Analoge und digitale Rhythmen in der Fortschrittsdebatte der beruflichen Lehrpersonenbildung. In K. Heinrichs, K. Kögler, & Ch. Siegfried (Eds.), *Berufliches Lehren und Lernen: Grundlagen, Schwerpunkte und Impulse wirtschaftspädagogischer Forschung. Digitale Festschrift für Eveline Wuttke zum 60. Geburtstag. bwp@ Berufs- und Wirtschaftspädagogik – Online* (pp. 1–49). https://www.bwpat.de/profil6_wuttke/sembill_kaerner_profil6.pdf. Accessed 19 Oct 2020.

Sembill, D., Wolf, K. D., Wuttke, E., & Schumacher, L. (2002). Self-organized learning in vocational education: Foundation, implementation, and evaluation. In K. Beck (Ed.), *Teaching-learning processes in vocational education. Foundations of modern training programmes* (pp. 267–295). Lang.

Sembill, D., Seifried, J., & Dreyer, K. (2008). PDAs als Erhebungsinstrument in der beruflichen Lehr-Lern-Forschung – Ein neues Wundermittel oder bewährter Standard? Eine Replik auf Henning Pätzold [PDAs as a tool for data collection in adult education research – New panacea or approved standard? A reply to Henning Pätzold]. *Empirische Pädagogik, 22*(1), 64–77.

Sluiter, J. K., Frings-Dresen, M. H., Meijman, T. F., & Van der Beek, A. J. (2000). Reactivity and recovery from different types of work measured by catecholamines and cortisol: A systematic literature overview. *Occupational and Environmental Medicine, 57*(5), 298–315. https://doi.org/10.1136/oem.57.5.298

Stawski, R. S., Cichy, K. E., Piazza, J. R., & Almeida, D. M. (2013). Associations among daily stressors and salivary cortisol: Findings from the National Study of Daily Experiences. *Psychoneuroendocrinology, 38*(11), 2654–2665. https://doi.org/10.1016/j.psyneuen.2013.06.023

Stevenson, A., & Harper, S. (2006). Workplace stress and the student learning experience. *Quality Assurance in Education, 14*(2), 167–178. https://doi.org/10.1108/09684880610662042

Stroud, L. R., Papandonatos, G. D., Williamson, D. E., & Dahl, R. E. (2004). Applying a nonlinear regression model to characterize cortisol responses to corticotropin-releasing hormone challenge. *Annals of the New York Academy of Sciences, 1032*, 264–266. https://doi.org/10.1196/annals.1314.034

Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A., & Van Huffel, S. (2011). Instantaneous changes in heart rate regulation due to mental load in simulated office work. *European Journal of Applied Physiology, 111*(7), 1497–1505. https://doi.org/10.1007/s00421-010-1776-0

Takahashi, T., Ikeda, K., Ishikawa, M., Kitamura, N., Tsukasaki, T., Nakama, D., & Kameda, T. (2005). Anxiety, reactivity, and social stress-induced cortisol elevation in humans. *Neuroendocrinology Letters, 26*(4), 351–354.

Tenenbaum, G., Edmonds, W. A., & Eccles, D. W. (2008). Emotions, coping strategies, and performance: A conceptual framework for defining affect-related performance zones. *Military Psychology, 20*(1), 11–37. https://doi.org/10.1080/08995600701804772

Thomas, B. L., Claassen, N., Becker, P., & Viljoen, M. (2019). Validity of commonly used heart rate variability markers of autonomic nervous system function. *Neuropsychobiology, 78*(1), 14–26. https://doi.org/10.1159/000495519

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*(1), 11–36. https://doi.org/10.1007/s12186-012-9091-z

Uusitalo, A., Mets, T., Martinmäki, K., Mauno, S., Kinnunen, U., & Rusko, H. (2011). Heart rate variability related to effort at work. *Applied Ergonomics, 42*(6), 830–838. https://doi.org/10.1016/j.apergo.2011.01.005

Van Dick, R. (1999). *Stress und Arbeitszufriedenheit im Lehrerberuf: Eine Analyse von Belastung und Beanspruchung im Kontext sozialpsychologischer, klinisch-psychologischer und organisationspsychologischer Konzepte*. Tectum Verlag.

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: Implications for the classroom. *npj Science of Learning, 1*, 16011. https://doi.org/10.1038/npjscilearn.2016.11

Von Davier, A. A., & Halpin, P. F. (2013). Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series, 2013*(2), i–36. Research Report ETS RR-13-41. https://doi.org/10.1002/j.2333-8504.2013.tb02348.x

Vygotskii, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Tobias Kärner (PhD) is a full professor for Economic and Business Education at the University of Hohenheim, Stuttgart, Germany. His research has focused, among other things, on participation and learning in vocational schools and the multimodal assessment and evaluation of learning and working environments. His recent publications include articles in *Learning and Instruction*, *Frontline Learning Research* and *Empirical Research in Vocational Education and Training*.

Detlef Sembill (PhD) is a professor emeritus for Economic and Business Education at the University of Bamberg, Germany. His research has focused, among other things, on complex problem-solving skills, emotions and motivation, self-organised learning and stress in teaching–learning arrangements. His recent publications include articles in *Frontline Learning Research*, *Frontiers in Psychology* and *Vocations and Learning*.

Chapter 7
Combining Physiological and Experiential Measures to Study the Adult Learning Experience

Minna Silvennoinen, Tiina Parviainen, Anita Malinen, Suvi Karjalainen, Mari Manu, and Mikko Vesisenaho

Abstract This chapter introduces methods for using an individual-level multimodal approach for studying the learning experience within the context of vocational education. There has recently been increased interest in recording physiological signals in pedagogical contexts. The current research literature on multimodal studies of adult learning experience is scarce and has been primarily applied and developed in studies of a preliminary nature and with varying combinations of modalities. Learning experience is a complex phenomenon which cannot be fully captured via a single-data modality. However, based on the reviewed literature, there is still a lack of larger datasets and strong empirical evidence to enable a comprehensive understanding of experiential learning as a phenomenon.

In addition to self-reported learning experiences, there is a need for theoretical development and a more holistic empirical approach that includes physiological and neurophysiological aspects involved in learning situation. We present a case example of simulation-based learning (SBL) of forestry skills, in which the modalities applied to explore the learning experience were video recordings, stimulated recall interviews, questionnaires, electrocardiography (ECG), and electroencephalography (EEG). Our example presents how multimodal research design can be used to study learning experience, by combining measurements of the human nervous system with subjective and observational data. It is too early to evaluate the practical impact of multimodal research for the field of adult education. Successful application of multimodal methods requires interaction across disciplines, harmonizing of

Authors Minna Silvennoinen and Tiina Parviainen have equally contributed to this chapter.

M. Silvennoinen (✉)
JAMK University of Applied Sciences, Jyväskylä, Finland
e-mail: minna.silvennoinen@jamk.fi

T. Parviainen · A. Malinen · S. Karjalainen · M. Manu · M. Vesisenaho
University of Jyväskylä, Jyväskylä, Finland
e-mail: tiina.m.parviainen@jyu.fi; anita.malinen@jyu.fi; suvi.k.karjalainen@jyu.fi; mari.j.manu@jyu.fi; mikko.vesisenaho@jyu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_7

conceptual frameworks and goals, as well as bringing together complementary, discipline-specific expertise to guarantee valid application of research methods. Opening of the disciplinary boundaries both at theoretical and methodological domains enables to increase discussion between researchers from different disciplines.

Keywords Multimodal measurements · Adult learning · Experiential learning · Combined methods · Research design · Physiological measurement · Heart rate variability · Electroencephalography · EEG · Learning experience · Vocational education

7.1 Introduction

Traditionally, adult learning is considered to be a primarily cognitive process that can be studied with various quantitative or qualitative methods. However, research emphasis has recently shifted to a more holistic approach, whereby an individual is perceived as a whole, consisting of both mind and body (Papastamatis & Panitsides, 2014). Within this perspective, the cognitive, emotional, and also physical (the so-called embodied aspects) need to be considered in learning and reflection, with consequences for the research of learning experience. Closer integration of different methodological approaches capturing the cognitive, emotional, and embodied reflections of behavior has the potential to critically increase our understanding of the adult learning experience and learning interaction. Learning has been approached from different theoretical and scientific traditions that, to a large extent, have been developed independently from one another, for example, in the fields of education and psychology. Within these different scientific disciplines, there are also divergent methodological approaches to studying learning, which can be roughly categorized according to three different perspectives. Methods focusing on the self-reported experience of the learner, such as reflective assignments, questionnaires, or interviews, can capture the subjective first-person dimension of the experience (see experiential measures, e.g., Lumma & Weger, 2021). Additionally, learning is often studied by collecting quantified variables (e.g., reaction times or number of errors) or qualitative descriptions and interpretations of performance or experience by an external observer. These data offer a second-person perspective on experience. Finally, neuroscience adds a new, third dimension in the exploration of the learning experience, as the recordings of the nervous system activity (i.e., quantitative third-person measurements that accompany the learning situation) form the biological basis of the various aspects within the learning experience (e.g., emotions, the level of effort or motivational state, and acquisition of new skills through experience; Ansari et al., 2011).

It is important to note that although we can collect these different data types, we have no straightforward way to directly map these different measurements with each other, and they need to be (and often are) treated as separate data approachable within distinct scientific paradigms (for a discussion of the case of neurophenomenology, see Varela & Shear, 1999). Therefore, neurophysiological measurements do not directly provide information about the experience itself but are rather correlates (i.e., external recordings of the underlying neural-level events) that give rise to the observed and experienced learning. These variables can, however, offer additional and complementary information to reach a more holistic understanding of learning and the learning experience.

Modern research technology allows for detailed monitoring of behavior outside laboratories in natural learning environments. Novel and fairly low-cost tools to record physiological and even neurophysiological signals in more natural settings have significantly increased the interest in recording these signals in pedagogical contexts (see van Atteveldt et al., 2018). In order to provide real added value, physiological measurements need to be sensibly linked to other indicators of the learning experience. However, this kind of research is challenged by a lack of methodological development in study design, data collection, and data analysis. Methodological approaches have not been developed to the same extent as the technological possibilities. To progress our understanding of the learning experience, research strategies need to combine laboratory and naturalistic research efforts (see Wilhelm & Grossman, 2010; Wilhelm et al., 2012). The combinations of quantitative and qualitative research designs, as well as different measurement technologies to explore learning and learning experiences (especially in natural contexts), are still fairly rare.

In this chapter, we introduce some new directions for using multimodal methods to study the adult learning experience in a natural context. We use a case example from a vocational education context of simulation-based learning (SBL) of forestry skills.

Before further discussing the different methodological ways to approach learning and specifically the learning experience, we must first define some key concepts used in this chapter. First, by the *learning experience*, we refer to a specific combination of elements that together constitute the experience of learning. Learning experience, therefore, forms in the short and long term as an outcome of processes that govern physiological, cognitive-psychological, and social elements that co-occur in learning situations. These elements are discernible by verbal or non-verbal expressions and recordable biophysical signals, for example, visible facial expressions and gestures as well as neurophysiological and autonomic nervous system (ANS) recordings.

Second, various terminology is used to refer to research approaches and designs that combine methodologies from different disciplines. The multi-method approach in education sciences, similar to mixed methods more generally (see Johnson & Onwuegbuzie, 2004; Patton, 2002), typically refers to combining qualitative and quantitative methods (Hammond, 2005) or applying two or more sources of data to investigate the same phenomenon (Brever & Hunter, 1989). The mixing/combining of approaches can take place at different stages of analysis, and in most studies, the qualitative and quantitative data are analyzed separately and combined only at the interpretation stage. In some studies, combining physiological and self-reported data (see, e.g., Eteläpelto et al., 2018; Harley et al., 2015) is referred to as multicomponential research. We have chosen to use the term *multimodal* research to reflect the application of a combination of modalities, namely experiential data (first person), observations or interpretations of behavior (second person), and physiological data (third person), in data collection and analysis. This concept has been used when more than one data collection method was needed to measure learning-related experiences or emotions (see Giannakos et al., 2019; Harley et al., 2015). In our case example—SBL of forestry skills—the modalities applied were video recordings, stimulated recall interviews, questionnaires, electrocardiography (ECG), and electroencephalography (EEG).

In the first part of this chapter, we review current literature on multimodal studies of learning experience, including those in the SBL context. Then, we present our SBL research case, within which we combined a variety of methods and disciplinary expertise to study learning experience. The chapter ends with a discussion and suggestions for future directions concerning theoretical and methodological developments in researching the adult learning experience.

7.2 Adult Learning Research: From Traditional to Multimodal Combinations of Methods in Researching Learning Experience

In the following section, we briefly summarize the traditional technique of combining methods for researching the adult learning experience and, subsequently, review the less familiar multimodal techniques for combining methods from these traditional disciplines with various other scientific fields, such as neuroscience.

7.2.1 Combining Methods to Explore Learning Experience

There has been growing interest in combining research approaches and methods in education, especially since the beginning of the previous decade (Gorard & Taylor, 2004). In education sciences, combining qualitative and quantitative approaches has a relatively short history compared to mono-method approaches. Appropriate combination and complementary utilization of methods provide increased value for both scientific and societal impact, resulting in stronger research arguments about the complex phenomenon in question (Gorard, 2002; Gorard & Taylor, 2004). Most of the studies researching the learning experience have relied on traditional qualitative approaches from education and social sciences (e.g., Daley et al., 2018; Gorard & Taylor, 2004), such as interviews and reflective writing. Of the quantitative research methods, questionnaires have mainly been used.

Thus far, few studies of the learning experience conducted in the educational context have complemented traditional educational research methods with recordings of body and brain signals. Brain information processing naturally provides the basis for learning, but this is primarily only possible in a laboratory context. The embodied learning view introduces concepts that are both attainable for more naturalistic, multimodal empirical research and also relevant for adult education scientists (Dirkx, 2008; Jordi, 2010). Embodied learning arises from the larger embodied cognition framework, which emphasizes the role of the entire organism (i.e., human body) and its immediate interaction with the environment, which can be used for studying cognitive functions (e.g., learning, experience, and emotions). From this point of view, learning as an experience is not only a rational analytical process but essentially involves the body as a producer of information related to learning and a tool for reflection (Dirkx, 2008; Fenwick, 2006; Jordi, 2010). Emotions are a good example of experiences (i.e., phenomena of the mind) that are intimately associated with physiological reactions (i.e., phenomena of the body; see Damasio, 1999; Hannaford, 1995). For example, memory formation is dependent on sensory (and motor) functions that ultimately bridge our brain with the contextual richness of experiential learning situations. Morris (2020) emphasized the importance of embodiment in the experiential learning process.

What kind of methodology is needed to understand the holistic nature of adult experiential learning? Daley et al. (2018) stated that "complex research questions are addressed by teams of researchers who understand and apply team science using quantitative, qualitative, and mixed methods" (p. 164). It is intuitively clear that adult learning poses complex research questions that cannot be comprehensively understood only by a single discipline. The challenge in designing this research is that it requires combining methodology with theories and concepts from different disciplines, such as education sciences, psychology, and cognitive neuroscience.

7.2.2 *Multimodal Research of the Adult Learning Experience*

We reviewed studies published between 2005 and 2021 that combined multimodal measurement techniques for studying adult learning-related experiences. We found 16 empirical research articles with varying modalities and learning contexts. The common ground for all studies was an interest in the learner experience and understanding this experience, in part by utilizing physiological recordings. In addition, we included a few papers that reported multimodal empirical studies without physiological recordings as well as a few that were more oriented toward methodology and technology. In addition, some purely theoretical papers on multimodal characterization of the learning experience are included in the literature review. We provide a brief overview of the literature found with the goal of introducing the current state of the topic rather than providing a deep analysis of the quality or validity of the findings.

7.2.2.1 Multimodal Studies on Learning Experience and Emotions

A holistic approach to learning (involving the whole body) must understand the role of emotions and affect as a natural and important part of the (reflective) learning experience (Dirkx, 2008; Jarvis, 2006). Researchers have shown increased interest in the contribution of emotions to learning (e.g., Calvo & D'Mello, 2011; Pekrun & Schutz, 2007; Picard et al., 2004). Indeed, the use of traditional self-reports are considered somewhat limited and unreliable in their scope, and therefore, various additional signals (face, voice, body, and learning environment) as well as a combination of quantitative and qualitative methods have been suggested for more comprehensive exploration of emotions related to learning (see Bahreini et al., 2016; Zembylas, 2007). Emotional reactions are the most likely source of variation in the physiological responses of the learner; for example, HRV and electrodermal activity (EDA; sometimes referred as skin conductance response/level) are physiological changes essentially correlated with emotions.

Researchers working with e-learning environments and learning applications have found that computerized learning environments (adaptive learning systems with sensors) could be improved by using affective or emotion-related information (e.g., Calvo & D'Mello, 2011; Fortenbacher & Yun, 2020). Multimodal physiological measurement settings have been developed for exploring the learning experience with the aim of customizing learning materials and improving the performance of e-learning environments based on students' emotional states (for a single-subject pilot, see Shen et al., 2009) or supporting learning applications with real-time biofeedback systems through wearable physiological sensors (see Schaaff et al., 2012). Furthermore, Hardy et al. (2013) tested the potential of EDA for providing additional insight into learners' affective and cognitive states in the context of learning programming skills.

It is, however, important to remember that an individual's own description is still the only way to capture an emotional experience, and the connection between experiential and physiological variables is still unclear. Eteläpelto et al. (2018) conducted a preliminary study aimed at understanding how self-reported and ANS indicators of emotions are related to each other in the context of professional learning. Hemingway et al. (2019) associated modulations of EDA during equine-assisted intervention with the significance of learning episodes (based on experiential interview data), but the data were only analyzed qualitatively. It is not always apparent that experiential measures of learners' emotional states co-vary with ANS-based assessments. Harley et al. (2015) showed that during complex learning situations, automatic facial recognition of emotions and self-report data were in agreement, but EDA was not correlated with these measures. Besides ANS recordings, one multimodal approach used cortisol measurements to explore students' emotions and coping in exam situations (Spangler et al., 2002).

It is known that specific affective states and emotions are related to learning experience (Pekrun & Schutz, 2007; Staus & Falk, 2017; Woolf et al., 2009). However, few studies on learners' emotional experiences have used multimodal recordings in real time or in informal learning contexts outside the classroom (see

LeBlanc, 2019; Staus & Falk 2017). Most of the aforementioned empirical studies were preliminary in nature, focusing on learners' emotional reactions and associated physiological reactions. Earlier studies also indicated a need for larger-scale research to multimodally explore the relationship between physiological measures and subjective, emotional experiences (Hardy et al., 2013; LeBlanc, 2019; Shen et al., 2009).

7.2.2.2 Multimodal Studies on Learning Experience and Cognitive Processes

Combined behavioral and body physiology measures have been used to explore metacognitive monitoring in the context of learning and cognitive behaviors, for example, in multimedia learning situations (Antonietti et al., 2015; Mudrick et al., 2019). Antonietti et al. (2015) recorded multimodal data within memory and problem-solving tasks with the aim of predicting correct responses according to the level of task-related cognitive effort as indicated by eye movement and EDA data. This study was not directly aimed at exploring learning experience but rather students' strategies during learning. The eye movement data were suggested to be useful in revealing participants' metacognitive monitoring (Antonietti et al., 2015). Similarly, Mundrick et al.'s (2019) study aimed at addressing how to identify behavior during meta-comprehension processes.

In a multimodal study on cognitive load during learning, Larmuseau et al. (2019) investigated the link between students' EDA and task-related self-reported cognitive load. They correlated physiological signals (EDA and skin temperature) recorded during a problem-solving task. Their preliminary results indicated no connection between the physiological data and task complexity, but a connection was found between EDA and mental effort during highly complex tasks. They thus suggested using EDA in combination with specific learning events for online determination of task-related cognitive load. Giannakos et al. (2019) explored the applicability of physiological recordings (via eye tracking, EEG, and video recordings) for understanding and supporting the learning experience when designing new learning technologies. Wang and Cesar (2015) explored the potential of physiological indicators (via GSR, video recording, and questionnaires) to provide feedback to teachers who teach in e-learning or distributed learning environments. Methodological combinations utilizing physiological data in relation to self-regulation and monitoring behavior have shown promise for developing tools for immediate learning research and support in collaborative settings (e.g., Dindar et al., 2020; Haataja et al., 2018).

To sum up, the authors of the aforementioned studies found it useful to combine multiple data modalities (e.g., facial expression, eye movement, learner experiences, and EDA) to better understand learning behavior, providing the potential to shape future learning technologies to support both individual and collaborative learning processes. As with the studies on emotional experiences, these studies were exploratory in nature, and further studies with larger sample sizes and hypothesis-driven designs are needed to strengthen the validity of these findings.

7.2.2.3 The Role of Multimodality in Studying SBL Experience

SBL is a widely used and rather well-researched experiential learning method in healthcare (Husebø et al., 2015; Poore et al., 2014; Rogers et al., 2019) as well as other safety-critical contexts or contexts requiring repetitive training. SBL provides an ideal environment to perform multimodal studies of learning experience, but empirical studies of this kind are still scarce. There has been recent interest in using physiological recordings to study stress associated with SBL in different contexts (see LeBlanc, 2019). For example, experienced stress and HR were measured to determine the level of workload during SBL and specifically its relation with self-perceived learning (Girzadas et al., 2009). Additionally, experienced stress and HR were measured to compare the level of stress in real patient encounters vs. SBL situations (Judd et al., 2016). Furthermore, Kocialkowski et al. (2020) used an HR-based evaluation of stress response in their study to clarify the effectiveness of static vs. dynamic observational learning scenarios in medical students. Bhoja et al. (2020) reviewed and piloted the potential of using EDA and HRV measuring devices and video recordings during SBL situations to evaluate the level of stress and alertness for optimal learning in SBL. Despite the lack of physiological measures, Rogers et al. (2019) emphasized the importance of exploring emotional experiences and their relationship to learning within SBL scenarios (see also LeBlanc, 2019).

There have been some initiatives to utilize combined physiological and experiential data outside the stress domain, for example, in the context of technology-enhanced learning and so-called serious game development. By combining data from physiological recordings (EEG and ECG) with self-reported data (questionnaires), Cowley et al. (2013) showed that HRV predicts learning effects in the serious game context. Simulated learning environments applying virtual reality (VR) or mixed reality technology have been shown to create powerful experiences (Vesisenaho et al., 2019), even though they are rarely designed for the process of experiential learning (Fromm et al., 2021; Birt et al., 2018). In their multimodal research, Aguayo et al. (2018) discussed the use of self-reported and biometric feedback data to design meaningful learning experiences in VR. While not directly related to learning experience, VR environments have been studied using multimodal methods to improve technologies and provide more immersive experiences (e.g., Marin-Morales et al., 2018; Stepanova et al., 2019). Thus, a more comprehensive understanding of learning experience via multimodal data could aid in creating SBL to better prepare learners for future professional practice. The presence of stress in particular should be acknowledged by instructors implementing SBL. Additionally, the physiological reactions in the SBL environment could provide an ideal basis for developing more reflective approaches to learning and for preparing for future performance in terms of stress/emotion regulation.

To conclude, the literature reviewed shows that recent developments in sensor technology and analysis algorithms have enabled the recording of learners' physiological signals along with their varying emotional and cognitive states. The suitable paradigms and reliable measures in these different modalities still need to be established to determine the real benefit of this type of research. The current evidence is still very narrow, with varying and often modest sample sizes. The analysis of the physiological data is often limited, and various methods are used to collect data on learning situations. Previously published studies have varied in both theme and impact, ranging from high-level journals to non-peer-reviewed conference papers.

Regarding the scientific fields of the previously reviewed literature, it is noteworthy that the research has been carried out mainly by representatives of disciplines other than education science or neuroscience. Information and communication technology, psychology, and combined disciplines—such as educational technology or human—computer interaction with SBL-related studies in the fields of medical or health care education—are the most represented. Our research team combined educational sciences, psychology, neuroscience, signal processing, and cognitive science to produce a comprehensive perspective for the empirical understanding of experiential learning.

7.2.3 Building a Theoretical Framework for the Multimodal Study of Learning

While studies have indicated the power of experience in learning (e.g., Illeris, 2018; Jarvis, 2005a, b, p. 1), the complex and diverse processes of human learning are not well enough understood to produce a single comprehensive theory (Illeris, 2018; Jarvis, 2006; Malinen, 2000; Yang, 2006). In this chapter, we approach learning as a complex and multifaceted phenomenon aligning with the fundamental questions asked by Illeris (2018): How does learning take place in the human brain and body? What are the supposed mechanisms of experiential learning? Additionally, we explore the question asked by Jarvis (2006, p. 198): Is it possible to have a comprehensive theory of learning? Recent research has provided refreshing and valuable empirical approaches and new theoretical considerations to answer these questions. Experiential learning theory currently provides the most holistic conceptualization of (adult) learning.

Theory building attempts concerning adult experiential learning need to be examined more thoroughly. Few authors have combined the concepts of neuroscience with experiential learning or experiential education. Kolb and Kolb (2017) examined the contributions of neuroscience research to our understanding of experiential learning. They combined learning cycles and the brain, especially memory functions, within the experiential learning processes. Theoretically, they integrated the constructivist and embodied cognition perspectives into one balanced learning process (Kolb & Kolb, 2017, pp. 78–79). They argued, however, that "the promise of an evidence-based practice of neuroscience education currently offers more provocative possibilities than proven practices" (Kolb & Kolb, 2017, p. 56).

Schenck and Cruickshank (2015) re-conceptualized experiential learning theory by building on cognitive neuroscience knowledge. They developed a biologically driven model of teaching—co-constructed developmental teaching theory—and argued for neurobiology as the foundation for experiential learning. With a focus on the neurobiological basis of learning, they emphasized the need for empirical testing: "We need to better inform ourselves about the mind-brain processes that affect experiential learning" (Schenck & Cruickshank, 2015, p. 90).

Hagen and Park (2016) aimed to bridge the gap between adult learning theory and more recent scientific human resource development research and cognitive neuroscience. They argued that the four core assumptions of andragogy have connection to the neural networks related to memory and cognition. They suggested a framework on how neuroscience knowledge can be used to understand prior experience in a learning process, problem-based learning, and experiential learning approaches, and they created a model of an adaptive cognitive neuroscience adult learning structure based on four andragogy assumptions: (1) self-directed learning, (2) experience-based learning, (3) adults' readiness for learning, and (4) application-focused learning (see also Lim et al., 2019). They also indicated that instructional techniques guided by andragogy can improve long-term memory and retention. They also suggested exploring the relationship of emotion and adult learning with neuroscientific methods.

To sum up, the current discussion around experiential learning theory clearly warrants efforts to develop a more holistic framework that can guide further experimental work at the intersection between educational sciences and cognitive neuroscience, especially regarding human physiology and neurophysiology. This development also needs to be informed by empirical evidence and research conducted more holistically on the process of learning, interfacing with research findings on the physiological and neurobiological basis of learning (Silvennoinen et al., 2020). In the next section, we introduce a case example of a multimodal study conducted in a vocational education context, SBL of forestry skills, that approached the learning experience in a holistic manner.

7.3 Case: Multimodal Study of the Forestry SBL Experience

Our literature review indicated the need for integration of the research methodologies and frameworks of different disciplines to understand learning experience. Our case example from a vocational education context introduces an interdisciplinary research project utilizing expertise and methods of educational sciences, psychology, neuroscience, signal processing, and cognitive science. In order to understand determinants and elements of learning experience in different contexts, we acquired reports of subjective learning experience and teacher observations as well as physiological and neurophysiological signals associated with ongoing learning experience. This study therefore provides an example of multimodal research methods and the experiential learning environment that can be used to answer broad questions in the field of the adult learning experience. Next, we introduce the case, the applied methodology, data collection techniques, and some relevant points for conducting analyses of multimodal learning experience data.

7.3.1 Context and Participants

The study participants were recruited from a vocational school at which students become qualified for forest-based energy production. Participants in this case were student—instructor dyads with six students and two instructors total. By the time of data collection in 2020, the students had already gained experience in the use of forestry simulators. SBL is an essential part of vocational studies, particularly when students begin to familiarize themselves with different forest machines. Pedagogically, SBL offers excellent possibilities in vocational education studies embodying strong elements of experiential learning (see, e.g., Clapper, 2014; Rogers et al., 2019). Each learning situation imitates an authentic environment, and learning tasks can vary according to the instructions (performed either by the simulator or instructor/teacher). Compared to traditional learning, such as lectures, simulations can create powerful experiences for learners due to their authentic connection to the emotions (Fromm et al., 2021) and reflections they stimulate and which are also debriefed (Bearman et al., 2019; Husebø et al., 2015; Lateef, 2010). From a research viewpoint, computer-based simulations offer optimally controlled situations to collect different kinds of data during the ongoing learning experience. Simulations combined with VR technology allow for data collection in a setting that is as close to natural and authentic as possible while still maintaining some control over the research design, providing maximal practical impact. However, learning in natural contexts is influenced by various uncontrolled variables, which poses a challenge for data collection and analysis.

7.3.2 Research Data Collection

We used video recordings, instructor observations (written notes), participating students' stimulated recall interviews and written notes, and interviews with teachers. For the autonomic and central nervous system reactions and ongoing signaling in the learning experience, we also recorded respiration using a respiratory belt HR and, HRV using ECG and ongoing brain activity using EEG.

7.3.2.1 Protocol

EEG, ECG, video recordings, and interviews were conducted simultaneously. The measurement period lasted on average 3 h for each dyad, including preparations. The student performed the task with a VR headset, which provided a close to realistic impression of being positioned in the cab of a forestry machine with a three-dimensional view of the control system and a virtual forest. The instructor sat beside the student and followed their performance (Fig. 7.1). After all tasks were completed, the instructor used screen-recorded videos recorded during the task performance to give feedback to the student. Two video cameras recorded the student's performance on the simulator screen as well as the interaction between the student and the instructor. Discussions between the student and instructor were audio recorded.

The SBL consisted of three phases: an introduction to the training task, the action (performing the training task within the simulator), and a debriefing discussion. During the introduction phase, the student and the instructor went through the structure and general instructions for the four forthcoming tasks. During the action phase, the student and the instructor watched the simulator's model video (one for each separate task) to gain an understanding of the optimal task performance, after which the student performed the task. The task difficulty progressively increased, and the students had little to no experience with the actions in the final task. Along with the tasks, the instructor kept notes on the student's performance, which were later included in the data analysis as supplementary information on the SBL process and student performance. After each task, the student and the instructor briefly reviewed the task and the numerical performance levels offered by the simulator.

Fig. 7.1 Illustration of the research situation while testing the equipment. The student sat on the right (operating the simulator) and the instructor on the left, and both wore the EEG caps

After the four tasks, an in-depth debriefing discussion supported by the screen-recorded videos was done to gain a joint understanding of the overall SBL situation. Each student participated in a videorecorded stimulated recall interview (see details in Sect. 7.3.2.2) either on the same day or the following day after the SBL situation.

7.3.2.2 Data Collection Methods

We used *video recordings* to gather detailed information on the timeline and events during the simulation and to enable further investigation of different modalities of data related to specific events. In addition, we used the videos to analyze SBL situations (e.g., behavioral analysis of student—instructor interaction) and in the stimulated recall interviews, where the video material was examined and annotated by each student. Video recording is a frequently used method in learning science research and are particularly beneficial for data collection in complex learning environments (Derry et al., 2010), such as simulations involving rapidly changing situations and interactions between both technology and humans as well as learners and instructors. Video recording offers a method of collecting, sharing, studying, presenting, and archiving learning-related cases to support teaching, learning, and intensive study of those practices. Generally, video recordings can (1) provide detailed data of the timeline of events during learning, enabling combinations (annotations) of various other datasets or events; (2) enable behavioral analyses of occurrences during interaction and learning situations; and (3) aid the reflection process when applied as a stimulating material for interviews (self-reported data of the learning experience).

Each student was *interviewed* individually after the SBL situation on the same or following day via a stimulated recall method. First, the student and the interviewer discussed the training tasks performed in the SBL situation. Second, the student and the interviewer watched screenshot recordings of the training tasks, and the student annotated episodes that they felt were meaningful. Students also wrote down notes relating to the episodes. The instructors were interviewed individually in order to discuss their conceptions and opinions about the SBL situation along with their pedagogical thinking concerning SBL in general. We also used standardized *questionnaires* for the students to acquire information about their temperament characteristics (adult temperament questionnaire, 77-item short form; Evans & Rothbart, 2007) and motivation (self-regulation questionnaire—learner; Black & Deci, 2000; Salmi et al., 2020) as well as non-standardized inquiries about the simulation training experience.

Interviews are a key source of experiential information on learning situations and often the most useful method for understanding experiences, opinions, values, etc. (Rowley, 2012). As another option, questionnaires may allow for faster collection of more precise or specific information, but they are always more formally structured and have limited space for participants to express information. These two data gathering methods can, however, complement each other. In *stimulated recall,* a video (or audio) recording is played to the participant to stimulate recollection of learning situation-related events in order to capture more accurate data explicated by the learner (Calderhead, 1981; Kagan et al., 1963). The stimulated recall method is a kind of think aloud technique, which has value especially in exploring cognitive strategies and learning processes (Lyle, 2003).

In recent years, the integrated efforts of neuroscience and educational sciences (i.e education neuroscience) have started to explore the brain basis of learning in increasingly natural settings (e.g., Ahonen et al., 2018; Dikker et al., 2017, van Atteveldt et al., 2018). In laboratory environments, the brain basis of learning and learning experience can be studied by several different types of brain research methods, such as brain structure via magnetic resonance imaging or electric signaling within networks of neurons via magnetoencephalography or EEG. In order to measure brain signaling during authentic learning experiences and interactions, the measurement equipment needs to be brought into natural learning environments. EEG is the most well-suited for this purpose, but it is only in the past 10 years that technological advances have enabled the measurement of reasonably good EEG signals from the brain during natural situations. In the forestry SBL case, we studied the ANS activity by recording *HR, HRV, and respiration*. Two HR measurement techniques were applied. First, continuous HRV measurement using a detection device (Firstbeat Technologies Ltd., Jyväskylä, Finland) was conducted for 3–5 days, including the day of the SBL. This measurement provided extensive information about the participants' baseline HRV over a long period of time (Firstbeat Technologies Ltd, 2014). Second, HR was recorded by ECG with the Bittium NeurOne system (Bittium Biosignals Ltd., Kuopio, Finland) during the SBL situation. A flexible respiratory belt (Spes Medica, Genova, Italy) recorded the frequency and phases of respiration.

HRV is the temporal variation between successive heart beats reflecting the involvement of the parasympathetic division of the ANS and can thus be used to evaluate intra-subject levels of arousal, stress, and recovery. EDA provides a direct measure of the activity of sympathetic division. Despite their apparent simplicity, the use of ANS measures requires good understanding of both the physiological basis and the computations used for extracting reliable measures of ANS activity. Respiration is closely linked with and mainly controlled by ANS. However, unlike heartbeat, breathing can also be controlled voluntarily. Furthermore, the phase of respiration (i.e., inspiration vs. expiration) influences the timing of cardiac signals. This synchrony between respiration and HRV is defined as respiratory sinus arrhythmia (Shaffer & Ginsberg, 2017). Respiratory variables can be measured using an accelerometer or a respiratory belt. *Brain signals* were simultaneously recorded for the student-instructor dyad (Bittium Biosignals Ltd., Kuopio, Finland) using a 64-channel standard EEG cap (EASYCAP, BrainProducts GmbH, Gilching, Germany) for instructors and a 13-channel EEG cap (Neoprene Headcap with NG Geltrode electrodes and press stud cables, Neuroelectrics, Barcelona, Spain) adapted for the VR headset for the students. Prior to the SBL session, EEG caps, along with the ECG electrodes and respiratory belt, were individually fitted and prepared for the measurement and monitoring of autonomic and central nervous system activity. Synchronizing the EEG, ECG, and respiratory signals together and with the experiential measures was crucial for later integration of different data types, and thus a timestamp annotation was added to both EEG and video recordings. Also, other relevant annotations were added, such as when the SBL tasks began and when the feedback stage occurred.

EEG is a non-invasive method for recording electrical activity of the brain with high temporal and adequate spatial resolution. It is based on monitoring the voltage differences between the electrodes placed on the surface of the scalp. Although typically used in laboratories, it is possible to conduct EEG measurements in naturalistic settings. The most robust element in EEG recording is the rhythmic activation occurring at a frequency of ~10 Hz (i.e., one cycle about 10 times/second), called alpha oscillation or alpha rhythm. Since alpha rhythm has been associated with arousal, attention allocation, and task engagement, it is a reasonable measure of brain activation for the study of learning experience. However, there are numerous other measures that can be calculated from the EEG signal that may be useful for studying the learning experience.

7.3.3 Data Analysis

The combination of methods described above, including interviews, questionnaires, learners' and instructors' notes, and physiological and neurophysiological measures, enabled us to explore the manifold nature of experiential learning. A multimodal study, however, requires tackling methodological challenges separately for each modality first and then integrating the data of the different modalities.

7.3.3.1 Analysis of Interview Data

We had two aims for the qualitative analysis of the interview data. First, we wanted to explore participants' experience of SBL through data-driven analysis, and second, we wanted to construct the course of the events as a temporal continuum.

The student interviews were first transcribed and placed into a table format as a linear continuum of events. It was important to maintain the order of the SBL event description from the perspective of the participants. The interview questions followed the SBL structure. Qualitative data, such as interview transcripts, are textual, non-numerical, and thus unstructured. Coding is probably one of the most sensible ways to organize and make sense of interview data. Coding allows researchers to communicate and connect with the data to facilitate the comprehension of the emerging phenomenon and to generate theory grounded in the data (Basit, 2003). Codes can be certain types of experiences, feelings, or emotions or more content-structured, such as task-related events. In our research process, the interview texts were first carefully scrutinized by one researcher to gather all expressed utterances into a table format. Each utterance was then supplemented with a short clarification and the phase of the SBL situation, for example: "The student describes success in a task. / Action, Task 1." Then, these entities were given a code that described them in terms of content. Two researchers independently generated the codes for the content and then jointly discussed each code to reach common agreement. Researcher triangulation is a recommended way to increase the credibility of data interpretations. Another way to increase credibility of the analysis is to consult professional experts in the vocational field. In this case, a forestry teacher could be consulted in cases where the interviewees used professional vocabulary and terms that were unfamiliar to the researchers. Thematic entities that emerged as a result of the coding process described the main elements of the SBL experience.

7.3.3.2 Preparing Interview Data for Multimodal Integration

For the purpose of multimodal integration of the research data, it is necessary to maintain the temporal dimension of the learning experience. We started by creating a timeline of the interview data (transcribed text) and continued with generating data-driven codes for participants' expressed utterances through researcher triangulation.

At this stage of the analysis, we used video data and participants' written notes to construct the timeline of the events. Some of the selected thematic entities, such as successes or failures elaborated by the interviewee, were placed in a temporal time continuum based on the video data and the time-stamped written notes. Thus, the temporal structure of the SBL experience was constructed for these thematic entities, such as task variation, successes, failures, and aspects of student—instructor interactions. The video recordings enabled us to review the authentic situation. Video data were utilized to check the accuracy of the timeline in relation to SBL events, such as guidance or interaction situations.

7.3.3.3 Analysis of Physiological and Neurophysiological Data

EEG data analysis consists of two main stages: preprocessing and the extraction of the signal of interest. Natural research environments cause specific challenges and potential caveats for both of these stages. First, the quality of the recordings in natural learning environments is usually poorer than in laboratory settings (van Atteveldt et al., 2018) since the recordings are conducted without any shielding against electromagnetic interference. Natural movements and interactions between the student and the instructor can also weaken the quality. Besides clearly recognizable artifact types, such as movement or eye blinks, there are many other sources of artifacts in the EEG recordings that can be easily mixed with those caused by neural activity. Therefore, advanced methods and experience in EEG data processing are needed to gain reliable results. Recent state-of-the-art analysis methods allow for precise pre-processing of the data to reduce these artifacts.

Second, no detectable features in the signals can be directly associated with a particular experience, let alone learning. However, some well-studied characteristics can be extracted from the physiological measurements, and to some extent also from neurophysiological recordings (see above), which are linked to particular states, such as arousal, vigilance, and stress (Berntson et al., 1997; Quintana et al., 2012).

Third, one of the core problems in analyzing and interpreting the neurophysiological data in complex learning situations is the continuous nature of the task, unlike in controlled experimental designs, where specifically defined stimuli or tasks are used. This requires careful design of the data analysis stage.

In our case example, we focused on extracting reliable, artifact-free electrophysiological signatures that reflect both the state and reactivity of the central nervous system and ANS during different phases of the SBL situation. The EEG data were first visually inspected, and electrodes with poor signal quality were excluded. Physiological artifacts such as ocular and cardiac activity were removed using independent component analysis. After filtering and re-referencing, alpha activity was extracted via fast Fourier transform, which converts the data from the time domain to the frequency domain. Heart rate-related data were preprocessed and analyzed using Kubios HRV software (Biosignal Analysis and Medical Imaging Group, University of Eastern Finland, Kuopio, Finland), which provides standard and validated protocols for calculating measures of HR and HRV (Tarvainen et al., 2014). Both time-domain (e.g., mean, minimum, and maximum HR; mean and standard deviation of the time interval between successive heart beats) and frequency-domain (e.g., high and low frequency components of HRV) measures of HR and HRV were calculated and extracted from the HR-related data.

After preprocessing the data, two analysis pipelines were designed to obtain central nervous system and ANS measures that are associated with learning experience (Fig. 7.2). First, based on the analysis of the video recordings, different states (e.g., rest, simulation task, and feedback) were timestamped to the neurophysiological and physiological data. This was supported by a well-defined structure for the SBL situation, reflecting behaviorally and pedagogically distinct phases. A rudimentary timeline was created, providing an anchor for integrating the qualitative and quantitative data for the *state-based analysis*. The purpose of the state-based analysis was to enable us to focus on physiological and neurophysiological characteristics in distinct behavioral states (i.e., time periods that were defined by the experimental protocol). Both EEG measures and HRV were investigated in these specific states at

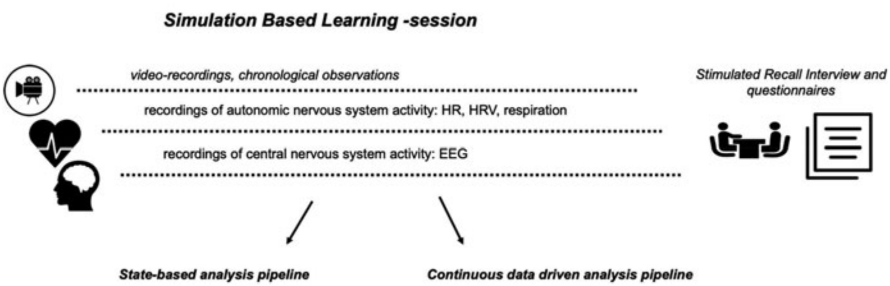

Fig. 7.2 Data collection methods

both the individual subject level and as group averages. Comparison of these different behavioral states can be used to extract information about the engagement of nervous system resources for attention allocation during cognitively different task requirements. Furthermore, if reliable measures can be verified, it is possible to examine how the nervous system indices of attention engagement during pedagogically different stages in SBL align with the observed or experienced indices of learning.

While the first analysis pipeline was based on the experimental conditions (rest, instruction, task performance, and feedback), the second *continuous data-driven analysis* pipeline focused on the time-varying nature of physiological and neurophysiological signals. In this approach, the data are represented as a continuously varying trajectory, enabling the investigation of possible intra- and inter-subject synchrony between the ongoing brain activity and ongoing bodily physiology during the evolution of the learning experience. Indeed, there is an emerging field of research utilizing synchrony measures across individuals to determine behaviorally meaningful associations (see Dikker et al., 2017).

7.3.3.4 Approaching Integration Across Data Modalities

These timelines representing changes in the physiological, neurophysiological, and self-expressed experiential level as well as in the pedagogic interaction were used to pinpoint causal associations and dependencies between events within and across individuals. The integration of multimodal data allowed us to identify reproducible elements reflecting phases of experiential learning and student—instructor interaction. This analysis was conducted separately for each student and instructor as well as for the interaction in each student—instructor dyad. This modality-specific analysis enabled us to test the reliability and reproducibility of measures collected during the naturalistic SBL situations (Fig. 7.3). Additionally, we were able to extract features from each modality that were linked to meaningful episode events during SBL and to structure the data in a way that enabled integration between modalities within the quantitative measures (EEG measures and HRV) and across modalities (self-expressed experiences as well as observational, physiological, and neurophysiological measures). The timestamp annotations (see Sect. 7.3.3.2) in the neurophysiological and video recordings were of critical importance, as temporal synchronization of datasets (EEG, HRV, observational data, and experiences described in the interview) enabled data integration.

7.4 Discussion

In this chapter, we introduced techniques to apply multimodal methods in the study of the adult learning experience. The learning experience is a complex phenomenon that cannot be fully captured via a single-data modality (Aguayo et al., 2018;

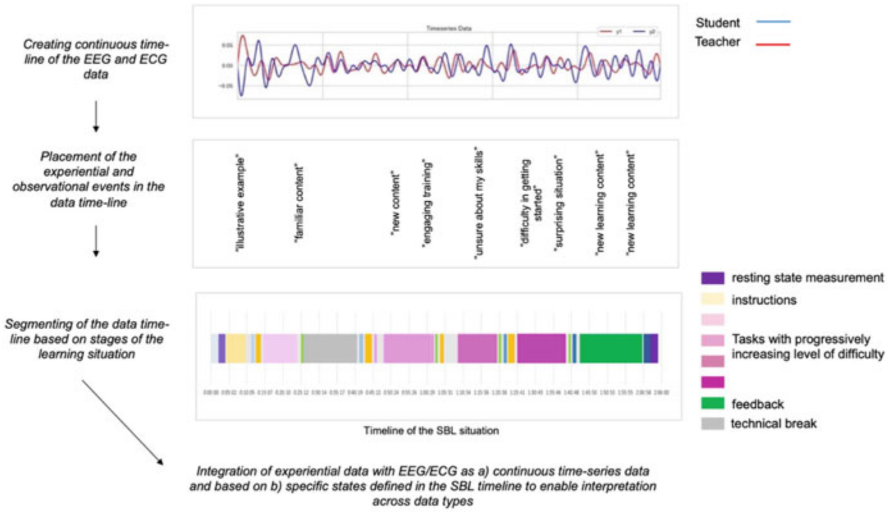

Fig. 7.3 Analysis pipeline

Giannakos et al., 2019; Larmuseau et al., 2019). Different methodological explorations have recently emerged using multimodal combinations of cognitive and physiological measures together with self-reports and interviews to capture various aspects of the learning experience. Despite increasing interest in methodological combinations, multimodal studies of the learning experience are still scarce and have mostly been preliminary in nature. Indeed, there is still a lack of larger-scale studies, and an accumulation of empirical evidence is needed to provide comprehensive understanding of experiential learning. Interestingly, many of the scholars who have approached the learning experience multimodally represented fields other than education and neuroscience, and the current focus seems to be more on technological aspects and learning environment development than on theoretically grounded empirical research. In our case example of SBL, we aimed to develop a multimodal research design that takes into account the complex nature of experiential learning and supports both theoretical and methodological development.

7.4.1 Understanding Learning Experience Requires Crossing Disciplinary Boundaries

Recently, cross-disciplinary and trans-disciplinary research has become more frequent, for example, to approach so-called wicked problems of societal issues and overarching research questions (Bore & Wright, 2009; Stadler et al., 2021). It is clear that the study of naturally occurring behaviors benefits from integration of cross-disciplinary understanding. Referring to Rosenfield's (1992) taxonomy of the level of integration between disciplines, it seems necessary to address learning experience with a transdisciplinary approach, with "researchers working jointly using a shared conceptual framework and drawing together disciplinary-specific theories, concepts, and approaches" (p. 1351). In other words, it is not enough to work in parallel (multidisciplinary) or even jointly (interdisciplinary) to address the questions of learning if the research still builds on a discipline-specific basis.

However, numerous challenges arise in transdisciplinary approaches, for example, from different research designs and traditions in data collection methods (Aagaard-Hansen, 2007). Methodologies are not separable from underlying theoretical assumptions, which inevitably differ between disciplines, for example, between natural sciences (neuroscience) and education sciences. Indeed, discipline-specific expertise is crucially needed for multimodal research approaches. Researchers are not and should not be experts in all fields, but high-quality multimodal research should be the result of cooperation between experts in the fields in which it takes place.

7.4.2 Combining Traditional Approaches with Physiological and Neurophysiological Recordings

The learning experience forms via a complex process. Indeed, multimodal studies of the learning experience present a good example of phenomena that would require both theoretical and methodological integration of lines of work across disciplines that seem far removed from each other (such as neuroscience, psychology, education science, and data science). Physiological measures may reveal, for example, the neural basis or arousal level that contributes to learning, but they do not capture the subjectively experienced elements revealed through the self-reports or via interviews. To best support learning, the experience needs to be understood as a phenomenon in which important factors can be captured at multiple levels, such as physiological contributors, emotional or cognitive significance, subjectively perceived implications, and the determinants of interaction. The multidisciplinary research design of our SBL case exemplifies how a combination of quantitative, third-person measures of human nervous system signaling can be combined with the first-person experiential data of the ongoing learning experience. We have also demonstrated the many challenges, including technological, methodological, and conceptual, in the multimodal research of the learning experience in natural settings.

Currently, technological possibilities seem to antecede methodological development. Advanced technologies are naturally needed to conduct measurements in natural learning situations, but for successful knowledge building, it is crucial to develop methods of valid data integration. There is, for example, a risk of making oversimplifications in the interpretation of physiological signals if discipline-specific background knowledge is not correctly applied in the data analysis. The problem of synchronizing across modalities goes beyond technological and methodological tools, as the meaningful phenomena in each modality differ in crucial ways. For example, the different dimensions of experience do not evolve and manifest themselves in similar ways, especially in the time domain. Unlike for physiological and neurophysiological signals, it is not conventional in educational research to organize data sequentially with detailed temporal dimensions. As pointed out by Eteläpelto et al. (2018), questions related to methodological complementarity, interchangeability, validity, and reliability should be addressed. The adoption of new multimodal methods and research designs will also necessitate new analytical and statistical techniques (Azevedo & Aleven, 2013). A multimodal approach also requires deeper interaction across disciplines at the conceptual level, which would in the long run facilitate the use of common terminology.

7.4.3 A New Theoretical Framework Requires Building Consistent Empirical Evidence of Learning Experiences in Natural Contexts

As discussed in Sect. 7.2.3, there exists no unifying theory of adult learning, which reflects the complexity of learning as a phenomenon (Yang, 2006). Experience itself is a complex phenomenon since it is both longitudinal and episodic and relates to various cognitive domains, such as awareness and perception (Jarvis, 2005a). It is also too early to evaluate the practical implications of multimodal research for the development of experiential learning and education. Each study and theory add a little more to our understanding of the phenomenon, and the relationships between theory, practice, and ideology should be made the central focus not just for philosophical purposes but also to achieve greater empirical precision (Baldwin et al., 2004).

The core challenges for future studies can be summarized as follows. *First*, multimodal approaches have thus far been used to study learning and related experiences (e.g., emotions), not specifically *experiential* learning. Consequently, the contexts vary from using the measures as indicators of stress in exam situations to emotion recognition in a gaming environment.

Second, these studies have been performed in various contexts and with different interpretations and meanings given both for experience and learning and for the physiological measures. Specifically, the concept of the learning experience and the elements involved in it varies across disciplines. Harmonizing the methodology and use of concepts would facilitate the accumulation of scientific knowledge.

Third, and perhaps most importantly, the experimental literature has so far not given rise to a novel theoretical framework in the field of education that would incorporate, and thus give integrative concepts and designs for, both experiential and physiological measures. There do exist theories that build on embodied learning and thus acknowledge the role of body (and brain) systems involved in the experience. These are, however, often rather unspecific, especially regarding the physiological and neurophysiological concepts and processes. The lack of a coherent theoretical basis is of course not a trivial problem, and development of novel theoretical concepts needs to be done in connection with experimental work. Therefore, now is the optimal time to increase discussion among researchers utilizing multimodal methods to understand learning and more broadly human experience and interaction.

References

Aagaard-Hansen, J. (2007). The challenges of cross-disciplinary research. *Social Epistemology, 21*(4), 425–438. https://doi.org/10.1080/02691720701746540

Aguayo, C., Dañobeitia, C., Cochrane, T., Aiello, S., Cook, S., & Cuevas, A. (2018). Embodied reports in paramedicine mixed reality learning. *Research in Learning Technology, 26*, 1–17. https://doi.org/10.25304/rlt.v26.2150

Ahonen, L., Cowley, B. U., Hellas, A., & Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports, 8*. https://doi.org/10.1038/s41598-018-21518-3

Ansari, D., Coch, D., & De Smedt, B. (2011). Connecting education and cognitive neuroscience: Where will the journey take us? *Educational Philosophy and Theory, 43*(1), 37–42. https://doi.org/10.1111/j.1469-5812.2010.00705.x

Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology, 40*(2), 187–209. https://doi.org/10.1080/17439884.2014.933112

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 1–16). Springer. https://doi.org/10.1007/978-1-4419-5546-3_1

Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments, 24*(3), 590–605. https://doi.org/10.1080/10494820.2014.908927

Baldwin, C., Persing, J., & Magnuson, D. (2004). The role of theory, research, and evaluation in adventure education. *The Journal of Experimental Education, 26*(3), 167–183. https://doi.org/10.1177/2F105382590402600307

Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research, 45*(2), 143–154. https://doi.org/10.1080/0013188032000133548

Bearman, M., Greenhill, J., & Nestel, D. (2019). The power of simulation: A large-scale narrative analysis of learners' experiences. *Medical Education, 53*(4), 369–379. https://doi.org/10.1111/medu.13747

Berntson, G. G., Thomas Bigger, J., Jr., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & Van Der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*(6), 623–648. https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

Bhoja, R., Guttman, O. T., Fox, A. A., Melikman, E., Kosemund, M., & Gingrich, K. J. (2020). Psychophysiological stress indicators of heart rate variability and electrodermal activity with application in healthcare simulation research. *Simulation in Healthcare, 15*(1), 39–45. https://doi.org/10.1097/SIH.0000000000000402

Birt, J., Stromberga, Z., Cowling, M., & Moro, C. (2018). Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information, 9*(2), 31. https://doi.org/10.3390/info9020031

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84*(6), 740–756. https://doi.org/10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Bore, A., & Wright, N. (2009). The wicked and complex in education: Developing a transdisciplinary perspective for policy formulation, implementation and professional practice. *Journal of Education for Teaching, 35*(3), 241–256. https://doi.org/10.1080/02607470903091286

Brever, J., & Hunter, A. (1989). *Foundations of multimethod research: Synthesizing styles*. Sage. https://doi.org/10.1080/02607470903091286

Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology, 51*(2), 211–217. https://doi.org/10.1111/j.2044-8279.1981.tb02474.x

Calvo, R. A., & D'Mello, S. K. (Eds.). (2011). *New perspectives on affect and learning technologies* (Vol. 3). Springer.

Clapper, T. C. (2014). Beyond Knowles: What those conducting simulation need to know about adult learning theory. *Clinical Simulation in Nursing, 6*(1), e7–e14. https://doi.org/10.1016/j.ecns.2009.07.003

Cowley, B., Ravaja, N., & Heikura, T. (2013). Cardiovascular physiology predicts learning effects in a serious game activity. *Computers & Education, 60*(1), 299–309. https://doi.org/10.1016/j.compedu.2012.07.014

Daley, B., Martin, L. G., & Roessger, K. M. (2018). A call for methodological plurality: Reconsidering research approaches in adult education. *Adult Education Quarterly, 68*(2), 157–169. https://doi.org/10.1177/0741713618760560

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Mariner.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3–53. https://doi.org/10.1080/10508400903452884

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology, 27*(9), 1375–1380. https://10.1016/j.cub.2017.04.002.

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies, 25*, 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

Dirkx, J. M. (2008). The meaning and role of emotions in adult learning. *New Directions for Adult and Continuing Education, 2008*(120), 7–18. https://doi.org/10.1002/ace.311

Eteläpelto, A., Kykyri, V.-L., Penttonen, M., Hökkä, P., Paloniemi, S., Vähäsantanen, K., Eteläpelto, T., & Lappalainen, V. (2018). A multi-componential methodology for exploring emotions in learning: Using self-reports, behaviour registration, and physiological indicators as complementary data. *Frontline Learning Research, 6*(3), 6–36. https://doi.org/10.14786/flr.v6i3.379

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality, 41*(4), 868–888. https://doi.org/10.1016/j.jrp.2006.11.002

Fenwick, T. (2006). Inside out of experiential learning: Fluid bodies, co-emergent minds. In R. Edwards, J. Gallacher, & S. Whittaker (Eds.), *Learning outside the academy. International research perspectives on lifelong learning* (pp. 42–55). Routledge.

Firstbeat Technologies Ltd. (2014). *Stress and recovery analysis method based on 24-hour heart rate variability*. https://assets.firstbeat.com/firstbeat/uploads/2015/11/Stress-and-recovery_white-paper_20145.pdf

Fortenbacher, A., & Yun, H. (2020). Can sensors effectively support learning? In N. Pickwart & S. Liu (Eds.), *Artificial intelligence supported educational technologies: Advances in analytics for learning and teaching* (pp. 93–114). Springer. https://doi.org/10.1007/978-3-030-41099-5_6

Fromm, J., Radianti, J., Wehking, C., Stieglitz, S., Majchrzak, T. A., & vom Brocke, J. (2021). More than experience? On the unique opportunities of virtual reality to afford a holistic experiential learning cycle. *The Internet and Higher Education*. https://doi.org/10.1016/j.iheduc.2021.100804

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management, 48*, 108–119. https://doi.org/10.1016/j.ijinfomgt.2019.02.003

Girzadas, D. V., Jr., Delis, S., Bose, S., Hall, J., Rzechula, K., & Kulsta, E. B. (2009). Measures of stress and learning seem to be equally affected among all roles in a simulation scenario. *Simulation in Healthcare, 4*(3), 149–154. https://doi.org/10.1097/SIH.0b013e3181abe9f2

Gorard, S. (2002). Can we overcome the methodological schism? Four models for combining qualitative and quantitative evidence. *Research Papers in Education Policy and Practice, 17*(4), 345–361. https://doi.org/10.1080/02671520220000031405

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. McGraw-Hill.

Haataja, E., Malmberg, J., & Järvelä, S. (2018). Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. *Computers in Human Behavior, 87*, 337–347. https://doi.org/10.1016/j.chb.2018.06.007

Hagen, M., & Park, S. (2016). We knew it all along! Using cognitive science to explain how andragogy works. *European Journal of Training and Development, 40*(3), 171–190.

Hammond, C. (2005). The wider benefits of adult learning: An illustration of the advantages of multi-method research. *International Journal of Social Research Methodology, 8*(3), 239–255. https://doi.org/10.1080/13645570500155037

Hannaford, C. (1995). *Smart moves: Why learning is not all in your head*. Great Ocean Publishers.

Hardy, M., Wiebe, E. N., Grafsgaard, J. F., Boyer, K. E., & Lester, J. C. (2013). Physiological responses to events during training: Use of skin conductance to inform future adaptive learning systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 2101–2105. https://doi.org/10.1177/1541931213571468

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48*, 615–625. https://doi.org/10.1016/j.chb.2015.02.013

Hemingway, A., Carter, S., Callaway, A., Kavanagh, E., & Ellis, S. (2019). An exploration of the mechanism of action of an equine-assisted intervention. *Animals, 9*(6), 303. https://doi.org/10.3390/ani9060303

Husebø, S. E., O'Regan, S., & Nestel, D. (2015). Reflective practice and its role in simulation. *Clinical Simulation in Nursing, 11*(8), 368–375. https://doi.org/10.1016/j.ecns.2015.04.005

Illeris, K. (2018). An overview of the history of learning theory. *European Journal of Education, 53*(1), 86–101. https://doi.org/10.1111/ejed.12265

Jarvis, P. (2005a). Towards a philosophy of human learning: An existentialist perspective. In P. Jarvis & S. Parker (Eds.), *Human learning: An holistic approach* (pp. 1–15). Routledge.

Jarvis, P. (2005b). *Towards a comprehensive theory of human learning* (1st ed.). Routledge. https://doi.org/10.4324/9780203001677

Jarvis, P. (2006). *Towards a comprehensive theory of human learning*. Routledge.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Jordi, R. (2010). Reframing the concept of reflection: Consciousness, experiential learning, and reflective learning practices. *Adult Education Quarterly, 61*(2), 181–197. https://doi.org/10.1177/0741713610380439

Judd, B. K., Alison, J. A., Waters, D., & Gordon, C. J. (2016). Comparison of psychophysiological stress in physiotherapy students undertaking simulation and hospital-based clinical education. *Simulation in Healthcare, 11*(4), 271–277. https://doi.org/10.1097/SIH.0000000000000155

Kagan, N., Krathwohl, D. R., & Miller, R. (1963). Stimulated recall in therapy using video tape: A case study. *Journal of Counseling Psychology, 10*(3), 237–243. https://doi.org/10.1037/h0045497

Kocialkowski, C., Rybinski, B., Hainsworth, L., & Yang, D. (2020). Simulation learning: Effectiveness and stressfulness in medical student teaching. *MedEdPublish, 9*(1). https://doi.org/10.15694/mep.2020.000143.1

Kolb, A., & Kolb, D. A. (2017). *The experiential educator: Principles and practices of experiential learning*. EBLS Press.

Larmuseau, C., Vanneste, P., Cornelis, J., Desmet, P., & Depaepe, F. (2019). Combining physiological data and subjective measurements to investigate cognitive load during complex learning. *Frontline Learning Research, 7*(2), 57–74. https://doi.org/10.14786/flr.v7i2.403

Lateef, F. (2010). Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma and Shock, 3*(4), 348–352. https://doi.org/10.4103/0974-2700.70743

LeBlanc, V. R. (2019). The relationship between emotions and learning in simulation-based education. *Simulation in Healthcare, 14*(3), 137–139. https://doi.org/10.1097/SIH.0000000000000379

Lim, D. H., Chai, D. S., Park, S., & Doo, M. Y. (2019). Neuroscientism, the neuroscience of learning: An integrative review and implications for learning and development in the workplace. *European Journal of Training and Development, 43*(7/8), 619–642. https://doi.org/10.1108/EJTD-03-2019-0033

Lumma, A. L., & Weger, U. (2021). Looking from within: Comparing first-person approaches to studying experience. *Current Psychology*. https://doi.org/10.1007/s12144-021-02277-3

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861–878. https://doi.org/10.1080/0141192032000137349

Malinen, A. (2000). *Towards the essence of adult experiential learning: A reading of the theories of Knowles, Kolb, Mezirow, Revans and Schön*. University of Jyväskylä.

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., . . . Valenza, G. (2018). Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports, 8*(1), 1–15. https://doi.org/10.1038/s41598-018-32063-4

Morris, T. H. (2020). Experiential learning – A systematic review and revision of Kolb's model. *Interactive Learning Environments, 28*(8), 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

Mudrick, N. V., Azevedo, R., & Taub, M. (2019). Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Computers in Human Behavior, 96*, 223–234. https://doi.org/10.1016/j.chb.2018.06.028

Papastamatis, A., & Panitsides, E. A. (2014). Transformative learning: Advocating for a holistic approach. *Review of European Studies, 6*(4), 74–81. https://doi.org/10.5539/res.v6n4p74

Patton, M. Q. (2002). Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative Social Work, 1*(3), 261–283.

Pekrun, R., & Schutz, P. A. (2007). Where do we go from here? Implications and future directions for inquiry on emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education: Educational psychology* (pp. 313–331). Academic. https://doi.org/10.1016/B978-012372545-5/50019-8

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., & Strohecker, C. (2004). Affective learning—A manifesto. *BT Technology Journal, 22*(4), 253–269. https://doi.org/10.1023/B:BTTJ.0000047603.37042.33

Poore, J. A., Cullen, D. L., & Schaar, G. L. (2014). Simulation-based interprofessional education guided by Kolb's experiential learning theory. *Clinical Simulation in Nursing, 10*(5), e241–e247. https://doi.org/10.1016/j.ecns.2014.01.004

Quintana, D., Guastella, A., Outhred, T., Hicki, I., & Kempf, A. (2012). Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition. *International Journal of Psychophysiology, 86*(2), 168–172. https://doi.org/10.1016/j.ijpsycho.2012.08.012

Rogers, T., Andler, C., O'Brien, B., & van Schaik, S. (2019). Self-reported emotions in simulation-based learning: Active participants vs. observers. *Simulation in Healthcare, 14*(3), 140–145. https://doi.org/10.1097/SIH.0000000000000354

Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science & Medicine, 35*(11), 1343–1357. https://doi.org/10.1016/0277-9536(92)90038-R

Rowley, J. (2012). Conducting research interviews. *Management Research Review, 35*(3/4), 260–271. https://doi.org/10.1108/01409171211210154

Salmi, E., Määttä, S., Vehkakoski, T., Aunola, K., Kairaluoma, L., & Pirttimaa, R. (2020). Oppimisvaikeuksien, motivaation ja oppijaminäkäsityksen merkitys ammatillisista opinnoista valmistumisessa. [The importance of learning difficulties, motivation and the student's self-perception in graduating from vocational studies.]. *Oppimisen ja oppimisvaikeuksien erityislehti: NMI-bulletin, 30*(3), 50–66. https://jyx.jyu.fi/handle/123456789/72613

Schaaff, K., Müller, L., Kirst, M., & Heuer, S. (2012). Affect–A modular framework for online affect recognition and biofeedback applications. In *7th European conference on technology enhanced learning (ECTEL 2012), MATEL workshop,* Saarbrücken, Germany.

Schenck, J., & Cruickshank, J. (2015). Evolving Kolb: Experiential education in the age of neuroscience. *The Journal of Experimental Education, 38*(1), 73–95. https://doi.org/10.1177/1053825914547153

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health, 5*, 1–17. https://doi.org/10.3389/fpubh.2017.00258

Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Journal of Educational Technology & Society, 12*(2), 176–189. https://www.jstor.org/stable/jeductechsoci.12.2.176

Silvennoinen, M., Vesisenaho, M., Manu, M., Kullberg, T., Malinen, A., & Parviainen, T. (2020). Methodology development in adult learning research: Combining physiological reactions and learning experiences in simulation-based learning environments. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *EDULEARN20 proceedings. 12th international conference on education and new learning technologies* (pp. 5037–5046). IATED. EDULEARN Proceedings. https://doi.org/10.21125/edulearn.2020.1316

Spangler, G., Pekrun, R., Kramer, K., & Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. *Anxiety, Stress & Coping, 15*(4), 413–432. https://doi.org/10.1080/1061580021000056555

Stadler, M., Graesser, A., & Fischer, F. (2021). Editorial: Transdisciplinary research on learning and teaching: Chances and challenges. *Frontiers in Psychology, 12*, 696219. https://doi.org/10.3389/fpsyg.2021.696219

Staus, N. L., & Falk, J. H. (2017). The role of emotion in informal science learning: Testing an exploratory model. *Mind, Brain, and Education, 11*(2), 45–53. https://doi.org/10.1111/mbe.12139

Stepanova, E. R., Quesnel, D., & Riecke, B. E. (2019). Space—A virtual frontier: How to design and evaluate a virtual reality experience of the overview effect. *Frontiers in Digital Humanities, 6*(7). https://doi.org/10.3389/fdigh.2019.00007

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine, 113*, 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

van Atteveldt, N., van Kesteren, M. T. R., Braams, B., & Krabbendam, L. (2018). Neuroimaging of learning and development: Improving ecological validity. *Frontline Learning Research, 6*(3), 186–203. https://doi.org/10.14786/flr.v6i3.366

Varela, F. J., & Shear, J. (1999). First-person methodologies: What, why, how. *Journal of Consciousness Studies, 6*(2/3), 1–14.

Vesisenaho, M., Juntunen, M., Häkkinen, P., Pöysä-Tarhonen, J., Fagerlund, J., Miakush, I., & Parviainen, T. (2019). Virtual reality in education: Focus on the role of emotions and physiological reactivity. *Journal of Virtual Worlds Research, 12*(1), 1–15. https://doi.org/10.4101/jvwr.v12i1.7329

Wang, C., & Cesar, P. (2015). Physiological measurement on students' engagement in a distributed learning environment. In *Proceedings of the 2nd international conference on physiological computing systems (PhyCS 2015)* (pp. 149–156). SCITEPRESS – Science and Technology Publications, Lda, Setubal, PRT. https://doi.org/10.5220/0005229101490156

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology, 84*(3), 552–569. https://doi.org/10.1016/j.biopsycho.2010.01.017

Wilhelm, F. H., Grossman, P., & Müller, M. I. (2012). Bridging the gap between the laboratory and the real world: Integrative ambulatory psychophysiology. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 210–234). Guilford.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology, 4*(3/4), 129–164.

Yang, B. (2006). A holistic conceptualization of adult learning and its critiques of selected concepts and theories. In *Proceedings of the 47th annual adult education research conference* (pp. 482–488). University of Minnesota. https://newprairiepress.org/aerc/2006/papers/79

Zembylas, M. (2007). Theory and methodology in researching emotions in education. *International Journal of Research & Method in Education, 30*(1), 57–72. https://doi.org/10.1080/17437270701207785

Minna Silvennoinen (PhD) is a Senior Researcher in the Scool of Professional Teacher Education JAMK University of Applied Sciences and researcher in KoKemus research project. Her research is mainly focused on skills learning, learning processes and learner experiences on both individual and group level as well as learner-teacher interaction in the contexts of adult education and simulation-based learning.

Tiina Parviainen (PhD) is Associate Professor of Neuroscience in the Department of Psychology and the director of the Center for Interdisciplinary Brain Research (CIBR) in the University of Jyväskylä. She uses neuroimaging methods and physiological recordings to study central and autonomic nervous system phenomena that contribute to individual experience and cognitive performance.

Anita Malinen (PhD) is a senior lecturer in the Department of Education at the University, Finland. She is a researcher in KoKemus research project. Her research is focused on experiential learning theory, its philosophical foundations and developmental work on experiential learning practices, especially in adult education contexts.

Suvi Karjalainen (M. Psych.) is a project researcher in the KoKemus research project at the Department of Teacher Education and a doctoral researcher at the Department of Psychology at the University of Jyväskylä, Finland. Her research is focused on the interaction between autonomic and central nervous system functioning, especially among individuals with neurodevelopmental conditions.

Mari Manu (MA, MEd) is a project researcher and a doctoral researcher at the Department of Teacher Education at the University of Jyväskylä, Finland. She works in Technological and Societal Innovations to Cultivate Critical Reading in the Internet Era (CRITICAL) and in KoKemus—research projects. Her research is focused on children's and adolescents' reading skills and reading development.

Mikko Vesisenaho (PhD) is a teacher educator and an educational technologist. He is also pioneering interdisciplinary research approaches by combining a variety of theories, methodologies, and contexts. Currently he leads e.g. KoKemus research project.

Chapter 8
Measuring Professional Competence Using Computer-Generated Log Data

Luca Spliethoff and Stephan Abele

Abstract One of the benefits of computer-based assessments lies in the automatic generation of log data. Such behavioural process data provide a time-stamped documentation of students' interactions with the assessment system (e.g., mouse clicks). This chapter explores the usefulness of computer-generated log data for the measurement of professional competence and their potential for the research on professional learning and development. Based on a selection of studies, we illustrate how interindividual differences in task completion processes can be analysed with the help of log data, e.g. to identify the use of certain problem-solving strategies, or to reveal subgroups of students with efficiency barriers. We further present our own research, where we applied a theory on the diagnostic process (Abele, Vocat Learn 11(1):133–159, 2018) in order to assess diagnostic strategies (Abele and von Davier, CDMs in vocational education: assessment and usage of diagnostic problem-solving strategies in car mechatronics. In: von Davier M, Lee YS (eds) Handbook of diagnostic classification models. Springer International Publishing, pp 461–488. https://doi.org/10.1007/978-3-030-05584-4_22, 2019) in the domain of car mechatronics using log data. A profound understanding of interindividual process differences may supplement a merely product-oriented competence measurement and pave the way for a more process-oriented approach. Challenges concerning the assessment, analysis and interpretation of log data will be discussed.

Keywords Process data · Log data · Competence measurement · Computer-based assessment · Vocational education

L. Spliethoff · S. Abele (✉)
Institute of Vocational Education and Vocational Didactics, Technical University of Dresden, Dresden, Germany
e-mail: luca.spliethoff@tu-dresden.de; stephan.abele@tu-dresden.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_8

8.1 Introduction

To generate a profound understanding of professional learning and development, we need to properly assess professional competence (see Beck et al., 2016; Winther, 2010). Competence refers to one's ability and readiness to solve specific problems in variable contexts and situations (Weinert, 2002). This ability involves successfully regulating a process towards a specific goal. Traditional competence measurement usually focuses upon the outcomes of the tasks – i.e., whether the intended goals have been reached or not (e.g., Abele et al., 2014). The outcomes serve as a basis for drawing conclusions about underlying competences. However, considering only the results of the problem-solving process neglects everything that occurs during the process before the final answer is given. Process data represent a window into this problem-solving process by providing us with insights into how and how well a user performs in a task. Furthermore, it has been shown by various scholars that process data allow conclusions about underlying cognitive processes (e.g., Richter et al., 2005; Stelter et al., 2015).

Computer-generated log data refers to process data that are collected as a by-product of a computer-based test. During the assessment, every action of the test takers, such as their mouse clicks, can be automatically recorded with a corresponding time stamp. The resulting log data are stored in log files for every test taker and provide an extensive documentation of their interactions with the assessment system (Goldhammer et al., 2017; Rausch et al., 2017). There are different types of log data: (1) log data that refer to information about a test taker's location, the used browser, or the participation status (login, logout, resume); (2) log data that involve information about keystrokes, mouse klicks, or answer-change events to items; (3) log data that involve information that is gathered beyond answers and inputs, like scrolling or zooming activities, or a test taker's request of additional information (see Kroehne & Goldhammer, 2018, for a more detailed categorisation of log data). An important characteristic of log data is their granularity and, accordingly, the amount of information the data provide. Traditional multiple-choice items produce rather basic log data, documenting the time users spend on a task, their reaction times, and changes in their responses. In contrast, more complex item formats such as highly interactive computer simulations provide log data at a much smaller grain size, including, for example, information about every mouse click, key stroke, and zooming activity (Goldhammer et al., 2017).

In vocational education and training (VET), computer-based test environments and the analysis of log data present an opportunity to measure what has been learned. Several computer-based test environments have recently been developed for this purpose. ALUSIM, for example, is an assessment system for commercial competences that visualises work processes in an industrial enterprise, requiring the participants to answer e-mails, edit texts, and process daily statistics (Sangmeister et al., 2018). Another example is a digital office simulation for students in commercial VET, involving complex problem scenarios specific to their field (Rausch et al., 2017). Within this simulation, in order to solve job-specific problems, students are

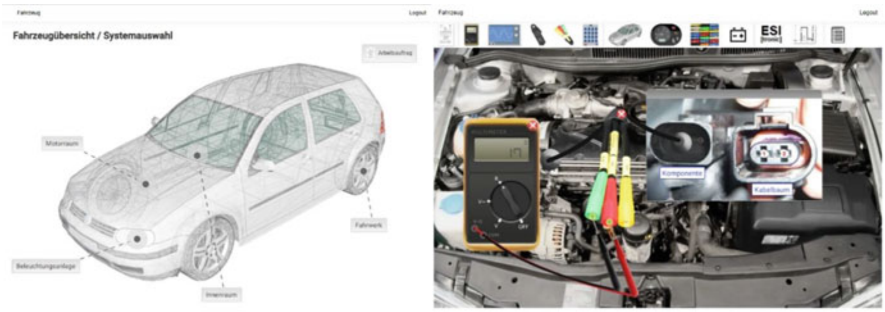

Fig. 8.1 Computer simulation for car mechatronics; Left: Overview of car systems; Right: Resistance measurement of the exhaust gas recirculation valve with a multimeter

invited to read messages, study authentic documents, and perform calculations. In the technical domain, Gschwendtner et al. (2009) have developed a computer simulation for capturing the diagnostic problem-solving competences of car mechatronics technicians (see Fig. 8.1, left, for the start of the simulation). The simulation contains genuine graphic material that represents different aspects of car mechatronics' work environment (e.g., a circuit diagram) and allows interactions that match their professional reality, such as resistance measurement (Fig. 8.1, right). Figure 8.2 shows an excerpt from a log file that resulted from the problem-solving process in this kind of computer simulation for car mechatronics.

In addition to computer-generated log data, other in situ approaches, such as *observations, think aloud protocols, experience sampling procedures*, or *diaries* exist to collect process data without a potential recall bias as in traditional retrospective self-reports (Pekruna, 2020; see also Rausch et al., 2022). However, by using these methods the problem-solving process itself can be interrupted or affected by the data elicitation (Kane, 1992; Napa Scollon et al., 2009; Zhang & Zhang, 2019; Zirkel et al., 2015). By contrast, log data are automatically captured online, without the need to influence or interrupt the task completion process or to reconstruct the process from fragmentary post-hoc declarations. Consequently, log data from computer-based assessments can deliver a more adequate picture of the task completion process (Veenman et al., 2014). Nevertheless, their interpretation often raises serious questions, which we will address later in this chapter. A detailed recording of the steps leading up to the final answer, as provided by log data, allows for a more precise gradation of professional competence levels and thus also for an accurate description of the learning success.

In this chapter, we aim to explore the value of computer-generated log data for the measurement of professional competence and therefore rather focus the process than outcome aspect of competence. The assessment of this process will help us, for example, to explore strategies, barriers and cognitive processes of professional experts and novices and, thereby, enhance our understanding of professional learning and development. To illustrate the value of log data for the examination of interindividual process differences in professional competence, we will present a

```xml
<?xml version="1.0"?>
- <activityLog>
    <activity lightState="00" motorState="000" timeStamp="12:15:19" id="1">lesson loaded: Fehler 6</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:00:02" id="2">menu open: SystemSelection</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:01:52" id="3">menu open: Document</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:01:52" id="4">Arbeitsauftrag page: 1</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:29" id="5">button close: Arbeitsauftrag</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:39" id="6">system loaded: Motorraum</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:42" id="7">system loaded: Cockpit</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:44" id="8">ignition on: true</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:45" id="9">shortCut open: DiagnoseSoftware</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:45" id="10">ESITRONIC screen: 00_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:47" id="11">ESITRONIC screen: 00_02</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:56" id="12">ESITRONIC screen: g4_00_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:02:58" id="13">ESITRONIC screen: g4_03_01_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:03" id="14">ESITRONIC screen: g4_03_01_13_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:10" id="15">ESITRONIC screen: g4_03_01_13_m01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:13" id="16">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:18" id="17">ESITRONIC Fehlerspeicher lesen:</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:22" id="18">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_02</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:03:57" id="19">ESITRONIC screen: g4_44A3_01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:12" id="20">button close: null</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:21" id="21">module open: Motorabdeckung</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:23" id="22">module zoom in: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:25" id="23">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:27" id="24">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:29" id="25">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:34" id="26">shortCut open: Multimeter</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:42" id="27">Messspitze aufgesetzt: red/01-04-03-01</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:45" id="28">Messspitze aufgesetzt: black/01-04-03-02</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:04:45" id="29">Multimeter measure: 01-04-03-01/01-04-03-02 (5 Udc)</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:05:20" id="30">button close: Multimeter</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:05:22" id="31">shortCut open: DiagnoseSoftware</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:05:27" id="32">ESITRONIC screen: g4_44A3_02</activity>
    <activity lightState="00" motorState="100" timeStamp="+00:05:38" id="33">button close: null</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:41" id="34">ignition on: false</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:45" id="35">shortCut open: Multimeter</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:54" id="36">Messspitze aufgesetzt: black/01-04-03-04</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:59" id="37">Messspitze aufgesetzt: red/01-04-03-03</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:59" id="38">Multimeter measure: 01-04-03-03/01-04-03-04 (OL R)</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:39" id="39">button close: Multimeter</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:40" id="40">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:42" id="41">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:43" id="42">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:44" id="43">module zoom out: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:46" id="44">button close: null</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:49" id="45">lesson FINISHED</activity>
</activityLog>
```


Fig. 8.2 Log file excerpt for a single test taker resulting from an assessment using the car computer simulation. Every row shows a time-stamped activity conducted by the student. The green box ('measuring tip attached') highlights potential measurement at the fuel temperature sensor with the multimeter. (Abele & von Davier, 2019, p. 465)

selection of studies that combined process and outcome data from computer-based assessments (Sect. 8.2). As the research on log data in VET is only at a preliminary stage (e.g., Abele & von Davier, 2019; de Klerk, 2016; Rausch et al., 2017), we also present related research from other educational sectors that we consider relevant to the domain of VET. As problem solving is very important in professional contexts, we mainly include research on this subject. Aiming to demonstrate in more detail how log data can be used in the context of VET, we present original research conducted in the domain of car mechatronics (Sect. 8.3). In our research, we used a theory around the diagnostic problem-solving process (Abele, 2018) to investigate the use of different diagnostic problem-solving strategies with the help of log data (Abele & von Davier, 2019). Certainly, the use of log data for the research on professional competence involves specific challenges and inherent boundaries. Therefore, in the fourth section, we will outline potential problems of this methodology for each step of the research process and present some suggestions on how these problems could be overcome (Sect. 8.4). We conclude with outlining future directions in the context of log data and the research on professional competence, learning and development (Sect. 8.5).

8.2 Investigating Differences in Problem-Solving Processes Using Log Data

8.2.1 Theory-Based Approaches: Identification of A Priori Defined Problem-Solving Strategies

Process-related informations in log files may enhance our understanding of interindividual differences in task outcomes. For example, this information may be used to explain differences between groups of students who successfully solved a problem and those who did not succeed (Goldhammer et al., 2017). Using log data, research investigating the relevance of specific process behaviours to problem-solving success has been conducted on the use of exploration strategies during complex problem-solving tasks (e.g., Greiff et al., 2015; Wüstenberg et al., 2014). MicroDYN is a common paradigm for the investigation of complex problem-solving (Greiff & Funke, 2009). The name *MicroDYN* is derived from the fact that typical MicroDYN elements consist of small *dynamic systems* (i.e., problems) that contain various input and output variables. The relationship of the input variables to the output variables, i.e. the problem's underlying causal structure, is opaque to the students at the beginning of the task. To find out the underlying relations, students need to apply adequate exploration strategies (*exploration phase*). The use of certain exploration strategies allows conclusions about how systematically someone has explored the problem. A well-investigated exploration strategy is the *vary-one-thing-at-a-time strategy* (VOTAT; Tschirgi, 1980), which involves separately manipulating the input variables, in order to determine the causal structure of a system. Researchers are able to identify participants' application of this strategy during the problem-solving process by searching for *a priori* defined patterns in the log data representing the VOTAT strategy. Either in parallel or after the exploration phase, students need to explicate their acquired knowledge by drawing a model of the problem's structure (*model-building phase*). The model drawn by the students serves as basis for their final problem-solving performance. Several studies have shown that the application of VOTAT, operationalised by specific activity patterns in the log data, explains a strong amount of variance in students' complex problem-solving performance outcomes (Greiff et al., 2015).

8.2.2 Data-Driven Approaches: Exploration of Unknown Data Patterns

However, looking for *a priori* defined behavioural patterns in the log data, such as the application of the well-established VOTAT strategy by the test takers, cannot always explain differences in task outcomes exhaustively. In such cases, exploratory data-driven analyses may help the researchers to identify previously unknown relationships in the data. The newly discovered patterns gained from this exploratory approach may serve to distinguish between successful and unsuccessful problem solvers beyond the well-known patterns (Kinnebrew et al., 2017; Martin & Sherin, 2013; Rausch et al., 2017). Computer-based log files provide a useful basis for this purpose, as they can provide a fine-grained recording of the problem-solving process. Exploring the relationship between the frequency of certain activities during the task completion process and the problem-solving outcome represents a simple example of such a data-driven analysis. For instance, in the so-called 'otter task', used in a study by Veenman et al. (2014), students can conduct experiments to determine the impact of various factors of influence on the development of the otter population. The authors found a positive relationship between the frequency of experiments that subjects had conducted during the task and specific metacognitive performance indicators. As these exploratory analyses are not based on sound theoretical assumptions, it is difficult to make assertions about the meaning of these results. The procedure can therefore be regarded as hypothesis-generating (Rausch et al., 2017). Subsequent experimental studies to validate the results are advisable.

Whereas studying the frequency of activities is an example of a rather simple log data analysis, data-driven methods often involve more complex statistical procedures, such as Hidden Markov Models, artificial neural networks, and n-grams (He & Von Davier, 2016; Lajoie et al., 2015; Stadler et al., 2019). The *n-gram* approach by Damashek (1995) was applied in a study by Stadler et al. (2019). The background to their study was that empirical findings could show the importance of the VOTAT strategy for complex problem-solving performance, but could not explain why some students applying this strategy fail to solve a problem while others succeed (Kuhn & Dean, 2005; Wüstenberg et al., 2014). Stadler et al. (2019) therefore aimed to achieve a more detailed picture of the problem-solving process that went beyond the mere application of VOTAT. They analysed the log data of students who solved several problem-solving tasks based on the MicroDYN paradigm (see Sect. 8.2.1). The goal was to identify distinct behavioural patterns in the log data that differentiate between students who applied the VOTAT strategy and were subsequently successful in solving the problem (i.e., drawing the correct model concerning the relationship between input and output variables), and those who applied VOTAT but were unsuccessful solving the problem.

The n-gram method decomposes a long string of actions (e.g., behaviours shown in a computer-based task) into small interpretable sequences of events (*n-grams*). Here, the authors chose two types of events (S and M) that could be displayed by students during the task. The action S (= working on the scenario) forms part of the exploration phase, in which students investigate the problem structure. It indicates that the student applies changes on the input variables. M (= working on the model) indicates that a student visualizes the relations that he/she assumes between the input and the output variables in a model (model-building phase). All possible bi-grams (e.g., MM, MS), tri-grams (e.g., MSM, SSS), and four-grams (e.g., MMMS, SMSS) were extracted for these two event types. Comparing the frequencies of the n-grams between the successful and the unsuccessful group using the VOTAT strategy revealed specific sequences of behaviours that were critical for success (e.g., MMM, MMMM) or failure (e.g., SS, SSS) during the problem-solving process. Students who did not succeed in solving the problem worked maximally long on the problem scenario (i.e., by exploring the effects of the input on the output variables), whereas students who applied VOTAT and were subsequently successful appeared to work maximally long on the model (i.e., by plotting their findings concerning the relationship between input and output variables). As stated by the authors, the results of this fine-grained log data analysis point to the importance not only of using the correct strategy (here: VOTAT) to solve a problem, but of drawing on meta-strategic competences in order to utilise the information gathered during the process. They presumed that students who did not succeed in solving the problem either did not understand the strategy correctly or could not manage the use of the strategy.

Similarly, He and von Davier (2016) conducted n-gram analyses on log data from problem-solving tasks of the *Programme for the International Assessment of Adult Competencies* (PIAAC; see Schleicher, 2008) that revealed significant differences in patterns of behavioural action sequences between performance groups. These studies show how log data analyses, such as the n-gram approach, can help to identify previously unknown patterns in data. In turn, this can broaden our understanding of the processes that underlie successful problem-solving behaviour. However, it should be noted that these exploratory results should be interpreted with caution and ideally be substantiated by subsequent experimental studies (Stadler et al., 2019).

8.2.3 Professional Problem Solving: Analysing Log Data from a Digital Office Simulation

In the field of VET, Rausch et al. (2017) performed analyses on trainees' log data from an office simulation with domain-specific problem scenarios (see Sect. 8.1). More specifically, they examined students' problem-solving behaviour over time and in relation to problem-solving outcomes. To allow for an easier handling of the significant amount of log data, Rausch et al. conducted several pre-processing steps. Firstly, individual interactions of a student (*log entries*) that belong to the same activity in terms of content and directly follow one another were grouped together and given the same activity label and time stamp. For example, all consecutive keystrokes of a student during an entry in the notepad can be grouped together and provided with the label 'use of notepad'. Secondly, the newly created variables (i.e., the summarised student-task interactions) were exported into a new dataset which is easier to handle during subsequent analyses. Thirdly, the authors grouped the problem-solving processes into short time intervals and determined the frequency of activities per interval.

The subsequent data analysis revealed that both the total number of activities and the frequency of specific activities (e.g., the use of the notepad) were positively related to the problem-solving performance. Furthermore, behavioural differences were identified between performance groups and between problem scenarios which differed in terms of their inherent procedural logic. For instance, one problem scenario required participants to first perform calculations in a spreadsheet before working with other documents. In another problem scenario, the inherent logic of the problem was the opposite: participants had to study various documents before performing calculations in the spreadsheet. The differences between performance groups that were found during the activities indicate that successful problem solvers were able to recognise the procedural logic of the problems and adapt their problem-solving behaviour accordingly, whereas poor problem solvers appeared to proceed rather unsystematically. Ultimately, the study by Rausch et al. (2017) shows how combining log data from highly interactive computer-based assessments with task outcomes can improve our understanding of key elements of trainees' strategies for solving professional problems.

8.2.4 Identification of Efficiency Barriers During the Problem-Solving Process

Analysing log data may also help to gather specific information about students who do not appear to differ when looking at the produced task outcomes. In such cases, a closer look at log data could be useful, for example, to identify potential barriers to efficiency during the task completion process. Following this approach, Tóth et al. (2013) applied a *k-means* cluster algorithm to the process data of a group of students who all succeeded in solving an information and communications technology literacy task. The k-means algorithm is a mathematical technique that classifies a given number of objects into homogeneous groups to ensure that the objects in each group are as similar as possible after classification (Jain et al., 1999). In their analysis, Tóth et al. focused on problem-solving efficiency, operationalised as a combination of different process measures of test takers' interactions with a simulated web environment (e.g., number of pages visited by participants, time spent on the relevant page, total time on task). Applying the k-means cluster algorithm, the researchers identified two sub-groups of successful problem solvers who differed in terms of their problem-solving efficiency. The more efficient subgroup was characterised, among others, by visiting fewer pages and spending less time on the relevant page. According to the authors, the less efficient subgroup may represent students with lower skills who were nevertheless able to successfully complete the task through compensatory behaviour (e.g., more page visits). Eventually, the results of this log data analysis conducted by Tóth et al. allow for a more differentiated picture of students' behavioural patterns by identifying participants who work particularly efficiently.

The selected examples described above demonstrate that there is a variety of approaches to analysing problem-solving processes with the help of computer-generated log data. In recent years, several software tools have been developed to analyse such data (e.g., by enabling the automatic extraction of predefined process indicators from raw log events). Among them are the *PIAAC LogDataAnalyzer* (OECD, 2019), *LogFSM* (Kroehne, 2021), and the R package *LOGAN* (Reis Costa & Leoncio, 2019). In the next section, we will present our own research on computer-based log data from the field of car mechatronics. We first give an overview of diagnostic problem-solving in the field of car mechatronics and derive from this the relevance of assessing problem-solving strategies (Sect. 8.3.1). The following is a detailed description of three different strategies and how we operationalised them (Sect. 8.3.2). We then present information on the statistical analysis as well as selected results of our study (Sect. 8.3.3). We conclude with a short discussion of the results (Sect. 8.3.4), before broadening the perspective and critically debating the use of log data for professional competence (Sect. 8.4). Finally, we outline future directions for the research on professional learning and development (Sect. 8.5).

8.3 Measuring Diagnostic Problem-Solving Strategies of Car Mechatronics Apprentices

8.3.1 Introduction: Why Do We Assess Diagnostic Problem-Solving Strategies in the Domain of Car Mechatronics?

Diagnostic problems are a common type of problem faced by workers in a variety of professional and vocational domains. These problems require workers to identify the causes of undesired states, such as machine malfunctions or diseases (Jonassen, 2010). In the technical domain, diagnostic problem solving is particularly relevant to the field of car mechatronics, where diagnosing the causes of car malfunctions is an important task at work (Baethge & Arends, 2009). According to Abele (2018), the diagnostic problem-solving process consists of the following sub-processes: (a) representing problem-relevant information, (b) formulating diagnostic hypotheses, (c) testing diagnostic hypotheses, and (d) evaluating diagnostic hypotheses. Diagnosticians can apply different diagnostic problem-solving strategies to regulate these sub-processes and identify the cause of an undesired state. These strategies involve different types of problem-solving activities, some of which are directly observable (= diagnostic problem-solving behaviour, e.g., conducting a voltage measurement), and others not (= mental problem-solving behaviour, e.g., mentally representing the symptoms of a diagnostic problem). To date, empirical research on how to differentiate between different types of strategies based on observable problem-solving behaviour is scarce and there is no agreement upon how to operationalise these strategies (Konradt, 1995; Schaper et al., 2004).

8.3.2 Methods: How Can We Assess Diagnostic Problem-Solving Strategies in the Domain of Car Mechatronics?

A promising opportunity for assessing diagnostic strategies lies in the analysis of computer-generated log data, as they provide an extensive picture of the problem-solving process. This approach was followed by Abele and von Davier (2019), who aimed to assess different diagnostic problem-solving strategies and investigate their usage by car mechatronics apprentices through log data analysis. They proposed a theory-based framework to distinguish between three different strategies by applying psychometric models to computer-based log data. The three strategies the authors focused on are *computer-based*, *case-based*, and *mental-model-based* strategies. The formulation of diagnostic hypotheses within the problem-solving process differs between the three strategies, implying that each strategy is associated with a distinctive pattern of mental problem-solving activities as well as observable problem-solving behaviours. These differences between the strategies provided the basis for their empirical distinction (Abele & von Davier, 2019). Abele and von Davier (2019) analysed log data derived from the work of 369 car mechatronics apprentices (mean age: 20.8 years; 96.6% males), who solved four different diagnostic problems in an authentic computer-based car simulation (see Fig. 8.1). Two pairs of problem scenarios were administered to each test taker. Here, to give an example, we only report on one problem pair that contains problems in the fuel temperature sensor (*sensor problems*).[1] The two problems contained in this pair are similar in terms of their symptoms, but different in terms of the cause of their symptoms and in their degree of difficulty, comprising a simple and a difficult problem. Based on these problems, the authors theoretically defined idiosyncratic patterns of problem-solving behaviour for each diagnostic strategy. How this was done is described in the next section (Sect. 8.3.2.1). The derived patterns can then in turn be used to determine which test takers exhibited each strategy (i.e., pattern).

8.3.2.1 A Priori Definition of Critical Problem-Solving Behaviour

The log file contains a long string of actions (see Fig. 8.2), some of which might be less relevant than others. It is therefore advisable to determine a selection of log entries that are essential for specific research purposes. Abele and von Davier (2019) decided to focus on critical test behaviour and critical information behaviour. In order to identify these two types of behaviours, the authors proceeded in a threefold manner. Firstly, they theoretically derived critical diagnostic hypotheses for each diagnostic problem. Critical hypotheses comprise potential and plausible causes of

[1] With regard to the sensor problems, the diagnostic classification model (see Sect. 8.3.3) supports a distinction between the three strategies. However, for the other problem pair assessed in the study, the results were less clear. For further details, see Abele and von Davier (2019).

an undesired state for a specific diagnostic problem. For instance, assuming that an empty tank is the cause of the 'check engine' light coming on would not be considered a plausible reason for the problem. The hypothesis that a broken fuel sensor is responsible for the defect, in contrast, does make sense and can therefore be considered as a critical hypothesis for this specific diagnostic problem. It is important to note that even though different critical hypotheses usually exist for each diagnostic problem, in each scenario of this study there is only one *true* hypothesis indicating the actual cause of the malfunction. Secondly, based on the relevant hypotheses, critical information and tests were determined for each hypothesis. Critical information could be, for example, the symptoms of a diagnostic problem that a diagnostician needs to be aware of in order to generate a hypothesis. Thirdly, based on the critical information and tests, the respective information behaviour (e.g., opening the sensor circuit diagram) and test behaviour (e.g., performing a sensor cable test) must be defined for each critical hypothesis. These behaviours must be visible within the computer simulation and hence identifiable using log data. Altogether, Abele and von Davier (2019) theoretically defined five critical information behaviours and four critical test behaviours for the sensor problems (see Table 8.1, left). In the following, we describe the three strategies and how the specific patterns of information and test behaviour are assumed to differ between them.

8.3.2.2 Deriving Behavioural Patterns for each Strategy

When using the *computer-based strategy* to diagnose the cause of a malfunction, diagnosticians follow the instructions that are provided by an external source, namely a computer-based expert system. This expert system provides the most relevant, but not all possible critical hypotheses for a diagnostic problem and may therefore be incomplete. Hence, for the computer-based strategy, only critical information behaviour and test behaviour should be found in the log data that relates to critical hypotheses provided by the computer.

By using the *case-based strategy*, diagnosticians retrieve knowledge from long-term memory about previous cases similar to the actual problem. In this strategy, the critical hypotheses are thus derived from experience. Based on previous studies, we assumed that learners in their third year of training only know a few of these hypotheses. This strategy is especially useful for simple diagnostic problems because the difficulty of a problem is related to the familiarity of a problem, meaning that problems that are common in real life are less difficult than other problems as hypotheses must not be generated but can be retrieved from memory. Both computer-based and case-based strategies require relatively low levels of mental effort and a shorter amount of time for information processing (i.e., low individual costs), as critical hypotheses and problem-solving templates are provided by the computer-based expert system or retrieved from memory. Therefore, in both cases, no profound knowledge about problem-related systems is required to apply these two strategies.

Table 8.1 Strategy-specific patterns of critical information and critical test behaviour for the sensor problems

Critical problem-solving behaviour (Activities in the computer simulation that are observable in the log files)		Computer-based strategy (Follow instructions of computer-based expert system)	Case-based strategy (Procedure based on experience with similar problems)	Model-based strategy (Derive hypotheses systematically based on mental models)
Information behaviour (Retrieving external information material)	Sensor problem description	+	+	+
	Expert system instruction 1	+	−	−
	Expert system instruction 2	+	−	−
	Sensor circuit diagram	−	−	+
	Sensor location diagram	+/−	+/−	+/−
Test behaviour (Conducting tests)	Sensor test	+	+	+
	Sensor cable test 1	−	−	+
	Sensor cable test 2	−	−	+
	Engine control unit test	−	−	+

Note: +: high probability; −: low probability; +/−: low, middle or high probability depending on previous experience with the location of the sensor (Adapted from Abele & von Davier, 2019, p. 475)

Unlike the first two strategies, the *model-based strategy* is a demanding and an analytical problem-solving procedure. Diagnosticians integrate both internal information (e.g., system knowledge) and external information derived from interactions with the problem environment to mentally represent parts of the car system. The mental models are used to derive critical diagnostic hypotheses systematically (Perez, 2012). Diagnosticians applying the model-based strategy can show a full range of critical information behaviour and critical test behaviour for the respective problem. In contrast, the first two strategies in our study are associated with a selection of these behaviours due to the limited information provided by the expert system and the limited experience of our participants (see above). The application of the model-based strategy requires a relatively high level of mental effort, as well as a longer amount of time for information processing. Diagnosticians need to possess deep knowledge about problem-related systems and furthermore know how to apply this knowledge. Consequently, the usage of this strategy involves considerably more mental effort than the other strategies.

Abele and von Davier (2019) expected that diagnosticians would use all three strategies to solve the diagnostic problems. The authors hypothesised that: (a) they would observe all three strategies (see Table 8.1); (b) the computer-based and the case-based strategies are employed more often than the model-based strategy, regardless of the problem difficulty, due to their relatively low mental effort; and (c) test takers use the case-based and computer-based strategies more often for simpler problems, as opposed to difficult problems, because simple problems are often familiar and can be solved using the expert system (Nickolaus et al., 2012). In contrast, test takers were expected to apply the model-based strategy more often when working on the difficult diagnostic problem compared to the simple problem.

8.3.3 Log Data Analysis and Selected Results

The log data resulting from the computer-based assessment provide a documentation of the apprentices' interactions with the car simulation for each individual and diagnostic problem. For this, the authors chose a binary scoring: if an apprentice displayed a critical information behaviour or test behaviour during the problem-solving process, the behaviour was scored as 1, otherwise as 0. The scoring was realised by applying a computer algorithm to the data, which is less prone to error than manual coding. The scoring procedure was performed for each diagnostic problem. In order to test their hypotheses, Abele and von Davier (2019) applied a diagnostic classification model to the log data. This model was selected from several psychometric models, as it was considered best suitable to describe the observed data.[2] Diagnostic classification models refer to a group of confirmatory multidimensional latent-variable models with categorical latent variables (e.g., Junker & Sijtsma, 2001; von Davier et al., 2008). These evaluate how multiple skills (here: strategies) are related to the observed outcome (here: problem-solving process data contained in the log data). With the help of these models, apprentices can be classified into *a priori* defined trait profiles (here: strategy profiles). As there are three binary strategy skill variables that represent the application of the computer-based, case-based and model-based strategies, the discrete distribution of the strategy types has $2^3 = 8$ potential outcomes. Importantly, the strategies are not mutually exclusive. This means that subjects may have applied more than one strategy, or, in other words, that their observed problem-solving behaviour may be compatible with more than one strategy.

Table 8.2 shows the distribution of the three strategies for the sensor problems. Taken together, for the easy problem, the case-based strategy was most frequently

[2] In addition to the diagnostic classification model, two other models (unidimensional item response theory and latent class analysis) were applied to the data. Model selection was based on information criteria that consider both model complexity (parsimony) and the probabilities of the data within the model (model fit). For a detailed description of the model selection procedure, see Abele and von Davier (2019).

Table 8.2 Distribution of strategy types for the easy and the difficult sensor problem. Multiple strategies or no strategy could be indicated, depending on the observed behaviour of diagnosticians

Computer-based strategy (Follow instructions of computer-based expert system)	Case-based strategy (Procedure based on experience with similar problems)	Model-based strategy (Derive hypotheses systematically based on mental models)	Proportion of participants using the strategy (%)	
			Easy problem	Difficult problem
–	–	–	22.4	22.7
+	–	–	14.1	16.8
–	+	–	37.0	25.9
+	+	–	23.3	19.2
–	–	+	0.7	4.1
+	–	+	0.4	3.0
–	+	+	1.2	4.6
+	+	+	0.7	3.4

Adapted from Abele and von Davier (2019, p. 481)

used by the respondents (37.0 + 23.3 + 1.2 + 0.7 = 62.2%). The computer-based strategy was applied by 38.5% (14.1 + 23.3 + 0.4 + 0.7). The model-based strategy was only applied by 3% (0.7 + 0.4 + 1.2 + 0.7). For 22.4% of the test takers, no strategy was assigned, meaning that none of them could be assigned to any of the predefined strategies. For the difficult sensor problem, the proportions were different, with a higher percentage of respondents using the model-based strategy (15.1%) and a lower percentage using the case-based strategy (53.1%) compared to the simpler problem. The computer-based strategy was applied by 42.4%, and in 22.7% of the cases, no strategy was assigned.

8.3.4 Discussion

The reported results are largely in line with the hypotheses. As assumed by the first hypothesis, the authors were able to differentiate between three *a priori* defined diagnostic strategies based on the log data. However, it is important to note that a substantial part of the test takers' observable behaviour could not be assigned to any of the strategies. Whether this indicates an unsystematic problem-solving procedure, or whether these test takers followed a strategy that was simply not captured within this study, requires further research. In line with the second hypothesis, diagnosticians applied the case-based and the computer-based strategies more often than the model-based strategy irrespectively of the problem difficulty. This can be explained by the higher mental effort of the model-based strategy in comparison to the other strategies. As predicted by the third hypothesis, the results suggest a shift in strategy from the easy to the difficult sensor problem. Compared to the easier diagnostic problem, more test takers seem to apply the model-based strategy when confronted with the difficult diagnostic problem. Contrary to the authors' assumptions, the results show that the proportion of diagnosticians using the computer-based strategy increased from the easy to the difficult sensor problem. One explanation for this finding could be that test takers who find a problem difficult are more likely to resort to an external source, as they may recognize that the task is beyond their ability and they cannot recall appropriate hypotheses from their memory (case-based), let alone generate them (model-based). Importantly, the study focused on the assessment of different diagnostic strategies without considering performance outcomes. The conducted analyses only allow for assertions about the usage of the strategies. For further insights, future analyses would need to investigate which strategy is associated with which solution probability for a specific problem. In addition, the study also highlights some specific challenges and inherent limitations that are associated with the analysis of computer-generated log data. These challenges are addressed in the next section.

8.4 Lessons Learned

In this chapter, we explored the value of process data from computer-based assessments for the measurement of professional competences and its implications for professional learning and development. Log data analyses can reveal specific activity patterns that are opaque when looking only on task outcomes, and shed light on process-specific aspects such as problem-solving efficiency and problem-solving strategies. Understanding differences in the task completion process can provide a basis for integrating process measures from log data into the measurement of professional competence, thereby paving the way for a deeper understanding of professional learning and development. However, competence measurement involving log data has some specific challenges, ranging from assessment, pre-processing, and analysis to the interpretation of the data, and is only at a preliminary stage in the field of VET.

8.4.1 Assessment and Pre-processing of Log Data

Whether the analysis of log data is useful depends largely on the concrete research goal (e.g., recording of distinct strategies, relation between strategy use and solution success, identification of efficiency barriers). Formulating an assessment objective and describing the type of log data needed to answer the research question (e.g., keystrokes, zooming) should therefore be at the beginning of the research process. Besides, it must be ensured that the computer system is able to capture the respective log data. In VET, computer simulations are more and more used as test and learning environments, as they allow for a realistic representation of professional problems. Simulation-based test environments authentically reproduce the complexity of professional problems and thus enable a high external validity of the competence measurement. However, highly interactive tests, such as simulations, usually offer test takers a large scope of actions and functionalities. Consequently, a variety of solution patterns exists and inferring clear competence assertions from the resulting log data is challenging (Kögler et al., 2020). Therefore, it is important to define which information from the log files is really needed to answer a specific research question. In alignment with this, the log data may be filtered to extract only the relevant data. The more assumptions about the data we have in advance, the more specifically we can proceed during this stage. But even though the selection of single log entries may be theoretically justified, it must be emphasised that this pre-processing step involves ignoring much potentially important information from the log files. This implies that many student-task interactions are not considered at all. Moreover, focusing on isolated events and choosing a binary scoring (as done by Abele & von Davier, 2019) only takes into account whether a certain action was shown or not. However, the significance of a specific student-task interaction may depend on the timing, frequency, duration or order in which that action was performed (Kögler et al., 2020). Studying event sequences (e.g., n-grams) and considering temporal dependencies between events could provide more detailed information about the problem-solving process and serve as more precise strategy indicators.

8.4.2 Analysis of Log Data

We presented two main approaches to the analysis of log data: (a) theory-driven and (b) data-driven. (a) Theory-driven approaches imply the assumption that observable behaviour stored in computer-generated log files is useful to test theories (e.g., on behaviour relevant to solve problems in professional contexts) and to assess individual characteristics (e.g., diagnostic strategies). These procedures require *a priori* assumptions about the log data. For instance, Abele and von Davier (2019) used a theory of the diagnostic problem-solving process (Abele, 2018) to model different problem-solving strategies and to define which behaviours from the log files are relevant for the distinct strategies. (b) When there is a lack of theoretical assumptions about the data, prior data-driven analysis techniques can help uncover solution-relevant activity patterns in the log data. Subsequent hypotheses-testing analyses can be conducted to empirically validate initial exploratory findings. Hence, theory-driven and data-driven approaches do not have to be seen as opposing and mutually exclusive approaches. Instead, they may be combined and even complement each other (Kögler et al., 2020).

In addition to deciding on a broader analytical approach, a statistical procedure must be selected for the analysis of the log data. We introduced a selection of procedures, such as n-grams (e.g., He & von Davier, 2016), k-means cluster algorithms (e.g., Tóth et al., 2013), and diagnostic classification models (for an overview, see von Davier & Lee, 2019). A detailed description of each of these techniques would go beyond the scope of this chapter.

8.4.3 *Interpretation of Log Data*

As emphasised in this chapter, conclusions about cognitive processes based on behaviour that is reconstructed from log data are always interpretative (Rausch et al., 2017). This is even the case when the behavioural indicators are developed on the basis of sound theoretical assumptions. For instance, when looking at the frequency of student-task interactions during the performance of a problem-solving task (operationalised as the number of keystrokes), this indicator can have different meanings. For high frequencies, it may represent disorientation, whereas low frequencies may correspond to task adequate problem-solving behaviour (Naumann et al., 2014). This illustrates the fact that the information contained in log files often cannot be interpreted unambiguously.

Our own study, presented above, further exemplifies the ambiguities in interpreting log data. The behaviour we considered critical to problem-solving success was assessed via computer-generated log files, taking into account, for example, whether a circuit diagram was opened or left closed. However, retrieving information material does not necessarily mean that the critical information has really been gained by the student (Abele et al., 2017). While reducing the problem-solving process on observable behaviour, mental processes are inferred. This indirect procedure must always be kept in mind when interpreting the results of log data analyses. Combining log data with other process measures – such as the fixation time revealed by eye tracking or think-aloud protocols – can help to solve the issue of unclear interpretation. These measures could substantiate the log file information, otherwise highly reliant on inference, thereby giving a more differentiated picture of the task completion process and increasing the internal validity of the competence measurement (Kögler et al., 2020).

8.5 Outlook

In addition to a more extensive assessment of already established constructs, such as diagnostic problem solving, the innovations in the field of computer-based testing and the integration of process data enable the measurement of completely new constructs that were difficult to assess using classical test formats (de Klerk et al., 2012). Among these constructs is collaborative problem-solving, which forms part of the so-called twenty-first century skills (e.g., Andrews-Todd et al., 2018; Dingler et al., 2017; Hesse et al., 2015; Pöysä-Tarhonen et al., 2018). Despite its high practical relevance across a range of professional domains, computer-supported collaborative problem solving is still an under-represented field in vocational educational research (Schwendimann et al., 2018) that could benefit from further methodological developments with regard to process data. Collaborative tasks require not only interactions between single test takers and the task, but also interactions between different test takers. The communication protocols can be stored in log files and analysed together with student-task interactions.

Beyond the assessment of well-established as well as new competence constructs, the analysis of computer-based log data holds potential for learning and instruction. For instance, knowing what characterises successful collaborative problem-solving processes is the basis for developing and evaluating interventions to promote this competence and thus for successful learning. Uncovering specific behaviours associated with successful or unsuccessful performance (e.g., efficiency barriers, strategy usage) that are not detectable when looking solely at task outcomes is useful for improving and developing interventions for VET that address the specific needs of students during the learning process (Abele et al., 2017; Stadler et al., 2019). Furthermore, log data are particularly valuable for formative assessments that take into account the student's professional development throughout the learning process (de Klerk, 2016). Additionally, analysing log data can help to diagnose individual needs in the learning process and professional contexts. This information can then be used for formative feedback: learners can be given feedback during VET that is in line with their needs.

Understanding interindividual process differences via the analysis of computer-generated log data paves the way for a more process-oriented approach to professional competence measurement. Supplementing merely product-oriented competence measurement with a process component will depend largely on whether the challenges associated with this type of data can be overcome. In this way, the possibilities of log data analysis for VET could be fully exploited and contribute to a more comprehensive understanding of professional learning and development.

References

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning, 11*(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier & Y. S. Lee (Eds.), *Handbook of diagnostic classification models* (pp. 461–488). Springer International Publishing. https://doi.org/10.1007/978-3-030-05584-4_22

Abele, S., Walker, F., & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker [Time-efficient and reliable diagnostics of professional problem-solving skills for trainees as automotive mechatronics technicians]. *Zeitschrift für Pädagogische Psychologie, 28*(4), 167–179. https://doi.org/10.1024/1010-0652/a000138

Abele, S., Ostertag, R., Peissner, M., & Schuller, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess: Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg [An eye-tracking study of the diagnostic problem-solving process: The importance of information representation for diagnostic problem-solving success]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 1*(113), 86–109.

Andrews-Todd, J., Forsyth, C., Steinberg, J., & Rupp, A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. In *Proceedings of the 11th International Conference on Educational Data Mining, USA* (pp. 239–245).

Baethge, M., & Arends, L. (Eds.). (2009). *Feasibility study VET-LSA: A comparative analysis of occupational profiles and VET programmes in 8 European countries–international report*. Federal Ministry of Education and Research. http://hdl.voced.edu.au/10707/236812

Beck, K., Landenberger, M., & Oser, F. (Eds.). (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based competence measurement in vocational education: Results from the BMBF funding initiative ASCOT]* (Vol. 32). W. Bertelsmann.

Damashek, M. (1995). Gauging similarity with n-grams: Language-independent categorization of text. *Science, 267*(5199), 843–848.

de Klerk, S. (2016). *Multimedia-based performance assessment in Dutch vocational education*. [Doctoral dissertation, Universiteit Twente]. https://doi.org/10.3990/1.9789036539975

de Klerk, S., Eggen, T. J. H. M., & Veldkamp, B. P. (2012). An overview of innovative computer-based testing. In T. J. H. M. Eggen & B. P. Veldkamp (Eds.), *Psychometrics in practice at RCEC* (pp. 151–161). RCEC. https://doi.org/10.3990/3.9789036533744.ch13

Dingler, C., Von Davier, A. A., & Hao, J. (2017). Methodological challenges in measuring collaborative problem-solving skills over time. *Research on Managing Groups and Teams, 18*, 51–70. https://doi.org/10.1108/S1534-085620160000018003

Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., & Tóth, K. (2017). Relating product data to process data from computer-based competency assessment. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education* (Vol. 40, pp. 407–425). Springer. https://doi.org/10.1007/978-3-319-50030-0_24

Greiff, S., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 157–163). Office for Official Publications of the European Communities. https://doi.org/10.11588/heidok.00015502

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers and Education, 91*(1), 92–105. https://doi.org/10.1016/j.compedu.2015.10.018

Gschwendtner, T., Abele, S., & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern [Computer-simulated work samples: A validation study using the example of error diagnosis performance of automotive mechatronics technicians]. *Zeitschrift für Berufs-und Wirtschaftspädagogik, 105*(4), 557–578.

He, Q., & Von Davier, M. (2016). Analyzing process data from problem-solving items with N-grams: Insights from a computer-based large-scale assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 750–777). IGI Global. https://doi.org/10.4018/978-1-4666-9441-5.ch029

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR), 31*(3), 64–323.

Jonassen, D. H. (2010). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge. https://doi.org/10.4324/9780203847527

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272. https://doi.org/10.1177/01466210122032064

Kane, M. T. (1992). The assessment of professional competence. *Evaluation & the Health Professions, 15*(2), 163–182. https://doi.org/10.1177/016327879201500203
Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2017). Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Transactions on Learning Technologies, 10*(2), 140–153. https://doi.org/10.1109/TLT.2015.2513387
Kögler, K., Rausch, A., & Niegemann, H. (2020). Interpretierbarkeit von Logdaten in computerbasierten Kompetenztests mit großen Handlungsräumen [Interpretability of log data in computer-based competency tests with large scopes of action]. B*wp@ Profil 6: Berufliches Lehren und Lernen: Grundlagen, Schwerpunkte und Impulse wirtschaftspädagogischer Forschung. Digitale Festschrift für Eveline Wuttke zum 60. Geburtstag* (pp. 1–21). https://www.bwpat.de/profil6_wuttke/koegler_etal_profil6.pdf
Konradt, U. (1995). Strategies of failure diagnosis in computer-controlled manufacturing systems: Empirical analysis and implications for the design of adaptive decision support systems. *International Journal of Human – Computer Studies, 43*(4), 503–521. https://doi.org/10.1006/ijhc.1995.1057
Kroehne. (2021). *LogFSM: Analysis of log data using finite-state machines.* https://logfsm.com
Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika, 45*(2), 527–563. https://doi.org/10.1007/s41237-018-0063-y
Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16*(11), 866–870.
Lajoie, S. P., Poitras, E. G., Doleck, T., & Jarrell, A. (2015). Modeling metacognitive activities in medical problem-solving with BioWorld. In A. Peña-Ayala (Ed.), *Metacognition: Fundaments, applications, and trends* (pp. 323–343). Springer. https://doi.org/10.1007/978-3-319-11062-2_13
Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences, 22*(4), 511–520. https://doi.org/10.1080/10508406.2013.840466
Napa Scollon, C., Prieto, C. K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In *Assessing well-being* (pp. 157–180). Springer.
Naumann, J., Goldhammer, F., Rölke, H., & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiereichen Umgebungen. Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen [Successful problem-solving in technology-rich environments. Interdependencies between interaction steps and task requirements]. *Zeitschrift für Pädagogische Psychologie, 28*(4), 193–203.
Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A., & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich-technischen Ausbildungsberufen –Modellierung, erreichte Niveaus und relevante Einflussfaktoren. [Subject-specific problem-solving ability in industrial-technical training occupations–modeling, levels achieved and relevant influencing factors.]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 108*(2), 243–272.
Organisation for Economic Co-operation and Development (OECD). (2019). *Beyond proficiency: Using log files to understand respondent behaviour in the survey of adult skills*. OECD Publishing. https://doi.org/10.1787/f25ca0cd-en
Pekruna, R. (2020). Commentary: Self-report is indispensable to assess students' learning. *Frontline Learning Research, 8*(3), 185–193. https://doi.org/10.14786/flr.v8i3.637
Perez, R. S. (2012). A view from troubleshooting. In M. U. Smith (Ed.), *Toward a unified theory of problem solving: Views from the content domains* (pp. 127–166). Routledge.
Pöysä-Tarhonen, J., Care, E., Awwal, N., & Häkkinen, P. (2018). Pair interactions in online assessments of collaborative problem solving: Case-based portraits. *Research and Practice in Technology Enhanced Learning, 13*(12). https://doi.org/10.1186/s41039-018-0079-7
Rausch, A., Kögler, K., Frötschl, C., Bergrab, M., & Brandt, S. (2017). Problemlöseprozesse sichtbar machen: Analyse von Logdaten aus einer computerbasierten Bürosimulation

[Uncovering problem-solving processes: Analysis of log data from a computer-based office simulation]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 113*(4), 569–594.
Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.
Reis Costa, D., & Leoncio, W. (2019). *LOGAN: Log file analysis in international large-scale assessments.* https://CRAN.R-project.org/package=LOGAN
Richter, T., Naumann, J., Brunner, M., & Christmann, U. (2005). Strategische Verarbeitung beim Lernen mit Text und Hypertext [Strategic processing in learning with text and hypertext]. *Zeitschrift für Pädagogische Psychologie, 19*(1/2), 5–22. https://doi.org/10.1024/1010-0652.19.12.5
Sangmeister, J., Winther, E., Deutscher, V., Bley, S., Kreuzer, C., & Weber, S. (2018). Designing competence assessment in VET for a digital future. In D. Ifenthaler (Ed.), *Digital workplace learning: Bridging formal and informal learning with digital technologies* (pp. 65–92). Springer. https://doi.org/10.1007/978-3-319-46215-8_5
Schaper, N., Hochholdinger, S., & Sonntag, K. (2004). Förderung des transfers von Diagnosestrategien durch computergestütztes Training mit kognitiver Modellierung [Promotion of the transfer of diagnostic strategies through computer-based training with cognitive modelling]. *Zeitschrift für Personalpsychologie, 3*(2), 51–62. https://doi.org/10.1026/1617-6391.3.2.51
Schleicher, A. (2008). PIAAC: A new strategy for assessing adult competencies. *International Review of Education, 54*(5), 627–650. https://doi.org/10.1007/s11159-008-9105-0
Schwendimann, B. A., De Wever, B., Prof, R. H., & Cattaneo, A. A. P. (2018). The state-of-the-art of collaborative technologies for initial vocational education: A systematic literature review. *International Journal for Research in Vocational Education and Training, 5*(1), 19–41. https://doi.org/10.13152/IJRVET.5.1.2
Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.
Stadler, M., Fischer, F., & Greiff, S. (2019). Taking a closer look: An exploratory analysis of successful and unsuccessful strategy use in complex problems. *Frontiers in Psychology, 10*, 777. https://doi.org/10.3389/fpsyg.2019.00777
Stelter, A., Goldhammer, F., Naumann, J., & Rölke, H. (2015). Die Automatisierung prozeduralen Wissens. Eine Analyse basierend auf Prozessdaten [The automation of procedural knowledge. An analysis based on process data]. In J. Stiller & C. Laschke (Eds.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2015. Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung* (pp. 111–131). Lang.
Tóth, K., Rölke, H., Goldhammer, F., & Kröhne, U. (2013). *Investigating students' ICT-skills with process data*. Paper presented at the DAILE13 workshop, Villard-de-Lans, France.
Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1–10. https://doi.org/10.2307/1129583
Veenman, M. V. J., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. P. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences, 29*, 123–130. https://doi.org/10.1016/j.lindif.2013.01.003
von Davier, M., & Lee, Y. S. (2019). *Handbook of diagnostic classification models*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05584-4
von Davier, M., DiBello, L., & Yamamoto, K. (2008). Reporting test outcomes using models for cognitive diagnosis. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 151–176). Hogrefe. https://doi.org/10.1002/j.2333-8504.2006.tb02034.x

Weinert, F. E. (2002). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit [Comparative performance measurement in schools – A controversial matter of course]. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (2nd ed., pp. 17–31). Beltz.

Winther, E. (2010). Kompetenzmessung in der beruflichen Bildung [Measurement of competence in vocational education]. In *BWP Berufsbildung in Wirtschaft und Praxis*.

Wüstenberg, S., Stadler, M., Hautamäki, J., & Greiff, S. (2014). The role of strategy knowledge for the application of strategies in complex problem solving tasks. *Technology, Knowledge and Learning, 19*(1–2), 127–146. https://doi.org/10.1007/s10758-014-9222-8

Zhang, L. J., & Zhang, D. (2019). Think-aloud protocols. In *The Routledge handbook of research methods in applied linguistics* (pp. 302–311). Routledge.

Zirkel, S., Garcia, J. A., & Murphy, M. C. (2015). Experience-sampling research methods and their potential for education research. *Educational Researcher, 44*(1), 7–16.

Luca Spliethoff (MSc) studied psychology at the Technical University of Dresden. Between September 2019 and September 2021, she was a research associate at the Chair of Vocational Education. As part of her work, she investigated how collaborative diagnostic competences of car mechatronics apprentices can be measured with the help of computer-generated process data.

Stephan Abele (PhD) is Professor of Vocational Education at the Technical University of Dresden. His research concentrates on vocational learning and instruction and the computer-based measurement of vocational competences. One focus is on competence measurement with process data, including from computer-generated log files.

Chapter 9
Investigating Interaction Dynamics: A Temporal Approach to Team Learning

Lida Z. David, Maaike D. Endedijk, and Piet Van den Bossche

Abstract Teams are at the core of every organisation, composed of individuals who continuously collaborate, exchange knowledge and ideas, and constantly learn from one another through formal or informal learning experiences. Team learning is therefore a continuously changing phenomenon that develops and evolves over time as teams interact. In this chapter, we aim to promote the investigation of team learning as a temporal phenomenon, and suggest that its temporality can be captured through team interaction dynamics, defined as continuously changing patterns of micro-behaviours that emerge and evolve as teams operate. We set three key steps for initiating and leading research that captures temporality: (a) identifying the interaction dynamics of interest, (b) figuring out the best way to collect and code these, and finally (c) choosing an analysis technique that helps capture continuously and sequentially unfolding patterns. We offer some 'food for thought' on interaction dynamics that relate to team learning and the added value of investigating them, and present some existing data collection and coding methods. We finally propose a framework for choosing an appropriate analysis technique based on the dynamic output that each analysis generates.

Keywords Team learning · Temporal phenomenon · Interaction dynamics · Patterns · Analysis techniques · Emergent states

---

L. Z. David · M. D. Endedijk (✉)
Department of Learning, Data-analytics, and Technology, University of Twente, Enschede, The Netherlands
e-mail: l.david@utwente.nl; m.d.endedijk@utwente.nl

P. Van den Bossche
Department of Educational Research and Development, School of Business and Economics, Maastricht University, Maastricht, The Netherlands

Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium
e-mail: piet.vandenbossche@uantwerpen.be

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_9

9.1 Introduction

Learning does not occur solely during the short period of attending school and other formal educational programs, but becomes a life-long process, needed to keep up with the rapid development and transformation of the world around us (Heaphy et al., 2018). Organizational structures become increasingly interconnected in complex and oftentimes unfamiliar ways (e.g. incorporation of remote teamwork in almost every workspace due to Covid-19), and individuals are called to learn from one another by experimenting and exchanging experiences and knowledge, while also constantly giving and receiving feedback from each other (Grant & Ashford, 2008; Van de Wiel et al., 2011). Consequently, teamwork becomes a continuous learning experience, a 'phenomenon' unfolding over time, which, as any temporal phenomenon (Roe, 2008), unravels through the implicit or explicit exchange of information as team members collaborate and solve problems together. This phenomenon of team learning through interaction has been strongly associated with innovation and performance in the workspace (Baker et al., 2006; Fay et al., 2015; Gambi do Nascimento et al., 2020).

Various concepts related to team learning, such as information exchange, help-giving and -seeking, and feedback-giving and -seeking have been associated with team performance (Gerken et al., 2016; Van der Rijt et al., 2013). However, while these concepts suggest a dynamic exchange of information, they are mostly studied and represented in static ways, through surveys, interviews, and other retrospective, self-reported data (Onnela et al., 2014). These methods not only lack objectivity and are prone to self-biases when it comes to measuring concepts such as the aforementioned, but are also limited in capturing the dynamic nature underlying team learning interaction processes.

Having conceptualized team learning as a process that takes place during interaction and over time, necessitates the investigation of how professionals actively construct, develop, and transform their interpersonal relationships throughout collaboration (Heaphy et al., 2018; Tynjälä, 2008). In other words, treating team learning as a temporal phenomenon calls for methods to grasp and analyse the continuously unfolding dynamic patterns of interaction between team members, thus capturing how learning unfolds over time. To promote the integration of the dynamic nature and temporality in researching team learning, we here provide a definition of interaction dynamics, the central tenets that can be used for capturing team temporal phenomena. We define interaction dynamics as *"micro-behaviours forming patterns that continuously change, emerge and evolve throughout a team's lifecycle of interaction"*.

Studying interactions in a dynamic way enables researchers to capture 'what happens' rather than 'what is' (see Roe, 2008). Through investigating real team interaction, research questions can evolve from how team learning behaviours are *perceived*, to how team learning behaviours actually *are*. For example, one may look for mechanisms and characteristics associated with team learning and performance under different contexts or interventions, and how these are affected, facilitated or hindered by internal or external variables such as shifts in workload, team membership changes, stress or unexpected events.

One might argue that dynamics can also be researched through higher-level team behaviours that are captured, for example, through observational research, thus also offering a means of understanding what happens. However, such a methodology might not offer the possibility of studying the connections, relations, and patterns between the exhibited behaviours to the level of scrutiny that would enable the detection of emergent states or changes in these states over time. By considering interaction dynamics as micro-behaviours that continuously change, and through using the appropriate methodological approaches to capture micro-behaviour patterns, researchers enable a detailed, effective, and efficient investigation of emergent states and otherwise undetectable details of team learning. What is more, such states and details can be captured and traced in both small-scale and more complex, interconnected teams and systems. For example, one can study a single small-scale team during one learning event (e.g. interaction within one flight controller team in NASA's Mission Control), but can also study, on a micro-level of detail, the interaction of larger-scale systems (e.g. interaction between all flight controllers in NASA's Mission Control) or even entire organisations (e.g. interaction between NASA's Mission Control, Astronauts, other support staff across all NASA field centres in the U.S.), and how these interact and learn from each other during the same event. Understanding team learning in such a micro-level of granularity helps develop and establish both formal and informal learning opportunities and interventions that can be introduced in the right place, at the right time. It is therefore important to find methodologies and develop apparatus that enable the procurement and analysis of such interaction dynamics.

Exemplary, technological developments enabling the collection of fine-grained interaction dynamics are sociometric badges (Kim et al., 2008, 2012), body sensors (Dong et al., 2012), or video and audio recording of team communication and behaviours (Klonek et al., 2020). Attempts are being made to incorporate these technologies in research across the domains of organisational psychology (Klonek et al., 2019), business management (Lehmann-Willenbrock & Allen, 2018), and educational sciences (Lämsä et al., 2021). However, investigation of interaction dynamics remains scarce due to the lack of accessibility to data collection or analysis methods, and the factual or perceived complexity of utilising the respective data analysis techniques.

It is our ambition to guide our readers towards a more practical and seamless application of researching and analysing fine-grained interaction data for modelling and understanding professional team learning. This chapter provides a rationale for designing and carrying out interaction research, by following the same order as this of the key steps that should be adopted by researchers: (a) defining the interaction dynamics of interest with respect to team learning, (b) choosing an appropriate data collection method, and (c) choosing an analysis technique that captures temporality.

9.2 Team Learning and Interaction Dynamics

A first step in designing research involving interaction dynamics is defining which patterns are relevant to our phenomenon of interest: team learning. It is thus vital to first envisage what team learning entails.

Research on team learning picked up after Edmondson (1999) defined the phenomenon as a behavioural cyclical process of seeking, collecting, experimenting, reflecting, and discussing information. The term has since then received multiple definitions, a lot of them embracing the idea that team learning is a temporal phenomenon, to be studied at the team-level, involving different processes, and leading to different outcomes (Decuyper et al., 2010; Wiese & Burke, 2019). Understanding how and which interaction dynamics can be used in the investigation of team learning involves first identifying what are the possible team learning processes whose interaction can be useful to researchers. An exemplary model is the integrative model of team learning by Decuyper et al. (2010). It delivers a model on team learning recognizing the importance of emergent states and positioning these in relation to team learning processes. Hereby, this model is exemplary in modelling teamwork in general, and team learning specifically (Van den Bossche et al., 2022).

Historically, research in team learning has been strongly influenced by an input-process-output model, where team processes describe the mechanism by which individual team members resolve tasks (Dillenbourg, 1999; Kozlowski, 2015). However, it was made clear that it is necessary to differentiate between the various types of process variables (Marks et al., 2001). Important variables such as group potency or cohesion, do not denote interaction processes. Decuyper et al. (2010) proposed to call them 'emergent states', constructs that describe cognitive, motivational and affective states of the team, and these are different from the team interaction itself. As such, emergent states do not represent team interactions, they are products of them and become new inputs to subsequent processes. For example, teams with low psychological safety (as an emergent state) may be less willing to share knowledge (as a process), which in turn may impact psychological safety.

In relation to the team learning process, the team learning model by Decuyper et al. (2010) presents seven categories of team learning processes: (1) sharing, (2) co-construction and (3) constructive conflict; (4) team reflexivity, (5) team activity and (6) boundary crossing; and (7) storage and retrieval. These team learning processes take the team towards adaptive, generative or transformative learning. These outputs are sometimes immediately observable in changing team performance. However, often they remain conceptual, as changes in the teams' capability to act differently. With regard to the emergent states as proximal outcomes of the team processes, this model points to exemplary variables such as shared mental models, team psychological safety, group potency, team efficacy, and cohesion.

Up until this point, we have presented the processes or concepts related to team learning without really focusing on their dynamic nature. It is important to understand here that all aforementioned processes are tied to time, meaning that they emerge, evolve, and develop differently throughout a team's single performance episode or entire life cycle. Raes et al. (2015) found that effective team learning is bound to the ability of a team to engage in sharing through "a sequence of successive and constructive verbal behaviours that construct meaning" (p. 491). Therefore, to investigate team learning dynamically, it is important to consider the interactions of team members related to these processes as they unfold during teamwork, revealing with temporal detail the structural combinations of behaviours desirable for effective team learning. But how can we do that?

It is important to first understand what different interaction dynamics exist, before connecting them to team learning.

9.2.1 What Are Interaction Dynamics?

Earlier, we defined interaction dynamics as micro-behaviours forming patterns that continuously change, emerge, and evolve throughout a team's lifecycle of interaction. Given the broad range of micro-behaviours that exist, we categorise them as either (a) feature-based, or (b) essence-based. The various combinations and development of both types of micro-behaviours into different patterns throughout interaction comprise interaction dynamics.

(a) *Feature-based micro-behaviours* include behaviours or events defined entirely by their inherent features (form, shape, proportion), and their classification as such depends on their concrete characteristics, independent of any interpretations. For example, speech acts (who talks to whom, for how long, how loud, etc.), physical proximity, movement energy (Onnela et al., 2014), or even positioning in the room (Ciolek & Kendon, 1980) are concrete behaviours, largely defined by specific inherent features, and may reveal a development or change in interaction (De Ruiter & Loth, 2016). Feature-based micro-behaviours may also include manipulation of objects or interactions with interfaces, such as clicks on a display, or eye gaze at a screen.

(b) *Essence-based micro-behaviours* consist of any behaviour or event for which the content of information is indispensable for its definition, and some kind of meaning attribution is necessary to classify it as such. Essence-based behaviours are of more abstract nature and constitute higher-level behaviours than feature-based ones, valuing the quality of information over the lower-level, concrete features of the behaviour itself. For example, 'providing positive feedback', 'suggesting', or 'directing', are essence-based micro-behaviours, since capturing such behaviours involves not merely collecting speech acts, but rather providing meaning to the content of the information exchanged in these acts.

We argue here that it is not only about which micro-behaviours are present during team interaction, but also how these micro-behaviours interconnect forming different interaction dynamics, and how various patterns change depending on environmental influences. Even though the stability, growth, and recurrence of structural patterns underlying temporal phenomena develop and change over time (Roe, 2008), little is known about how these unfold and relate to the phenomenon of team learning. Dynamic patterns of interaction can give insights into *how* teams learn (e.g. pattern combinations, emergence of new patterns, etc), or *when* teams learn (e.g. moments where emergence of different patterns occurs), thus helping to inform the design of interventions for the promotion of team learning processes.

For example, spotting the development of learning triggers (Wiese & Burke, 2019) can help in facilitating learning or manoeuvring away from delays in the learning process. Learning triggers are presented in the work of Wiese and Burke (2019) as catalyst events for team learning, where exposure to an event causes a spark or a disruption of the learning process. For instance, communication patterns involving positive statements by the team leader followed by agreement from all team followers might cause an increase in a team's knowledge state, which may be temporarily hindered when patterns revealing differences in opinion from the followers emerge. In other words, studying patterns of behaviours temporally can help in answering research questions around what mechanisms pose "learning triggers" (e.g. agreement between leader-follower), as well as whether and how the emergence or disappearance of these learning triggers is associated with specific contextual or procedural events during team meetings (e.g. disagreement between leader-follower after a leader makes a mistake).

It is therefore important not only to associate micro-behaviours to team learning, but also investigate their dynamic nature in greater detail. Research relating micro-behaviours to phenomena or constructs that are closely tied to team learning, such as sharing or co-construction (Decuyper et al., 2010; Van den Bossche et al., 2011; Wiese & Burke, 2019), set a strong starting point for developing assumptions on dynamics that facilitate or impede learning. By doing that, researchers open the door towards the investigation of questions related to the exact moments at which team learning is promoted or impeded (*when*), as well as to the mechanisms by which this promotion or decline is manifested (*how*).

In the following sections, we examine how team learning can be conceptualized through feature and essence-based micro-behaviours, to inspire our readers to choose the appropriate micro-behaviours for their research. We also explore through examples the added value that interaction dynamics captured via these micro-behaviours may have on team learning research.

9.2.2 Connecting Team Learning Processes to Interaction Dynamics

Any research strategy that attempts to understand temporal phenomena such as team learning involves first identifying which phenomena should be studied, before exploring their temporal relationships and their long-term stability and change (Roe, 2008). Therefore, any sampling decisions on how and which data to gather should be directly related to which interaction dynamic is of interest to the research's aim.

As discussed above, both essence- and feature-based dynamics can provide information on team phenomena, and both can give insights into the question of 'what happens' during interaction. Depending on the chosen research question, researchers may opt to either focus on essence or feature micro-behaviours, or combine both and possibly study their interaction. For example, a researcher interested in understanding the process of how *sharing* unfolds during an event may investigate the essence-based patterns formed between different information sharing micro-behaviours (e.g. fact sharing, interpretation sharing, and projection sharing; see Uitdewilligen & Waller, 2018). In addition, that same researcher may choose to also assess how this process of sharing combines and evolves across team members, and so also investigate feature-based micro-behaviours of actor relationships (e.g. conversational turn-taking patterns across team members; see Gorman et al., 2019). Other feature-based interaction dynamics, such as information flow, flexibility, complexity, or high mimicry during learning (Kim et al., 2012), can also be used as proxies for understanding processes of team learning. Below, we provide certain pilar examples of research done in team interaction dynamics and connect these to the team learning emergent states and processes mentioned earlier.

An interesting example of research investigating the interaction dynamics of essence-based micro-behaviours was done by Kolbe et al. (2014) in healthcare, who researched sequences between essence-based behaviours revealing implicit coordination behaviours (e.g., providing assistance, giving information without request) or explicit ones (e.g. giving information upon request, instructing). They found that higher team performance is associated with significantly more patterns combining implicit coordination behaviours followed by explicit coordination behaviours. Even though their study was not directly associated to team learning, the behavioural patterns researched may also reveal interesting information with respect to our phenomenon of interest. More specifically, implicit behaviours are related to the existence of team mental models and trust (Burtscher et al., 2011; Rico et al., 2008), while explicit behaviours reveal opportunities for the development of shared understanding and co-construction (Salas et al., 2007). Therefore, understanding the dynamic interchange between the two types of behaviours reveals the undergoing structure through which teams inform, develop, and use their shared mental models throughout interaction.

Another example, this time using feature-based micro-behaviours to capture interaction dynamics comes from the research of Gorman et al. (2019), who investigated action teams by modelling the actors speaking at each second in a team meeting. The researchers calculated the recurrence of turn-taking patterns throughout teams' meetings, to determine the reorganisation of communication in response to environmental perturbations. Findings showed that novice teams were not as quick or as successful in reorganising their information as compared to experienced teams. This finding possibly reveals that reorganising information is not only an important skill for performance, but it is a skill revealing that preceding effective team learning has occurred.

Within the educational domain, interaction patterns were researched to identify phase transitions during collaboration. Ricca et al. (2019) coded feature-based turn-taking patterns of team members and how these reoccurred and changed throughout interaction. They also qualitatively marked the points of phase change within collaboration (e.g., problem scoping, generating ideas, reporting, testing) and found that the points where turn-taking patterns changed collided with the points of phase transitions they had noted, thus showing that turn-taking behaviour not only changes during interaction but is also a good indicator of how and when a team moves from one phase to another. Team learning research can use such indicators of phase transition to research the effectiveness of interventions or note the times at which a team requires an intervention.

Comparing interaction dynamic peaks using turn-taking behaviours is also a promising way of monitoring environmental changes (transitioning from task to task), and understanding team interaction structures that lead to increased effectiveness, cognitive stability, or flexibility (Stevens, 2012). Such interaction dynamics can also be used to identify boundary crossing, as patterns have been found to collapse when a team goes through a phase transition (Kugler & Turvey, 1987; Wiltshire et al., 2018).

Interaction dynamics may also be used to reveal the emergence of joint attention, which is useful for creating a common ground and engaging in active collaboration. For example, in aviation research, joint visual attention has been used to understand pilots' coordination strategy during teamwork. Gaze alignment was used to indicate engagement in joint activity or overlap in information acquisition, while misalignment indicated divergent activity and sharing of task-load (Gontar & Mulligan, 2016). The investigation of different eye-gaze patterns development during task-work can further reveal learning behaviours such as adaptive coordination through early acknowledgement of changing situational demands requiring joint or divergent activity.

We have provided some examples of how both feature- and essence-based micro-behaviours can be used to examine interaction dynamics and to understand processes and states related to team learning. We now move on to presenting some ways through which this data can be gathered and coded, before delving into the analysis techniques that enable the investigation of the dynamic nature of interactions.

9.3 Researching Interaction Dynamics: How Do I Gather Data?

Having first identified the interaction dynamic of interest to the research's aim, the second decision step of investigating interaction dynamics includes choosing a data gathering and coding method (Lehmann-Willenbrock & Allen, 2018) by assessing their appropriateness in capturing the corresponding temporal granularity of the phenomenon of interest.

Interaction dynamics may differ with respect to their temporal resolution, and some data collection and coding methods may lead to representation of the dataset that is of a higher or lower temporal resolution; that is, the detail at which the dataset is represented and the phenomenon at hand captured and broken down. High resolution refers to a more exhaustive representation of the dataset. For example, dynamics based on behavioural changes of milliseconds (e.g. change in eye movements, change in loudness level of speech) are of higher temporal resolution compared to dynamics based on higher-level constructs (e.g. communication types such as defending one's own position).

9.3.1 Data Gathering Considerations

There are multiple ways of gathering team interaction data. Direct observation is one means of gathering data, however, human observers may not be as fast or as reliable in recording the necessary behaviours as these are being observed, while semi- or fully-automated real-time data transcription or coding methods have only recently started getting validated (Bokhove & Downey, 2018). Other than direct observations, video and audio recording are so far two key means through which temporality in interaction can be captured, as they offer the choice to replay, check, and reuse of the data. Especially with the rise of small-scale devices, such as action cameras and 360-degree cameras, etc. this opens up a world of possibilities to record team meetings for longer periods of time, in an unobtrusive and low-cost manner. In addition, in the past decade, methods from computer science have been developed to unobtrusively capture and analyse interpersonal behaviour, also called social sensing (Schmid et al., 2015). As many smartphones and smartwatches already are equipped with camera, microphone, accelerometer and GPS, the challenge is not anymore on recording the data, but on developing the right data extraction methods (Schmid et al., 2015). Moreover, sensors fully dedicated towards recording (and analysing) feature-based micro-behaviours in social settings, such as the sociometric badges (Onnela et al., 2014), have been developed and are further discussed below.

9.3.2 Data Coding Considerations

Software developed for video and audio manipulation offer the ability to either transcribe and then code (e.g. Atlas.ti; AmberScript), or code directly on the video noting the temporal stamp of each behaviour (e.g. Observer-XT; Noldus et al., 2000). Coding schemes differ with respect to the number of behaviours they include, the extent to which these behaviours relate to the task at hand or to a specific phenomenon, and the level of abstraction (Waller & Kaplan, 2018). Coding of interaction data, especially when it comes to essence-based behaviours, may involve to some extent a reduction from the complexities of actual behaviours to broader categories, thus inhibiting an exhaustive representation of interaction, and following a top-down, theory-driven quantitative approach rather than a bottom-up exploration of interaction (Stivers, 2015).

It is important to consider the extent to which the research requires the identification and analysis of micro-behaviours with high or low temporal resolution. Depending on the research aims, the combination of essence-based coding schemes with feature-based data (e.g., speaker-receiver information) can also be used to account for any disregarded essence-based changes in interaction dynamics that might be overlooked. For example, a researcher can use essence-based micro-behaviours of problem solving processes (e.g. information provision, information request, option generation or solution evaluation) to spot phase transitions during team collaboration (Wiltshire et al., 2018). Transitions can also be spotted via coding and analysing actor turn-taking patterns across the meeting (Ricca et al., 2019).

Validated coding schemes that can help with the incorporation of interaction dynamics related to team learning are already available (see review of Brauner et al., 2018). For example, the Co-ACT coding scheme by Kolbe et al. (2013) can be used to research implicit and explicit coordination processes (Kolbe et al., 2014). The coding scheme of Raes et al. (2015) can be used to code different types of basic team learning behaviours with lower temporal resolution. The Act4Teams coding scheme of Kauffeld et al. (2018) includes four types of team interaction (problem-focused, procedural, socioemotional, and action-oriented) and especially the analysis of the problem-focused communication shows how teams share knowledge, generate new ideas, etc.

All aforementioned coding schemes can be applied to data records (video, audio, or text based) using software such as the Observer-XT (Noldus et al., 2000), Interact (Mangold, 2020), or CAT (Klonek et al., 2020). Other examples of coding schemes for micro-coding, as well as available software can be found in Waller and Kaplan (2018).

9.3.3 Towards a Combined Data Gathering and Coding Methodology

The coding process of datasets of high-temporal resolution can be very labour-intensive. So far, most research on team interaction dynamics involves manual coding of the data before analysis. However, more recently, technological innovations such as behaviour sensor systems, or machine learning algorithms (e.g. Bonito & Keyton, 2018) are attempting to ease the coding process by automating the acquisition of meaningful information. Here we exemplary introduce two options for (semi-)automated process of coding behaviours.

9.3.3.1 Behaviour Sensor Systems

Sociometric badges are a sensor technology device enabling the automatic collection and measurement of interactions (Kim et al., 2012; Olguín Olguín et al., 2009). The installed sensors in the device enable the collection of data identifying face-to-face interactions (infrared sensor), speech features (microphone audio-signal), and mimicry patterns (accelerometer). Sociometric badges offer a bundle of services since collection and coding are both done automatically via the badges. However, lately the downsides of these commercial all-in-one devices have come to light: as the underlying algorithms are company-secret information, researchers have tried to validate these sensors with disappointing results (Kayhan et al., 2018) next to the problem of malfunctioning badges and synchronization problems (Endedijk et al., 2018).

9.3.3.2 Supervised and Unsupervised Machine Learning

With digitalization pushing forward, automated coding solutions have also become available, such as automatic coding filters to code different communicative functions (Erkens & Janssen, 2008). More recently, advancements in automated content analysis have also started to incorporate supervised machine learning (SML) as a technique to automatically code transcribed datasets. SML is the process of training an algorithm to identify relationships between input (text) and output variables (codes) from a specific dataset, so well that the machine is then able to identify any relationships between text and codes in new datasets. For example, assume that you code one transcript, thus associating text content to a specific code, (e.g. a sentence including “I suggest we incubate her” is coded under the label “suggestion”). This association between content–code is used to train your algorithm to automatically spot content that falls under this code in other, unknown transcripts. For an illustration of how SML can be used in practice, see Bonito and Keyton (2018). Note that SML is a deductive, top-down process as it requires training algorithms based on existing codebooks.

Although more challenging, unsupervised machine learning can be used as an inductive, bottom-up process of identifying text clusters and structures within a dataset, enabling a researcher to choose the most interpretable set (Lambert, 2001) and thus offering the chance for exploratory research with no pre-existing codebook. Unsupervised machine learning can also be applied to datasets including feature-based micro-behaviours such as posture or eye gaze (Huang et al., 2019).

9.4 Researching Interaction Dynamics: How Do I Analyse My Data?

9.4.1 Developing a Framework for Choosing a Data Analysis Technique

Having defined interaction dynamics as patterns that continuously change, researchers are called to use analysis techniques that can help them capture this continuous, or near-continuous nature of the data they gather. Different techniques can be used, depending on the research's aim and purposeful insights. There is an increasing call for the incorporation of such techniques in research (David et al., 2021; Herndon & Lewis, 2015; Leenders et al., 2016). To ease their incorporation, we present our readers with the DATS framework (referring to Data Analysis Technique Selection); a framework for choosing an analysis technique based on two key aspects: the *units of analysi*s, and the *dynamic output generated*. The framework is described below and illustrated in Fig. 9.1.

The *unit of analysis,* referring to the key units or data points that are modelled in the analysis, can be:

(a) *Actor-oriented*, where actors (team members) in the dataset are incorporated in the analysis
(b) *Behaviour-oriented*, where the behaviours (essence-based or feature-based) are incorporated in the analysis

The two are not mutually exclusive, meaning that one analysis technique can combine both actor-oriented and behaviour-oriented units of analysis.

The dynamic output generated can be on:

(a) *Qualitative dynamic patterns*, referring to understanding the composition of each pattern. The term 'qualitative' here refers to the fact that the content and interpretation of each behaviour that makes up the pattern are of central focus, and the term dynamic refers to considering the order in which behaviours occur. It can be based on essence-based behaviours, for example examining whether an action-oriented behaviour is being followed by an information-oriented behaviour (Kolbe et al., 2014), or in relation to feature-based behaviours, for example exploring the development of inertia or recency patterns in actor sequences (Butts, 2008; Leenders et al., 2016).

Actor-oriented	Analysis techniques that focus on actors and can provide qualitative information between actor patterns	Analysis techniques that focus on actors and can provide quantitative information on actor patterns
Behaviour-oriented	Analysis techniques that focus on behaviours and can provide qualitative information on behaviour patterns	Analysis techniques that focus on behaviours and can provide quantitative information on actor patterns
	Structural qualitative dynamics	**Structural quantitative dynamics**

Fig. 9.1 DATS Framework for temporal analysis technique selection

(b) *Quantitative dynamic patterns*, referring to number of occurrences of each pattern, number of components, frequencies, distributions. Information on the timing of each quantitative output is necessary for a method to fall under that category (hence labelled 'dynamic').

Again, the two are not mutually exclusive and can both be generated from the same analysis technique.

In the following section, we list the main temporal techniques that enable the investigation of interaction dynamics and their temporal underpinnings. We use the DATS framework to categorise each technique based on the units analysed and the output generated (see Fig. 9.2). To highlight the use and added value of each technique, we provide some examples of research that can be done using each technique in the context of team learning.

9.4.2 Pool of Temporal Analysis Techniques

We apply the DATS framework on a representative, yet not exhaustive pool of five temporal analysis techniques, namely sequential analysis, relational event modelling, process mining, non-linear time series analysis, and T-pattern analysis. In Fig. 9.2, the techniques are classified based on their actor or behaviour-oriented units of analysis, as well as their qualitative or quantitative dynamic output. We also provide an overview of each technique, however getting into details on how each can be applied is beyond the scope of this chapter. For more information on the requirements of the presented techniques see David et al. (2021).

Actor-oriented	Relational Event Modelling T-pattern analysis	Non-linear, time-series analysis T-Pattern analysis
Behaviour-oriented	Sequential Analysis (LSA, FSM) Process Mining T-Pattern analysis	Non-linear, time-series analysis T-Pattern analysis
	Structural qualitative dynamics	**Structural quantitative dynamics**

Fig. 9.2 DATS Framework applied on pool of temporal analysis techniques

Sequential Analysis Analysis techniques that fall under this category are lag sequential analysis (LSA; Bakeman & Quera, 2011; Quera, 2018) and frequent sequence mining (FSM; (Zaki, 2001). LSA considers the order of behaviours in the data and identifies statistically significant direct or indirect transitions from one behaviour to another (A→B or A→C respectively), thus identifying pattern sequences. The main use of LSA is to identify sequential patterns of behaviours, thus providing structural *qualitative* information on *behavioural-oriented* patterns. It also calculates frequencies, probability and significance of transitions as well as information on the strength of association.

Different from LSA, FSM detects sequences of behaviours that occur more often than a minimum level in a dataset (e.g., 10%). Combinations of different sequences of behaviours are introduced by the coder and the analysis identifies sequences and sub-sequences that occur above a customized threshold in all defined sequences. FSM thus focuses on *behavioural* units and generates frequencies of varying sequences (Chen et al., 2017). Even though the order of the behaviours is considered, LSA and FSM do not provide information on the exact moment in time at which certain sequences are present, or how they evolve differently throughout an interactive meeting, thus neglecting structural quantitative aspects as we defined them. A research paper that can be used as tutorial for applying this technique in the context of team learning is by Kolbe et al. (2012).

Sequential analysis can be used to answer questions relating to the likelihood of event sequences to exist at different phases during a meeting, at different points throughout a team's life-cycle (comparison of different meetings), or at different teams in an organization exposed to similar experiences. A simplified example of how sequential analysis can be used in the context of team learning is to assess the process of *co-construction* during a learning event. During training, co-construction involves engaging in "repeated cycles of acknowledging, repeating, paraphrasing, enunciating, questioning, concretizing, and completing the shared knowledge, competencies, opinions or creative thoughts" (Decuyper et al., 2010). Researchers can use these behaviours to code a communication dataset, and explore the nature of co-construction with LSA by seeing how these behaviours form different sequences, and which sequences are more likely to occur at different moments in time.

Relational Event Modelling (REM) REM shares ideas with network systems theory (Herndon & Lewis, 2015; Klonek et al., 2019; Lehmann-Willenbrock & Allen, 2018), which views teams as a system consisting of nodes (actors) and links (relationships between actors). REM assumes that past relationships between actors shape future relationships and interactions; that is, every sequence of events between actors depends on the immediately preceding events (Butts, 2008; Butts & Marcum, 2017). Relational events consist of a sender, a receiver and a timestamp, and reveal interaction dynamics such as inertia, referring to the degree to which a member's past contact will be their future contact, or recency, referring to the degree to which a member's most recent contact will be their future contact (for more patterns see Marcum & Butts, 2015). REM gives actor-oriented information of *qualitative* nature (different ways in which relationships occur). The frequency and significance of sequences are also provided, but there is no output related to their exact temporal stamp. A paper that can be used as guide for applying REM in team research is by David and Schraagen (2018).

REM could be used to assess how actor relationships are affected by immediately preceding events and how these change for example after the incorporation of a new team member. Different patterns of actor relationships can be examined and modelled on the dataset, such as the aforementioned inertia (e.g. negative inertia would indicate searching behaviours on whom to share information with).

Even though we only focus on REM in this chapter, note that other techniques and stochastic actor oriented models (SAOM; Snijders, 2001) of the social network analysis family could also be used in interaction dynamics research (for review and examples see Borgatti & Foster, 2003; Lusher et al., 2010; Robins, 2015).

Process Mining Process mining includes data mining methods such as Petri Nets or Finite State Machines (Robero et al., 2010). It includes modelling the data in process models that consist of nodes (event classes) and edges (connections between nodes). The two key parameters examined within these models are the significance (how important nodes and edges are) and correlations (how closely related events following one another are) between data points. Process mining can be inductive, including model discovery (e.g. Heuristic Miner; Weijters et al., 2006) and thus enabling the investigation of the temporal structure of events and discovery of unknown sequences between them. The technique also offers conformance checking, where one can assess the extent to which a process model generated from one dataset differs from another dataset (Van der Aalst, 2012). Process mining is mostly behaviour-oriented as it assesses the structural composition between micro-behaviours without considering the involved actors. Its output reveals qualitative dynamics by offering the ability to assess the content of the patterns discovered. Software support for process mining is included in the ProM framework (Van Dongen et al., 2005) providing guidance and tool support for process analysis.

An example from research that can be used as a tutorial for applying process mining to team learning context is the research by Engelmann and Bannert (2021), who investigated cognitive and metacognitive events in self-regulated learning, and found that unfolding interaction sequences between team members, unlike popular belief, included weak connections to metacognitive behaviours.

Non-linear Time-Series Analysis This category includes a variety of methods such as Recurrence analysis, Phase or state space analysis, Hurst exponent, Lyapunov exponent and entropy (Guastello, 2017). They all deal primarily with interval data, although variations of each analysis also enable their application in discrete interaction sequences of ordinal spacing (e.g. Gorman et al., 2012, 2019). This type of analysis assesses whether data points in a time-series are recurrent or random. Depending on the data modelled, the unit of analysis can be both actor or behaviour-oriented, with output revealing structural quantitative dynamics.

While non-linear time-series analysis is performed through different techniques, a paper that discusses and guides its readers through the application of state space grids is that by Meinecke et al. (2019). In their paper, Meinecke and colleagues showcase step-by-step how to investigate the development of team problem-solving behaviours by comparing patterns of interaction between the first and final team meeting. Non-linear time-series analysis can also be used to model existing or known team learning processes, or spot new, emergent states within a team learning event.

T-Pattern Analysis TPA is a mixed method, aimed at detecting patterns in a dataset that would be undetectable by a naked eye (Magnusson, 2000, 2018). Any T-pattern is made up of a criterion and a target behaviour that constitute a T-pattern if these fall adjacent to one another within a time-dependent ‘critical interval’. Small T-patterns (*ab*) may also be part of larger patterns (*abcd*). TPA detects hierarchical pattern compositions and visualises how these unfold in different ways over time, thus providing both structural *qualitative and quantitative* information on *behavioural* patterns. It can also provide both types of structural information for actor-oriented units, as it is also able to yield information regarding the involvement of actors and actor switches. However, it is not dynamically sensitive in terms of actors, as no information is available on the ordering of actors switches or how they unfold differently over time. TPA also generates further structural quantitative output regarding Monte Carlo validation, number of pattern occurrences, and number of components. It is the only method that applies to all four aspects of our framework, enabling the analysis of both behavioural and actor-oriented units, while yielding qualitative and quantitative structural outputs for both. For T-pattern analysis breakdown and step-by-step application see Magnusson (2017, 2018).

T-pattern analysis offers very good opportunities for the exploration and modelling of emergent states. Different or more complex patterns that would otherwise be undetectable could be traced to specific moments in time, thus suggesting possible emergence of a shared mental model, or a recognition of another catalyst emergent state.

Note that any of the methods can provide structural quantitative dynamics if a performance episode is segmented into different phases, thus enabling the comparison of pattern development across the different phases. Phases can be determined either based on certain events that occur within the episode (e.g. unexpected surprise), or based on standard time intervals. For example, SNA can provide dynamic output in interaction by comparing different networks across multiple phases within a performance episode. Also, depending on the needs of the research, different analysis techniques can be used isolated or in combination. For example, a non-linear time-series analysis can be used in combination with sequential analysis, to capture both structural qualitative and quantitative aspects of the data.

9.5 A Temporal Approach to Team Learning: Closing Remarks

Teams make up the nucleus of every organisation, collaborating and coordinating their activities to meet organisational goals. In the ever-changing environments in which teams interact, either due to changes in team membership, task demands, or other organisational structures, teams are called to learn and adapt. Organisations are thus called to embrace a culture of lifelong professional learning, integrating, and promoting formal, non-formal and informal learning experiences in their employees (Carr et al., 2018). Research has already shown that the processes making up team learning are continuous in nature. It is thus important for learning research to move past the static representation of team learning and start capturing its dynamic nature, which stems from the continuous team member interaction.

Throughout our chapter, we aimed to show our readers how interaction dynamics that reveal temporal changes in teamwork are a proxy for investigating and understanding team learning, its underlying processes, such as co-construction and constructive conflict, and its emergent states, such as shared mental models or psychological safety (Decuyper et al., 2010). Through examples, we tried to illustrate the added value and opportunities in researching on a detailed level 'what happens' both in small-scale and large-scale, complex, interconnected systems. As a closing remark to this chapter, we would like to highlight the three key takeaways for ensuring that team learning is studied as a temporal phenomenon:

9.5.1 Key Takeaways

(i) *Define the phenomenon of interest and determine any relevant interaction dynamics*

Processes and emergent states related to team learning can be studied through the investigation of interaction dynamics, which we define in this chapter as sequential patterns created between micro-behaviours that change throughout interaction. We break micro-behaviours down to two different categories, essence- and feature-based, thus easing the process of identifying relevant micro-behaviours that ought to be investigated in order to capture the temporality of the chosen phenomenon, process, or emergent state.

(ii) *Choose a data collection and coding method that can capture the chosen interaction dynamics*

We outline certain possibilities for data gathering and coding of micro-behaviours that can help detect interaction dynamics in the data. Data collection techniques like the ones mentioned in this chapter enable the investigation of team learning in the field rather than in controlled contexts, thus capturing interaction in actual organisational contexts. We see that there is an evolution from the time and labour-intensive coding, moving towards automatic ways of attaining data. But challenges still exist. State-of-the-art technologies such as sociometric badges are yet to be validated, while developers are not opening up in an understandable way regarding the transformation that occurs in automatically coded or analysed data. We call for the incorporation of more transparent technological methodologies, so as to ensure the adoption and incorporation of such techniques in team learning research.

(iii) *Choose an analysis technique that can help you capture temporality*

Lehmann-Willenbrock and Allen (2018), pointed out how researchers should start embracing existing methodologies and analysis techniques that capture temporality in the investigation of team learning. We stand by this claim and offer a framework for choosing such techniques based on their unit of analysis and the output they generate, alongside a pool of existing techniques. We also provide some examples and a study 'tutorial' that can be used for the application of these techniques and the incorporation of interaction dynamics in team learning research. Some of the studies that we refer to already deal with aspects that can also be investigated within the context of team learning, such as self-regulated learning by Engelmann and Bannert (2021), while others provide a detailed description of the analysis technique they used that can be applied by researchers in the field of team learning in the future.

We hope that our framework for data analysis technique selection will assist and foster the investigation of team learning, its processes, and its emergent states as these unfold over time, in actual organisational contexts where teams are faced with, and learn through true organisational challenges.

References

Bakeman, R., & Quera, V. (2011). Sequential analysis and observational methods for the behavioral sciences. In *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9781139017343

Baker, D. P., Day, R., & Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health Services Research, 41*(4p2), 1576–1598. https://doi.org/10.1111/J.1475-6773.2006.00566.X

Bokhove, C., & Downey, C. (2018). Automated generation of 'good enough' transcripts as a first step to transcription of audio-recorded data. *Methodological Innovations, 11*(2), 205979911879074. https://doi.org/10.1177/2059799118790743

Bonito, J. A., & Keyton, J. (2018). Introduction to machine learning: Teaching computers to code group interaction data. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis* (pp. 387–404). Cambridge University Press. https://doi.org/10.1017/9781316286302.020

Borgatti, S., & Foster, P. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management, 29*, 991–1013. https://doi.org/10.1016/S0149-2063_03_00087-4

Brauner, E., Boos, M., & Kolbe, M. (2018). Coding schemes for interaction research. In *The Cambridge handbook of group interaction analysis* (pp. 419–638). Cambridge University Press.

Burtscher, M. J., Kolbe, M., Wacker, J., & Manser, T. (2011). Interactions of team mental models and monitoring behaviors predict team performance in simulated anesthesia inductions. *Journal of Experimental Psychology: Applied, 17*(3), 257–269. https://doi.org/10.1037/A0025148

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology, 38*(1), 155–200. https://doi.org/10.1111/j.1467-9531.2008.00203.x

Butts, C. T., & Marcum, C. S. (2017). A relational event approach to modeling behavioral dynamics. In A. Pilny & M. S. Poole (Eds.), *Group processes: Data-driven computational approaches* (pp. 51–92). Springer International Publishing.

Carr, A., Balasubramanian, K., Atieno, R., Onyango, J., Carr, A., Balasubramanian, K., Atieno, R., & Onyango, J. (2018). Lifelong learning to empowerment: Beyond formal education. *Distance Education, 7919*, 1–18. https://doi.org/10.1080/01587919.2017.1419819

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments, 25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Ciolek, T. M., & Kendon, A. (1980). Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry, 50*, 237–271.

David, L. Z., & Schraagen, J. M. (2018). Analysing communication dynamics at the transaction level: The case of Air France Flight 447. *Cognition, Technology and Work, 20*(4), 637–649. https://doi.org/10.1007/s10111-018-0506-y

David, L. Z., Schraagen, J. M., & Endedijk, M. (2021). Toward the incorporation of temporal interaction analysis techniques in modeling and understanding sociotechnical systems. *Human Factors and Ergonomics in Manufacturing, 32*(1), 35–49. https://doi.org/10.1002/hfm.20930

De Ruiter, J. P., & Loth, S. (2016). *Understanding social signals: How do we recognize the intentions of others?* Frontiers Media SA 2016. http://lib.ugent.be/catalog/ebk01:3710000001092144

Decuyper, S., Dochy, F., & Van den Bossche, P. (2010). Grasping the dynamic complexity of team learning: An integrative model for effective team learning in organisations. *Educational Research Review, 5*, 111–133. https://doi.org/10.1016/j.edurev.2010.02.002

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Elsevier.

Dong, W., Olguin-Olguin, D., Waber, B., Kim, T., & Pentland, A. (2012). Mapping organizational dynamics with body sensor networks. In *Proceedings – BSN 2012: 9th international workshop on wearable and implantable body sensor networks* (pp. 130–135). https://doi.org/10.1109/BSN.2012.16

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Endedijk, M., Hoogeboom, A. M. G. M., Groenier, M., De Laat, S., & Van Sas, J. (2018). Using sensor technology to capture the structure and content of team interactions in medical emergency teams during stressful moments. *Frontline Learning Research, 6*(3), 123–147. https://doi.org/10.14786/flr.v6i3.353

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction, 72*, 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning, 3*(4), 447–470. https://doi.org/10.1007/s11412-008-9052-6

Fay, D., Shipton, H., West, M. A., & Patterson, M. (2015). Teamwork and organizational innovation: The moderating role of the HRM context. *Creativity and Innovation Management, 24*(2), 261–277. https://doi.org/10.1111/caim.12100

Gambi do Nascimento, L., Lizarelli, F. L., & Boer, H. (2020). The effect of teamwork on innovation performance: Does organizational culture have some influence? In *Proceedings of the 21st CINet conference on practicing continuous innovation in digital ecosystems* (Vol. 21, pp. 267–278).

Gerken, M., Beausaert, S., & Segers, M. (2016). Working on professional development of faculty staff in higher education: Investigating the relationship between social informal learning activities and employability. *Human Resource Development International, 19*(2), 135–151. https://doi.org/10.1080/13678868.2015.1116241

Gontar, P., & Mulligan, J. B. (2016, September). Cross recurrence analysis as a measure of pilots' visual behaviour. In *Proceedings of the 32nd conference of the European association for aviation psychology*. Groningen, NL.

Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Ddetecting perturbations to team communication. *Ergonomics, 55*(8), 825–839. https://doi.org/10.1080/00140139.2012.679317

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2019). Measuring real-time team cognition during team training. *Human Factors, 62*(5), 825–860. https://doi.org/10.1177/0018720819852791

Grant, A. M., & Ashford, S. J. (2008). The dynamics of proactivity at work. In *Research in organizational behavior* (Vol. 28, pp. 3–34). Elsevier. https://doi.org/10.1016/j.riob.2008.04.002

Guastello, S. J. (2017). Nonlinear dynamical systems for theory and research in ergonomics. *Ergonomics, 60*(2), 167–193. https://doi.org/10.1080/00140139.2016.1162851

Heaphy, E. D., Byron, K., Ballinger, G. A., Gittell, J. H., Leana, C., & Sluss, D. M. (2018). Introduction to special topic forum: The changing nature of work relationships. *Academy of Management Review, 43*(4), 558–569. https://doi.org/10.5465/amr.2018.0222

Herndon, B., & Lewis, K. (2015). Applying sequence methods to the study of team temporal dynamics. *Organizational Psychology Review, 5*(4), 318–332. https://doi.org/10.1177/2041386614538276

Huang, K., Bryant, T., & Schneider, B. (2019). Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data. In *EDM 2019 – Proceedings of the 12th international conference on educational data mining* (pp. 318–323). Edm.

Kauffeld, S., Lehmann-Willenbrock, N., & Meinecke, A. L. (2018). The advanced interaction analysis for teams (act4teams) coding scheme. In *The Cambridge handbook of group interaction analysis* (pp. 422–431). Cambridge University Press. https://doi.org/10.1017/9781316286302.022

Kayhan, V. O., Chen, Z., French, K. A., Allen, T. D., Salomon, K., & Watkins, A. (2018). How honest are the signals? A protocol for validating wearable sensors. *Behavior Research Methods, 50*(1), 57–83. https://doi.org/10.3758/s13428-017-1005-4

Kim, T., Chang, A., Holland, L., & Pentland, A. S. (2008). Meeting mediator: Enhancing group collaboration using sociometric feedback. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (pp. 457–466). https://doi.org/10.1145/1460563.1460636

Kim, T., McFee, E., Olguin, D. O., Waber, B., & Pentland, A. "Sandy". (2012). Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior, 33*(3), 412–427. https://doi.org/10.1002/job.1776

Klonek, F., Gerpott, F. H. G., Lehmann-Willenbrock, N., & Parker, S. K. (2019). Time to go wild: How to conceptualize and measure process dynamics in real teams with high-resolution. *Organizational Psychology Review, 9*(4), 245–275. https://doi.org/10.1177/2041386619886674

Klonek, F., Meinecke, A. L., Hay, G., & Parker, S. K. (2020). Capturing team dynamics in the wild: The communication analysis tool. *Small Group Research, 51*(3), 303–341. https://doi.org/10.1177/1046496420904126

Kolbe, M., Burtscher, M. J., Wacker, J., Grande, B., Nohynkova, R., Manser, T., Spahn, D. R., & Grote, G. (2012). Speaking up is related to better team performance in simulated anesthesia inductions. *Anesthesia & Analgesia, 115*(5), 1099–1108. https://doi.org/10.1213/ANE.0b013e318269cd32

Kolbe, M., Burtscher, M. J., & Manser, T. (2013). Co-ACT-a framework for observing coordination behaviour in acute care teams. *Research and Reporting Methodology*. https://doi.org/10.1136/bmjqs

Kolbe, M., Grote, G., Waller, M. J., Wacker, J., Grande, B., Burtscher, M. J., & Spahn, D. R. (2014). Monitoring and talking to the room: Autochthonous coordination patterns in team interaction and performance. *Journal of Applied Psychology, 99*(6), 1254–1267. https://doi.org/10.1037/a0037877

Kozlowski, S. W. J. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review, 5*(4), 270–299. https://doi.org/10.1177/2041386614533586

Kugler, P. N., & Turvey, M. T. (1987). Information, natural law, and the self-assembly of rhythmic movement. In *Information, natural law, and the self-assembly of rhythmic movement*. Lawrence Erlbaum Associates, Inc.

Lambert, B. L. (2001). Automatic content analysis of pharmacist-patient interactions using the Theme machine document clustering system. In M. West (Ed.), *Progress in communication sciences, volume 17: Applications of computer content analysis* (pp. 103–122). Ablex.

Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Lampi, E. (2021). What do we do when we analyse the temporal aspects of computer-supported collaborative learning? A systematic literature review. *Educational Research Review, 33*, 100387. https://doi.org/10.1016/j.edurev.2021.100387

Leenders, R. T. A. J., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review, 6*(1), 92–115. https://doi.org/10.1177/2041386615578312

Lehmann-Willenbrock, N., & Allen, J. A. (2018). Modeling temporal interaction dynamics in organizational settings. *Journal of Business and Psychology, 33*(3), 325–344. https://doi.org/10.1007/s10869-017-9506-9

Lusher, D., Robins, G., & Kremer, P. (2010). The application of social network analysis to team sports. *Measurement in Physical Education and Exercise Science, 14*(4), 211–224. https://doi.org/10.1080/1091367X.2010.495559

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, and Computers, 32*(1), 93–110. https://doi.org/10.3758/BF03200792

Magnusson, M. S. (2017). *THEME 6 user's manual*. http://patternvision.com/wp-content/uploads/2017/06/Theme-Manual-7-June-2017.pdf

Magnusson, M. S. (2018). Temporal patterns in interactions. In *The Cambridge handbook of group interaction analysis* (pp. 323–353). Cambridge University Press. https://doi.org/10.1017/9781316286302.017

Mangold. (2020). *INTERACT user guide*. Mangold International GmbH (Ed.). www.mangold-international.com

Marcum, C. S., & Butts, C. T. (2015). Constructing and modifying sequence statistics for relevent using informR in R. *Journal of Statistical Software, 64*(5), 1–36. https://doi.org/10.18637/jss.v064.i05

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review, 26*(3), 356. https://doi.org/10.2307/259182

Meinecke, A. L., Hemshorn de Sanchez, C. S., Lehmann-Willenbrock, N., & Buengeler, C. (2019). Using state space grids for modeling temporal team dynamics. *Frontiers in Psychology, 10*(863), 863. https://doi.org/10.3389/fpsyg.2019.00863

Noldus, L. P. J. J., Trienes, R. J. H., Hendriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, and Computers, 32*(1), 197–206. https://doi.org/10.3758/BF03200802

Olguín Olguín, D., Waber, B. N., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 39*(1), 43–55. https://doi.org/10.1109/TSMCB.2008.2006638

Onnela, J. P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports, 4*(1), 1–9. https://doi.org/10.1038/srep05604

Quera, V. (2018). Analysis of interaction sequences. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis* (pp. 295–322). Cambridge University Press. https://doi.org/10.1017/9781316286302.016

Raes, E., Boon, A., Kyndt, E., & Dochy, F. (2015). Measuring team learning behaviours through observing verbal team interaction. *Journal of Workplace Learning, 27*(7), 476–500. https://doi.org/10.1108/JWL-01-2015-0006

Ricca, B. P., Bowers, N., & Jordan, M. E. (2019). Seeking emergence through temporal analysis of collaborative-group discourse: A complex-systems approach. *The Journal of Experimental Education, 88*(3), 431–447. https://doi.org/10.1080/00220973.2019.1628691

Rico, R., Sánchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *The Academy of Management Review, 33*(1), 163–184. https://doi.org/10.5465/AMR.2008.27751276

Robero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *Handbook of educational data mining*. Chapman & Hall/CRC.

Robins, G. (2015). *Doing social network research: Network-based research design for social scientists*. SAGE Publications Ltd. https://doi.org/10.1007/s10588-016-9236-y

Roe, R. A. (2008). Time in applied psychology: The study of "what happens" rather than "what is". *European Psychologist, 13*(1), 37–52. https://doi.org/10.1027/1016-9040.13.1.37

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation Space and Environmental Medicine, 78*(5 Suppl), B77–B85.

Schmid, M. M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science, 24*(2), 154–160. https://doi.org/10.1177/0963721414560811

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31*(1), 361–395. https://doi.org/10.1111/0081-1750.00099

Stevens, R. (2012). Charting neurodynamic eddies in the temporal flows of teamwork. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*, 208–212. https://doi.org/10.1177/1071181312561020

Stivers, T. (2015). Coding social interaction: A heretical approach in conversation analysis? *Research on Language and Social Interaction, 48*(1), 1–19. https://doi.org/10.1080/08351813.2015.993837

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154. https://doi.org/10.1016/j.edurev.2007.12.001

Uitdewilligen, S., & Waller, M. J. (2018). Information sharing and decision-making in multidisciplinary crisis management teams. *Journal of Organizational Behavior, 39*(6), 731–748. https://doi.org/10.1002/JOB.2301

Van de Wiel, M. W. J., Van den Bossche, P., Janssen, S., & Jossberger, H. (2011). Exploring deliberate practice in medicine: How do physicians learn in the workplace? *Advances in Health Sciences Education, 16*(1), 81–95. https://doi.org/10.1007/s10459-010-9246-3

Van den Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: Building shared mental models. *Instructional Science, 39*(3), 283–301. https://doi.org/10.1007/s11251-010-9128-3

Van den Bossche, P., Gabelica, C., & Koeslag-Kreunen, M. G. M. (2022). Team learning. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning – Insights from a growing field*. Springer.

Van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems, 3*(2), 1–17. https://doi.org/10.1145/2229156.2229157

Van der Rijt, J., Van den Bossche, P., van de Wiel, M. W. J., de Maeyer, S., Gijselaers, W. H., & Segers, M. S. R. (2013). Asking for help: A relational perspective on help seeking in the workplace. *Vocations and Learning, 6*(2), 259–279. https://doi.org/10.1007/s12186-012-9095-8

Van Dongen, B. F., De Medeiros, A. K. A., Verbeek, H. M. W., Weijters, A. J. M. M., & Van Der Aalst, W. M. P. (2005). The ProM framework: A new era in process mining tool support. *Lecture Notes in Computer Science, 3536*(i), 444–454. https://doi.org/10.1007/11494744_25

Waller, M. J., & Kaplan, S. A. (2018). Systematic behavioral observation for emergent team phenomena: Key considerations for quantitative video-based approaches. *Organizational Research Methods, 21*(2), 500–515. https://doi.org/10.1177/1094428116647785

Weijters, A. J. M. M., Aalst, van der, W. M. P., & Alves De Medeiros, A. K. (2006). *Process mining with the HeuristicsMiner algorithm.* (BETA publicatie : working papers; Vol. 166). Technische Universiteit Eindhoven.

Wiese, C. W., & Burke, C. S. (2019). Understanding team learning dynamics over time. *Frontiers in Psychology, 10*, 1417. https://doi.org/10.3389/fpsyg.2019.01417

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive Science, 42*(1), 129–167. https://doi.org/10.1111/cogs.12482

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning, 42*(1–2), 31–60. https://doi.org/10.1023/A:1007652502315

Lida Z. David (MSc) is a PhD Candidate at the Department of Data analysis, Learning and Technology at the University of Twente. Her research concerns team resilience, and how dynamic interaction patterns related to adaptive mechanisms emerge and change as teams work in challenging environments. Lida is editor in chief of the Resilience Engineering Association Newsletter.

Maaike D. Endedijk (PhD) is Professor of Professional Learning and Technology at the University of Twente. Her research interests are workplace and professional learning, self-directed learning, and team learning. Recently started projects focus on learning in relation to the energy transition, health care transformation and the digital transformation.

Piet Van den Bossche (PhD) is Professor of Learning in Organizations at the University of Antwerp (Faculty of Social Sciences) and Professor of Team Learning at Maastricht University (School of Business & Economics). He is co-director of the Antwerp Social Lab. His research activities are centered on learning, cognition and performance in teams and collaborative work.

Part II
Methods for Data Analysis

Chapter 10
Bayesian Statistics in the Research Field of Professional Learning and Development

Petri Nokelainen, Tahani Z. Aldahdouh, and Alaa A. Aldahdouh

Abstract This chapter will discuss issues related to analysing empirical data in the research field of professional learning and development using Bayesian statistics. It will start by briefly explaining why the frequentist (so-called classical) approach to analysing empirical data in professional learning and development research is so popular. Also, some conditions when this approach might not be the best solution is discussed and contrasted to the Bayesian approach. The essentials of Bayesian statistics are described and some practical examples of its application are provided. A major part of this chapter is devoted to an example of applying Bayesian statistics in the context of multilevel path analysis. This chapter concludes with a discussion of using Bayesian methods in professional learning and development research and its potential future views.

Keywords Bayesian statistics · Multilevel path analysis · Professional learning and development

10.1 Introduction

Frequentist statistical research methods are preferred for several practical reasons, from a researcher's point of view. First, many journal editors like them, as the results usually contain *p*-values and confidence intervals, and they can be assured that the majority of reviewers are familiar with a large variety of frequentist techniques (e.g., linear regression based on ordinary least squares). Thus, a researcher does not need to "waste" the limited number of words (e.g., typically 6000–8000) in the manuscript for *the methods section* to explain how the results should be interpreted.

P. Nokelainen (✉) · T. Z. Aldahdouh
Faculty of Education and Culture, Tampere University, Tampere, Finland
e-mail: petri.nokelainen@tuni.fi; tahani.aldahdouh@tuni.fi

A. A. Aldahdouh
Independent Researcher, Tampere, Finland

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_10

Second, the application of frequentist methods is quite easy to comprehend, as they dominate the contents of university-level courses, online tutorials, and textbooks. There is a lot of support and examples available, and their application and results are easy to discuss with colleagues at conferences. Third, frequentist models are usually quite fast to calculate, as they are based on the assumption of fixed parameters that need to be fitted into any existing “universal” distribution. Due to the central limit theorem, a widely applied normal distribution does not require a normal distribution of the actual sample values but instead assumes normally distributed sampling distributions (which are obtained if the sample size is 30 or more).

However, certain challenges are associated with frequentist methods that make them a less attractive option for some researchers. In the research field of professional learning and development (PLD), typical quantitative data are based on self-assessments, cross-sectional design, multiple scale survey(s), and regression, path, or latent variable analysis. Although the sample sizes seem to be typically quite impressive (e.g., more than 300 participants), taking model complexity into account may make them look somewhat modest. A study by Lecat et al. (2018) serves as a good example of a typical multivariate study in this field. The authors conducted path analysis on an opportunity sample of 301 teachers to investigate the relationship between (in)formal learning activities, employability (through competences, e.g., occupational expertise), and innovative working behaviour (e.g., ideas generation). As the path model included 15 variables, there were 120 available observations ((15 (15 + 1))/2) of which 71 were specified to be estimated in the final path model. A sample could be considered small when the ratio between the number of participants and the number of unknown parameters in the model is around 4–5, and very small if the ratio is close to 3 or less (Lee & Song, 2004). In Lecat et al.’s study, we find the ratio to be approximately 4.2, and thus the study could be labelled as “a small sample study”.

The last two decades have witnessed a growing interest in employing *Bayesian statistics* in multiple domains (van de Schoot et al., 2014). Many scholars have turned into Bayesian statistics over the years, as some complex models cannot be estimated using frequentist statistics (e.g., to improve convergence issues, aid in model identification, and produce more accurate parameter estimates). Further, the Bayesian approach allows the incorporation of existing knowledge into the estimation process via informative prior distributions, and it views the population parameters as random (instead of fixed), leading to a more intuitive interpretation of results and allowing the modeling of uncertainty as a part of the statistical model. Bayesian methods also welcome a mixture of measurement levels (e.g., continuous and discrete). Lastly, as they are not based on the large sample assumption (i.e., the central limit theorem), they may produce reasonable results, even with small to moderate sample sizes (when strong and defensible prior knowledge is available) (e.g., van de Schoot et al., 2017). Using a series of simulation studies with confirmatory factor analysis (CFA) and structural equation models (SEM), Lee and Song (2004) showed that the Bayesian results were robust even with small samples (ratios of 4 and 5) compared to the frequentist (maximum likelihood) approach.

A pertinent question that must be addressed is, what are the most essential features of Bayesian statistics for PLD researchers? *First*, 'updating beliefs' regarding the research phenomena under investigation via a priori information resonates well with the dynamic nature of the real (working) life (Gill & Witko, 2013). This is supported by Bayesian methods, as they are built on the assumption that we update our views (or beliefs) when we become more educated about the research topic or when we receive more evidence (data). However, we are not tied to using only informative priors (e.g., results of previous studies); instead, we can choose to use non-informative (e.g., various programs "default") priors and, thus, 'know very little a priori' except what the data at hand tells us. It is worth noting that the influence of any kind of prior information (informative or non-informative) on the resulting posterior distribution is greater with small sample sizes (the ratio of n of participants and n of estimated parameters in the model is ≤ 5) than with larger sample sizes (Kruschke & Liddell, 2018). Smid and Winter (2020) demonstrated this by estimating an SEM model with 13 parameters and various sample sizes ($n = 26$, $n = 52$, $n = 325$). They found that the use of "non-informative" priors produced different results with very small (26/13 = 2) and small (52/13 = 4) sample sizes compared to using "partial informative" or "full informative" prior information. The results based on these three different prior settings were quite identical when the sample size was large (325/13 = 25).

Second, the result of the Bayesian analysis, the posterior distribution, answers directly to the research question set by the researcher: How certain are we, after declaring our prior beliefs (non-informative/informative) and observing the data, about the population values (of interest)? (Gill & Witko, 2013). One of the first publicly available online tutorials (released in 2002) on data structure learning, *B-Course* (http://b-course.hiit.fi/obc, see Myllymäki et al., 2002), demonstrates this in a simple and user-friendly way (however, with non-informative priors). A more complete picture of the topic is available via the *bnstruct* package (Franzin et al., 2017) that operates in the *R* environment (R Core Team, 2020).

Third, Bayesian methods release some of the assumptions compared to frequentist methods, as they can operate robustly with small samples and do not require normality when variances are estimated (Hox, 2010). Currently, developments within the analysis of the monotonic effects of ordinal predictors are interesting to empirical researchers who use survey data. This method, implemented in *R brms* (Bürkner, 2017) allows ordinal level (e.g., Likert scale items) predictor variables to have a monotonically increasing or decreasing relationship with the response (dependent) variable instead of falsely treating them either as continuous or as unordered categorical predictors (Bürkner & Charpentier, 2020).

In summary, Bayesian inference is different from frequentist (classical) inference, mainly because it (a) treats the model parameters as random variables (rather than fixed) and (b) allows for prior information to be formally taken into account in the analysis (Puza, 2015). The first point practically suggests that we are allowed to make conclusions based on the actual data that we have collected (e.g., to report a 95% probability that a correlation in the population will be between .31 and .57), and the second point allows us to combine the findings of an existing relevant research body with our own empirical data.

10.2 Introduction to Bayesian Statistics

Bayesian statistics has gained popularity over other (most often frequentist) statistical methods not only for the aforementioned reasons but also for the fact that it takes prior knowledge into account when conducting the analysis, while other statistical methods fail to integrate the findings from previous studies into the current data analysis. According to König and van de Schoot (2018), contributing to building robust accumulative knowledge entails the incorporation of results from comparable prior research. To illustrate this, we consider a researcher who is conducting several studies on a topic, such as the goal orientation of university teachers. For each study, the analysis is done via conventional analyses on the data in hand, while ignoring the results yielded from previous samples or even from the earlier studies done by other researchers. Doing so deprives the researcher of using existing previous knowledge and improving the estimation of the magnitude and uncertainty of the study parameters. Combining the current data with the findings from previous comparable studies enhances the accuracy of parameter estimates and thus leads to more precise results and richer interpretations. Bayesian statistics is like a meta-analysis in the sense that it is able to use previous knowledge in data analysis. However, what distinguishes the Bayesian approach from the meta-analysis is that it repeatedly revises the current knowledge as new data become available, while in the meta-analysis, there should be sufficient studies published to be able to conduct the meta-analysis (König & van de Schoot, 2018).

Principles and advanced details of Bayesian statistics have been presented in numerous excellent textbooks over the years (e.g., Berger, 1985; Bernardo & Smith, 2000; Congdon, 2001). One seminal work that focuses on Bayesian statistics in contexts relevant to PLD researchers was that of Jeff Gill (2007). An easy-to-read introduction to the topic is a book by Dennis Lindley (1971) and Francisco Samaniego (2010), who made an exhaustive comparison of frequentist and Bayesian estimation methods. The recommended current theoretical book was written by Andrew Gelman et al. (2013), and for applied researchers who use *R,* we recommend Kruschke's book (2015).

What follows next is a basic-level introduction to the Bayesian process of data analysis that has three main characteristics: (a) assigning prior distributions to unknown parameters, (b) using Bayes' rule to obtain a posterior distribution for unknown parameters and missing data conditioned on observable data, and (c) describing inferences in probabilistic terms.

10.2.1 Bayes' Rule

Bayesian inference uses conditional probabilities to represent uncertainty (Congdon, 2001). Conditional probability refers to the probability that one event will occur, given that another has occurred (Hastings, 1997). Therefore, researchers' interest lies in $P(H|D)$—the probability of unknown things (H = hypothesis) given the evidence (D = data). The initial uncertainty about H is also represented as a conditional probability $P(H|I)$, where I stands for information. For example, we could have some initial belief that some answers are more likely than others.

Bayesian statistics derive their name from the initial works of a British reverend, Thomas Bayes (1702–1761), who was interested in studying what the data (evidence) reveal about the probability of an event of interest (e.g., the probability of an existence of a dragon if a villager claims to have seen one on the village road). The essence of Bayesian inference is in the Bayes' rule (or theorem), which tells us how to update our initial beliefs (probabilities) $P(H)$ if we see some evidence (data) D, to find out what we truly believe in (i.e., posterior probability) $P(H|D)$:

$$P(H|D) = \frac{P(D \vee H)P(H)}{P(D \vee H)P(H) + P(D \vee H)P(H)} \tag{10.1}$$

Consequently, Bayesian inference comprises the following three principal steps:

1. Obtain the initial probabilities $P(H)$ for unknown things. These probabilities are called *prior (distribution) knowledge* (prior information);
2. Calculate the probabilities of the data D given different values for the unknown things, that is, $P(D|H)$. This function of the unknowns is called the *likelihood function*;
3. Calculate the probability distribution of interest, $P(H|D)$ using Bayes' rule given above. This so-called *posterior (distribution)* will then express what is known about our hypothesis (H) after observing the data (D) and using the information from prior knowledge $P(H)$.

Prior distribution $P(H)$ represents the previous knowledge (or lack thereof) that we have about the parameter (hypothesis) to be estimated. This parameter is an unknown variable for which the researcher aims to find an estimation (e.g., the average level of employees' job satisfaction). It could be that we know nothing about workers' job satisfaction in a certain field, or that we have previous assumptions (or beliefs) about how much the workers are satisfied in their job. The prior distribution reflects our prior knowledge, assumptions, or beliefs. Prior distributions can be seen as informative (informed) or non-informative (noninformed) (Gelman et al., 2013). *Informative priors* are based, for example, on expert interviews, the results of previous studies, or meta-analyses. Their role is to have an effect on the outcome of the analysis (posterior probability) in cooperation with the effect of the actual evidence (data). However, it should be noted that the effect of prior distributions on posterior probability diminishes when the sample size of the data increases (Gelman et al., 2013). *Non-informative priors* are set in such a way that their effect is as small as possible, and the actual observed data play a major role in the analysis. A uniform prior distribution (i.e., the same probability is attached to all outcome values) may serve as an example of a non-informative prior (of course, with reservations, see Myllymäki et al., 2002; Gill & Witko, 2013).

Likelihood function $P(D|H)$ is the probability of data given the hypothesis (parameter); thus, it represents the evidence coming from the data itself or the observed data (e.g., the mean of the sample in hand). It clarifies how well the data distinguish the occurrence of the parameter from its absence. It is worth mentioning that frequentists (see, e.g., Mayo & Spanos, 2011) and likelihoodists (see, e.g., Blume, 2011) also use maximum likelihood (ML) estimation to find this value (the probability of data given the parameter).

Posterior distribution $P(H|D)$ is obtained by multiplying the prior distribution and the likelihood function: $P(H|D) = P(D|H)\ P(H)/P(D)$. Unlike the frequentist method of using likelihood analysis (producing a single point estimate and its variance), this approach of using data through the likelihood function to update prior distribution into posterior knowledge produces a more informative distributional summary (mode, median, mean, standard deviation) (Gill & Witko, 2013).

As learning from data is the basic idea in Bayesian statistics (van de Schoot et al., 2014), Bayes' rule can be used sequentially. If we first receive some data D, calculate the posterior $P(H|D)$, and at some later point in time receive more data D', the calculated posterior can be used in the role of prior to calculate a new posterior $P(H|D, D')$ and so on. The posterior $P(H|D)$ expresses all the necessary information to perform predictions. The more data we get, the more certain we will become of the unknowns, until all but one value combination for the unknowns has probabilities so close to zero that they can be neglected.

10.2.1.1 An Example of Applying Bayes' Rule

A business organisation B wants to employ workers who have fluent oral and writing skills in a specific language X. Publicly available demographic statistics indicate that the criterion is valid for 1% of the persons living in the population. The organisation's psychometrician has been given a task to develop a questionnaire that would pre-screen the most suitable applicants for a job interview. The psychometrician estimates that the sensitivity of the questionnaire (true positive rate: it will correctly identify applicants who really possess the required skills) is 95%. She also states that the questionnaire's specificity is 10% (false positive rate: it will falsely identify applicants who do not possess the required skills as potential candidates). Based on this information, we are interested in determining whether an applicant gets enough points to participate in the interview. What is the probability that he/she will be hired for the job (after an interview)?

A priori probability $P(H)$ is described as the number of people in the target population who are able to meet the requirements of the job (1 out of 100 = .01). The counter assumption of the a priori is $P(\sim\mathrm{H})$, which equals $1\text{-}P(H) = .99$. The psychometrician's belief about the instrument's sensitivity is the conditional probability, $P(D|H) = .95$, where D refers to the fact that the instrument reports that a person meets the specific language X writing skill requirement. The instrument's tendency to indicate non-valid applicants as potential candidates for the job, $P(D| \sim \mathrm{H})$, is stated as .10.

Next, we input the values to the Bayes' rule to get the posterior probability $P(H|D)$:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D|H)P(H) + P(D|H)P(H)} = \frac{.95 \times .01}{.95 \times .01 + .10 \times .99} = .088$$

The probability that a person who passes the psychometrician's test will actually get the job is .088 (approximately 9%). Why is the number so low? There are two plausible reasons: First, although the screening instrument is valid, it is not perfect (sensitivity: 95% true positive rate; specificity: 10% false positive rate). Second, there are not that many valid persons in the target population (only 1%). What if we make the screening instrument better, increasing its sensitivity (true positive rate) to 99% and reducing its specificity (false positive rate) to 5%? The posterior probability will then increase to .167 (16.7%). Further, if we increase the prior probability (the number of people in the population who are able to meet the requirements of the job) to 5% (instead of earlier 1%), we will get a much higher posterior probability of .667 (66.7%).

10.2.2 Comparison of Frequentist and Bayesian Correlation Analysis

To illustrate the difference between frequentist and Bayesian analysis, we calculated a correlation between two summative scale variables from the data of a previous study by Pylväs and Nokelainen (2020) using the *jamovi* (The jamovi project, 2021) *jsq* package (The JASP Team et al., 2022) in the *R* environment (R Core Team, 2020). The survey part of Pylväs and Nokelainen's (2020) multi-method study investigated academics' ($n = 59$) perceptions of their intercultural competence after an international exchange period and the influence of academic international mobility on their professional development. In this analysis, we used two composite (average) variables from the Cultural Intelligence Scale (CQS, see van Dyne et al., 2009): cognitive cultural intelligence (CQ), which consists of 6 items, and behavioural cultural intelligence, which consists of 5 items. A sample item for cognitive CQ is "I know the cultural values and religious beliefs of other cultures", and a sample item for behavioural CQ is "I vary the rate of my speaking when a cross-cultural situation requires it". The response scale for all items ranged from 1 = totally disagree to 5 = totally agree. A random sample ($n = 25$) was drawn from the data to investigate the effect of prior distributions on the results.

Usually, a bivariate correlation (linear relation between two variables) is calculated to estimate a parameter (rho: ρ) and its confidence interval, and/or to test whether a two-sided null hypothesis is more evident than an alternative hypothesis

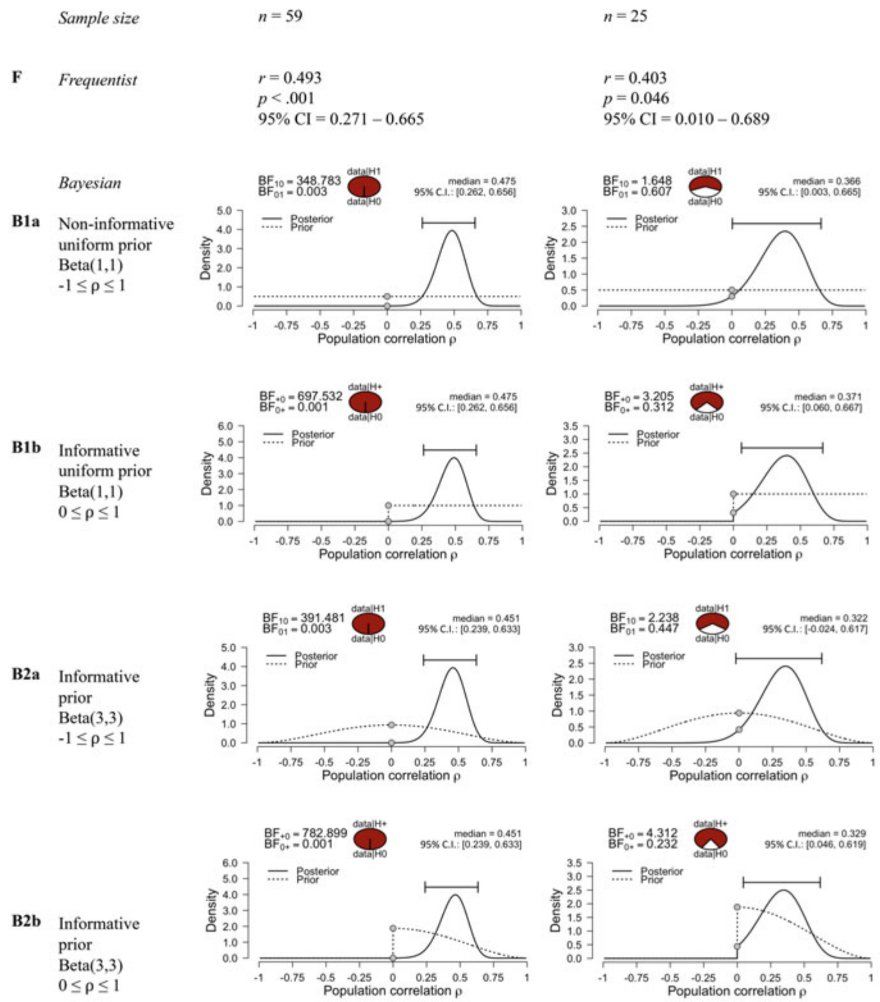

Fig. 10.1 Pearson correlation between two variables and related frequentist and Bayesian significance reports with two different sample sizes

(H_0: $\rho = 0$; H_1: $\rho \neq 0$). The "F" row in Fig. 10.1 shows the results of the frequentist correlation analysis for the two sample sizes: $r(59) = .493$, $p < .001$, 95%CI = .271–.665; $r(25) = .403$, $p = .046$, 95%CI = .010–.689. The effect sizes of the correlations are at the medium level (i.e., r is between .3 and .5; see Cohen, 1988). The p-value is *the probability of observing the same or more extreme data, assuming that the null hypothesis is true in the population* (e.g., van de Schoot et al., 2014). Following this, a "significant" p-value below the traditionally applied .05 threshold level (e.g., $p < .001$ for the first sample) does *not* disprove the null hypothesis (i.e., absence of evidence is not evidence of absence), say anything about the probability of a hypothesis, or indicate that whether repeating the experiment a great number of times would yield a significant result on 99% of occasions (Gigerenzer, 2004). However, frequentist 95% CIs align with the p-values, as there is no zero correlation value in the interval when the p-values stay under the .05 level. A typical 95% CI is interpreted as *if we were to collect repeatedly similar samples to our existing sample, then 95% of the time, the CIs (of randomly obtained samples) would contain the true population mean that we do not know* (e.g., Hoekstra et al., 2014). The main point is that a 95% CI is *not* the probability that the true value (i.e., population level r) is within the interval, as CIs do not make any probability statements about parameters or hypotheses. Following this, for the larger sample ($n = 59$), we would report that a small ($r > .1$) to large ($r > .5$) effect size positive relationship between the two variables would be found in 95 out of 100 (95% *proportion*, not *probability*) random samples that would be collected from the population. In other words, the true population correlation between the two variables was between .271 and .665 in 95% (of a large number) of samples drawn from the population. The result related to the smaller sample ($n = 25$) are more problematic to report: we have a positive correlation of .403 with medium effect size, but it is barely below the desired .05 level ($p = .046$), and the 95% CI ranges from very small correlation of .010 to very high correlation of .689.

Having described the findings in a frequentist style, we turn to Bayesian reporting practices (see, e.g., Appelbaum et al., 2018; van Doorn et al., 2021). Bayesian correlation analysis (as any Bayesian analysis) starts by specifying the absence (or presence) of prior knowledge, in this case regarding the latent correlation ρ (rho) of interest. Figure 10.1 shows that we used four different prior distribution settings (rows B1a, B1b, B2a, B2b) to illustrate their possible effects on the outcome (posterior distribution of ρ) with varying sample sizes (two columns labelled "$n =$ 59" and "$n = 25$").

The first prior is the so-called uniform prior, which is displayed with a flat dashed line in the two plots (based on sample sizes of 59 and 25) on the B1a row. This beta (1,1) prior distribution assigns equal weight to all possible correlation coefficients before we analysed the data. Although it has been argued that all priors are more or less "informative" (e.g., Gelman et al., 2013), this type of "non-informative" (a.k.a. "uninformative", "reference", "vague" or "diffuse") prior is used to reflect the lack of advanced knowledge on the topic. The second prior (B1b) is still uniform beta(1,1), but this time it is an "informative" prior as we have indicated that the range of expected population correlations is $0 \leq \rho \leq 1$, referring to an existing research on cultural intelligence that indicates positive correlations among CQ factors (van Dyne et al., 2009). The third prior beta(3,3) distribution in row B2a is also informative, as the population ρ values are expected to follow a distribution that concentrates around the zero value. This prior is based on existing research suggesting that cultural intelligence correlations are seldom extreme (van Dyne et al., 2009), indicating that our initial belief is to see correlations mostly from −.5 to .5 with a peak value of zero. The last row in Fig. 10.1 (B2b) displays the results when the previous beta (3,3) prior is set to be positive, with an expectation of correlations concentrating between 0 and .5.

The results with the original larger sample ("$n = 59$" column in Fig. 10.1) show that the four different priors have very little effect on the posterior distribution of correlations (median ρ range from .451 to.475). Instead of only reporting median ρ (comparable to Pearson correlation), the Bayesian approach allows us to investigate its posterior distribution visually and report a related credible interval (C.I.). Compared to the lower and upper bounds of previously discussed frequentist CIs, the interpretation of their Bayesian counterparts are quite different: *C.I. provides a 95% probability that the population ρ is between its lower and upper bounds* (van Doorn et al., 2021). For this larger sample (with four different priors), we observe that there is a 95% probability that the population correlation is between .239 and .656.

In addition to learning about the point estimates, it is usually important to know if the null (H_0 = no correlation) or alternative (H_1 = there is a correlation) hypothesis is more plausible. Instead of a *p*-value, we will report a Bayes factor (BF; see, e.g., Kass & Raftery, 1995), which is a measure that compares the likelihood of the data under the alternative hypothesis (H_1: $\rho \neq 0$) and the null hypothesis (H_0: $\rho = 0$). A comparison of the two hypotheses (i.e., models) posterior probabilities (posterior odds = BF × prior odds) is shown in Eq. 10.2. The first term on the right-hand side is the BF_{10} (a.k.a likelihood ratio), where the probability of the data given an alternative hypothesis $P(D|H_1)$ is divided by the probability of the data given a null hypothesis $P(D|H_0)$.

$$\frac{P(H_1|D)}{P(H_0|D)} = \frac{P(D|H_1)}{P(D|H_0)} \times \frac{P(H_1)}{P(H_0)} \tag{10.2}$$

The first diagram (B1a row in Fig. 10.1) shows that the BF_{10} value for the larger sample ($n = 59$, $r = .493$) is 348.783. This result may be communicated, as *the data are approximately 349 times more likely under H_1 (the two variables are correlated) than under H_0 (there is a zero correlation).* The other BF_{01} is the likelihood of the data under the null hypothesis compared to the alternative hypothesis ($1/BF_{10} = 1/348.783 = 0.003$). If we follow the B1a row to the second diagram ($n = 25$, $r = .403$), we notice that this time, the BF values are close to each other, an indication of a very small difference. This is also reflected in the frequentist *p*-value of .046. Based on Harold Jeffreys' initial work in the 1960s, Kass and Raftery (1995, p. 777) suggested the following BF_{10} threshold values: BF_{10} = 1–3 (not worth more than a bare mention); BF_{10} = 3–20 (positive/moderate); BF_{10} = 20–150 (strong); $BF_{10} > 150$ (very strong). Corresponding BF_{01} values are as follows: .1–.333 (not worth more than a bare mention); .332–.050 (positive/moderate); .049–.007 (strong); < .007 (very strong).

Figure 10.2 shows the evidence for the alternative hypotheses (BF_{10} values for two-sided H_1/H_0 and BF_{+0} values for one-sided H_+/H_0 likelihoods) with beta(1,1) and beta(3,3) prior distributions. The x-axis shows all possible prior width values from 0 to 2, and the curve line represents the related BF value. We observe from the figure that BF is robust, except for very small prior widths (e.g., a prior width of .01 would produce a beta(1/.01,1/.01) = beta(100,100) prior distribution that would

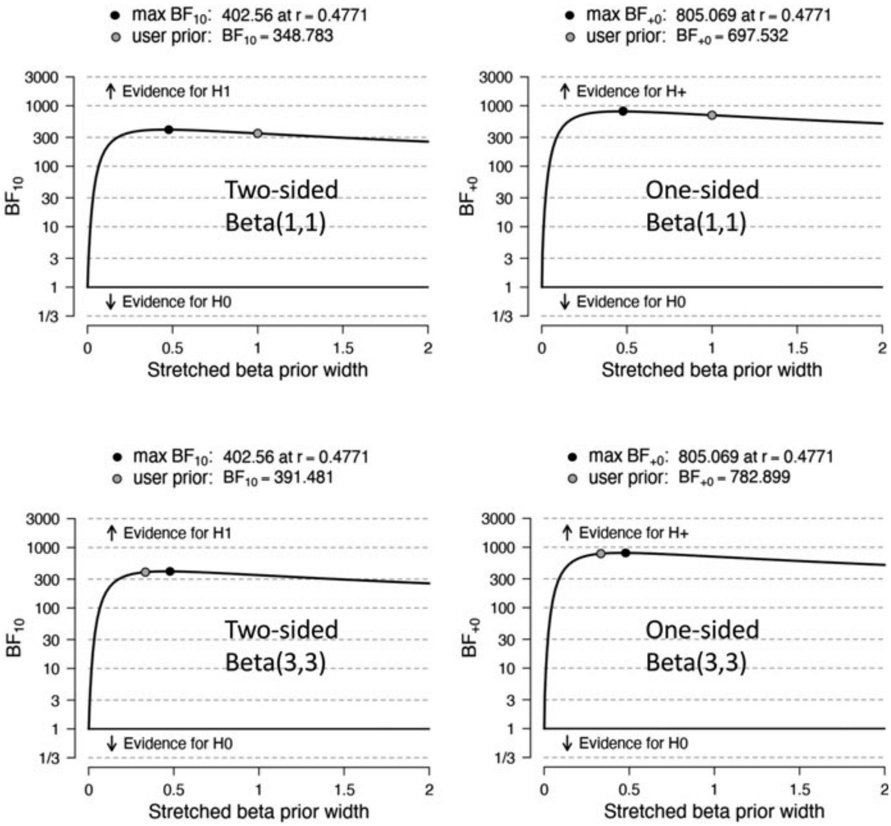

Fig. 10.2 Bayes factor robustness plots for four different prior settings ($n = 59$)

indicate a peaky concentration of correlations around zero). Computational prior values that produce maximum BF for each scenario are displayed with black dots and user-defined priors with grey-filled dots. The first row in Fig. 10.2 shows that our uniform prior beta(1,1) was not the most optimal choice to produce the most evidence to support the alternative hypotheses ($BF_{10} = 348.783$; $BF_{+0} = 697.532$). Instead, the prior width setting of beta(1/.4771, 1/.4771), i.e., beta(2,2), would have produced slightly higher likelihood ratios (403 and 805) in favour of alternative hypotheses (H_1 and H_+, respectively). However, we will not make any adjustments to our initial prior settings based on *this sample,* as that would be called “double dipping” or “Bayes hacking”.[1] The second row in Fig. 10.2 indicates that our more normal (bell-shaped) prior beta(3,3) distribution is really close to the computational maximum of beta(2,2).

[1] Temptation of “*p*-hacking” is also present in the frequentist context, referring to selecting data or statistical analyses until nonsignificant results become significant (Head et al., 2015).

As presented earlier, the use of prior information is one of the main factors that sets Bayesians apart from frequentists (and likelihoodists; see Blume, 2011). In this example, we used four different beta distributions to reflect different levels of prior knowledge about the relationship between academics' cognitive and behavioural cultural intelligence. Uniform (non-informative) prior showed that it was not actually that non-informative after all with the smaller sample ($n = 25$), as it drew the posterior distribution towards the zero correlation by giving the same weight for the (relevant) existent ($\rho > 0$) and (non-relevant) virtually non-existent ($\rho < 0$) correlations. In other words, similar prior probabilities were assigned for both negative (not expected in the real world) and positive correlations. This is clearly visible in row B1a (Fig. 10.1): The BF values for the alternative hypothesis H_1 drop rapidly when the sample size decreases ($348.783 \rightarrow 1.648$), giving more room (i.e., white space in the probability wheel) for the null hypothesis.

A takeaway message from this is that priors need to be carefully considered, especially when the sample size is small (e.g., Smid et al., 2020; Smid & Winter, 2020). However, despite the obvious dangers, they should be seen as an exciting apparatus for any PLD researcher to make the analysis more attached to real working life (by recognising expert knowledge, earlier research, etc.). This example demonstrated the power of scientifically relevant "prior positive correlation expectation" with both uniform beta(1,1) (row B1b in Fig. 10.1) and (more or less) normal beta (3,3) (row B2b) priors: with these two samples, the alternative one-way hypothesis (H_+) was approximately two times more probable than the two-way (asymptotic) alternative hypothesis (H_1). For example, the one-sided hypothesis testing with informed beta(3,3) for the smaller sample (see Fig. 10.1, rightmost column of the rows B2a and B2b) resulted in $BF_{+0} = 4.312$ (indicating moderate support for the alternative hypothesis) while the two-sided BF value was below the threshold level of 3 ($BF_{10} = 2.238$), an indication of no sufficient evidence for the alternative hypothesis.

Next, we will present an example of Bayesian multilevel path analysis of empirical data using *Mplus* software (Muthén & Muthén, 2017).

10.3 Bayesian Multilevel Path Analysis

In PLD research, hierarchical data are almost everywhere and can present themselves in individuals within groups, classes, schools, departments, or institutions. Typical examples are employees nested within organisations (or departments) and organisations nested within countries (or municipalities). It is well established that the hierarchical structure of the data should be taken into account to avoid erroneous results and misleading interpretations (Hox, 2010). Most of the statistical analyses that seek to identify the differences between variables are, in one way or another, based on the assumption of the observation independence condition (Curran, 2003). This condition states that the responses of one participant do not affect and should not be affected by others' responses. Quite often in studies on PLD, this constraint is violated. For example, in the example presented below, participants who belong to a specific department could have something in common in terms of their views on organisational culture. Thus, the response of an employee is not independent of the response of his/her colleague working in the same department.

Multilevel modeling (MLM), also known as mixed-effects models or hierarchical analysis, is an evolved statistical method for analysing hierarchical (clustered) data (e.g., Pinheiro & Bates, 2000; Gelman & Hill, 2007). This multilevel form of SEM (e.g., Kaplan & Depaoli, 2012) was developed to deal with the violation of observation independency by removing the common variance among members of the same cluster (e.g., departmental-level variance). MLM is commonly used in PLD research, where the focus is on filtering out the variation that is attributed to each level from the others (i.e., individual and group levels). By doing so, the researcher can examine the relationships within each level and between levels (interactions).

Now that we have quite good groundings to use MLM for our data, we might next ask, why do we use Bayesian instead of the traditional frequentist version of MLM? In empirical PLD research, sample sizes tend to be small (e.g., due to naturally small populations), the models quite complex (to model complex reality), and the distribution of data skewed (e.g., towards positive response scale values in self-assessment surveys). Unfortunately, frequentist estimation, as the name says, needs larger samples as the models become more complex, leading to nonconvergence, inadmissible parameter solutions, and inaccurate estimates (Smid et al., 2020). Bayesian estimation offers a solution to these problems, as it does not rely on large-sample theory (that assumes the distribution of the parameter estimate to be normal) but rather provides the whole distribution of the parameter and assumes a symmetric distribution; it also offers credible intervals based on the percentiles of the posterior (allowing the presence of strongly skewed distributions) (Muthén & Asparouhov, 2012).

The following example of Bayesian multilevel path analysis shows how to examine the effect of psychological (implicit theory and goal orientations) and organisational factors (organisational culture) on university staff members' innovativeness (Aldahdouh, 2020). The dependent variable in this study is individual innovativeness, which refers to one's willingness to change and embrace novel ideas. Innovativeness has been defined in three approaches: behavioural, psychological, and domain-specific (see Aldahdouh, 2020 for more details). The presented study example focuses on innovativeness as a psychological characteristic.

The innovativeness concept sprang from Rogers' (1962) diffusion of innovation theory. In its early stages, innovativeness was studied in the fields of marketing and business to predict consumer behaviour when purchasing new products. Later works have been extended to include professional development in higher education (Aldahdouh et al., 2019; Gökçearslan et al., 2017), management (Jong & Hartog, 2007), and the health sector (Park & Kim, 2010). The idea of studying innovativeness in the PLD context was to investigate the employee's behaviour or responses when they encounter new challenges or changes in their regular work practices (Messmann & Mulder, 2017). The concept of innovativeness overlaps with multiple closely related and well-researched concepts in the field, such as creativity (Tierney

Table 10.1 Study variables, abbreviations, and definitions

Variable	Abbreviation	Definition
Innovativeness	INNOV	Refers to an individual's willingness to change
Entity theory of ability	ETA	Refers to an individual's beliefs that the human attributes are fixed, innate and stable
Mastery goal orientation	MAS	Refers to an individuals' tendency to engage in a task in order to improve their own capacities and to sharpen skills
Performance-approach goal orientation	PAP	Refers to an individuals' tendency to engage in a task in order to show others how well they can do or to overtake their peers
Performance-avoidance goal orientation	PAV	Refers to an individuals' tendency to engage in a task in order to avoid appearing incompetent in comparison to their peers
Clan culture	CLN	Refers to a culture that focuses on internal flexibility
Hierarchy culture	HRC	Refers to a culture that focuses on internal stability
Market culture	MRK	Refers to a culture that focuses on external stability
Adhocracy culture	ADH	Refers to a culture that focuses on external flexibility

& Lanford, 2016) and entrepreneurship (Goller & Paloniemi, 2022). Table 10.1 shows a definition of the study variables, including innovativeness, and the independent variables used to predict innovativeness.

Based on the previous studies, we proposed the following hypotheses:

Hypothesis 1. The entity theory of ability and performance-avoidance goal orientation contribute negatively to predicting innovativeness, while mastery goal orientation contributes positively to predicting innovativeness.

Hypothesis 2. The entity theory of ability is negatively associated with mastery goal orientation and positively associated with performance-avoidance goal orientation.

Hypothesis 3. The clan and adhocracy cultures contribute positively to predicting innovativeness, while the hierarchy culture contributes negatively to predicting innovativeness.

10.3.1 Participants and Measures

The data were collected in 2019 using a non-probability sampling method from 315 staff members who were working in 34 different departments of a Finnish university via an online survey. The survey consisted of demographic questions and the following established scales: innovativeness (Hurt et al., 1977), organisational culture (Cameron & Quinn, 2006), goal orientations (Midgley et al., 2000), and implicit theories of ability (Levy et al., 1998).

10.3.2 Analysis Approach

We conducted a Bayesian multilevel path analysis using *Mplus* version 8.0 (Muthén & Muthén, 2017) to test the four above-mentioned hypotheses. A multilevel approach was warranted, given that our data had a nested structure: 315 responses of individuals working in 34 schools/departments. We opted to use path analysis because the study variables were assumed to have structural dependencies among the predictor variables, apart from their effects on the outcome variable. We followed the within-and-between approach in multilevel path analysis, wherein estimates for the within-covariance matrix (individual level) and the between-covariance matrix (group level) were determined separately (Hox, 2010). This made it possible to partial out the group-level variance from individual-level variables. Multilevel path analyses of both levels were conducted separately but simultaneously. Summary scores of the variables were used in the analysis because, due to the small number of groups in this study, it was not feasible to conduct an analysis on latent variables. The Bayesian approach was chosen because of its superior performance in small samples (Stegmueller, 2013). Unlike frequentist (inferential) techniques, the Bayesian approach does not rely on any distributional assumptions about the data, such as normality (Finch et al., 2014).

10.3.3 Analysis Settings

The Markov Chain Monte Carlo (MCMC; e.g., van Ravenzwaaij et al., 2018) methodology was implemented to obtain the parameter estimates in the analysis. The convergence of parameter estimates was assessed by the potential scale reduction (PSR; e.g., Gelman et al., 2013) convergence criterion, where values below 1.1 for each parameter indicate that the convergence of the MCMC sequence has been reached. The convergence was also monitored using trace plots, in which quick oscillations indicate convergence. Autocorrelation plots were used to check for the correlation between two adjacent MCMC draws and to set the thinning value. If the values are highly autocorrelated, it may take many iterations (a long time) before a sample that sufficiently represents the entire range of the posterior distribution is created (Erler et al., 2021). We used a thinning value of 4 (i.e., kept every fourth value and discarded all other values) to minimise the correlation to near zero.

Model fit was assessed using the posterior predictive p-value (PPP; Meng, 1994) and the credible interval (C.I.; see, e.g., Kaplan & Depaoli, 2012). A plausible PPP value is within the interval 0.3–0.7 (Lee & Song, 2004), and a value close to 0.5 indicates optimal fit (Finch et al., 2014). A 95% C.I. that contains a zero (ideally, at the centre of the interval) indicates that the actual and simulated data do not differ from each other (i.e., producing a desirable output). The deviance information criterion (DIC) was used for model comparisons (the model with the lowest DIC value is preferable; see, e.g., Gelman & Hill, 2007).

The analysis was conducted using *Mplus* default priors (normal distribution with a mean hyperparameter of zero and a variance of 10^{10}) as there were no previous data available on the topic. It is well known that the default “non-informative” priors may, in fact, act as highly informative priors with small samples, as they can heavily influence the posterior distribution and impact the conclusions of a study (for further discussion, see, e.g., Smid & Winter, 2020). As the sample size of the current study is 315, the ratio of sample size to the number of estimated parameters for the two models (Model 1: 11.7 = 315/27; Model 2: 10.2 = 315/31) are well above the value of 5 (Lee & Song, 2004), indicating that the sample size is not small, but instead relatively large.

Using Gibbs sampling, two MCMC chains of a minimum of 45,000 iterations (discarding half of the iterations as burn-in; see Muthén & Muthén, 2017) were used to draw samples from the posteriori distribution of the model parameters.

10.3.4 Statistical Procedures

We tested two models while analysing the data: random intercept and random slopes models. First, we tested the random intercept model (Fig. 10.3) and assessed the fitness of the model by computing the PPP. Then, we tested the random slopes model (Fig. 10.4) by allowing the slopes to vary across departments. This process had two objectives. One was to ensure simplicity by testing the models from simple to more complex structures. The other was to check the PPP model fit value, which was not available except for the random intercept model. We then compared the DIC value of the random intercept model with the DIC values of the random slopes model.

10.3.5 Data Aggregation

The intra-class correlation coefficient (ICC1; Bliese, 2000) was calculated to examine whether there were department-level variances in the study variables that necessitated their inclusion in the between-level model. ICC1 represents the proportion of group-level variance with respect to the total variance of the variable. Variables showing ICC1 > 0.05 were included in the between-level model (LeBreton & Senter, 2008). Although we considered the entity theory of ability and goal orientations individual characteristics, we computed ICC1 for their respective variables because they may differ significantly across departments due to the study’s sampling method. As hypothesised, however, the ICCs for those individual variables showed almost no variance according to department membership (all ICC1s < 0.03). Thus, they were included only in the within-level model. Furthermore, we calculated ICC1 for the outcome variable (innovativeness) to determine whether individual innovativeness was affected by department membership (Bliese, 2000). The results revealed that 10% of the variance in innovativeness was due to department membership.

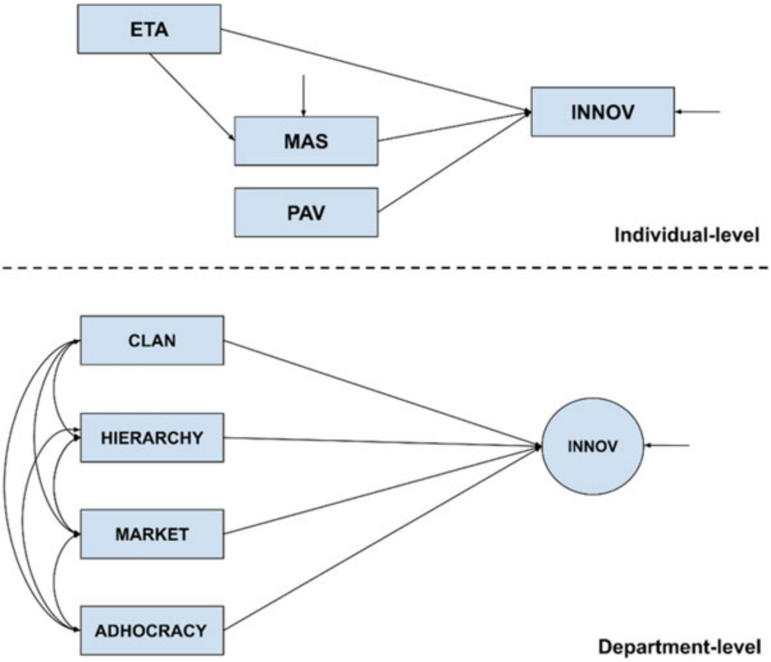

Fig. 10.3 Random intercept model
Note: *INNOV* Innovativeness, *ETA* Entity Theory of Ability, *MAS* Mastery goal orientation, *PAV* Performance-Avoidance goal orientation. Performance-Approach goal orientation as well as the path from ETA to PAV were omitted based on the correlation findings

Our intention for the cultural variables was to measure the common perceptions of culture in each department. We were interested in the means of each department, and not individual perspectives on what the departmental culture was. However, cultural variables were measured through the ratings given by individuals in the department. To justify the aggregation of these cultural variables to their departments' means, we used a calculator developed by Biemann et al. (2012) to compute median r_{WG} values using the null uniform distribution (Bliese, 2000). The r_{WG} value indicates the degree of agreement among staff members within a department. Values greater than 0.70 indicate generally accepted agreement among raters (LeBreton & Senter, 2008). Using the same tool, we determined ICC1 in addition to the reliability of the group means (ICC2). The results were as follows: for clan culture, $r_{\mathrm{WG}} = 0.88$, ICC1 = 0.08, and ICC2 = 0.44; for hierarchy culture, $r_{\mathrm{WG}} = 0.88$, ICC1 = 0.08, and ICC2 = 0.45; for Market culture, $r_{\mathrm{WG}} = 0.89$, ICC1 = 0.16, and ICC2 = 0.64; and for adhocracy culture, $r_{\mathrm{WG}} = 0.88$, ICC1 = 0.09, and ICC2 = 0.47. The F-ratios associated with the ICC values were all statistically significant at the 0.05 level. The ICC1 and r_{WG} values of all cultural variables were above the cut-off values. The ICC2 values ranged between 0.44 and 0.64, classified by Fleiss (1986, p. 7) as fair-to-good reliability estimates (ICC2 values <0.40 are poor, those between 0.40 and

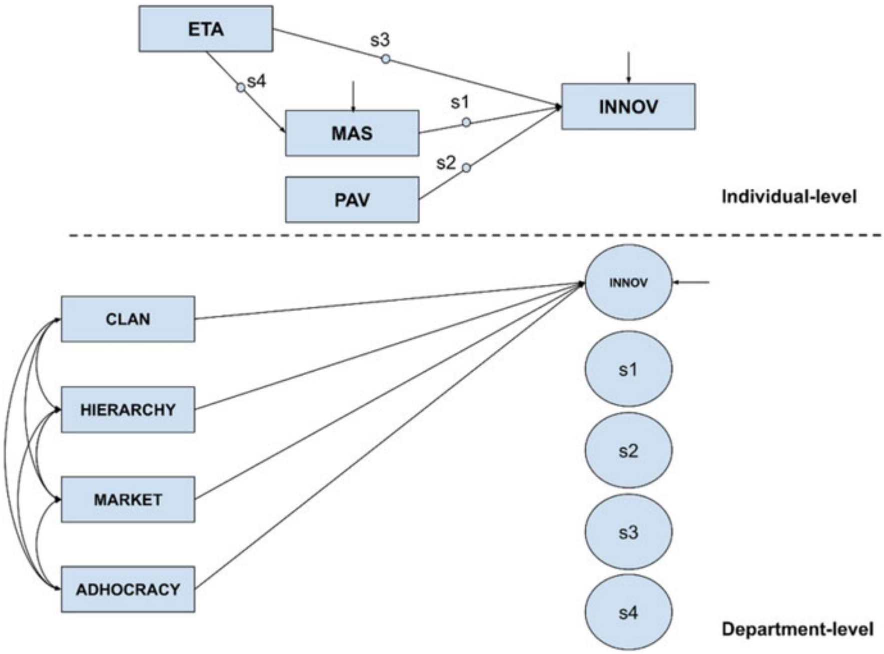

Fig. 10.4 Random slopes model
Note: *INNOV* Innovativeness; *ETA* Entity Theory of Ability; *MAS* Mastery goal orientation; *PAV* Performance-Avoidance goal orientation. Performance-Approach goal orientation as well as the path from ETA to PAV were omitted based on the correlation findings

0.75 are fair-to-good, and those >0.75 are excellent). Based on the results, we decided to aggregate the cultural variables.

We followed the recommendation of Enders and Tofighi (2007), who suggested centring the individual-level (level 1) variables (e.g., "mastery goal orientation") on their group mean (CWC, centring within cluster) when the focus is on inspecting associations between level 1 variables. In CWC, each participant's department mean value was subtracted from the participant's score. Department-level variables (e.g., "clan culture") were centred on the grand mean (CGM, see Hox, 2010) by subtracting the mean of the full sample from each variable.

10.3.6 Descriptive Statistics

Table 10.2 displays the means, standard deviations, and correlations among the variables at the individual and department levels. Bayes factor (BF_{01}) values in Table 10.2 less than .333 are bolded, indicating moderate to very strong evidence for the existence of correlation (Kass & Raftery, 1995). An inspection of the correlations revealed that mostly the relationships between innovativeness and the psychological variables were significant, except for PAP that was excluded from further analysis. Similarly, ETA showed a non-significant relationship with PAV. Thus, the regression coefficient between ETA and PAV was cancelled out in the examined models.

10.3.7 Results

10.3.7.1 Random Intercept Model

We were guided by the hypotheses and the correlation matrix in specifying the paths between the variables. At the individual level, we examined the model in which ETA, MAS, and PAV were predictors of innovativeness, while ETA was a predictor of MAS. At the department level, we examined the extent to which clan, hierarchy, market, and adhocracy cultures explained the variance in the random intercept of innovativeness. Equations (10.3), (10.4), and (10.5) below represent the model:

$$\mathrm{INNOV}_{ij} = \beta_{0j} + \beta_{10}\mathrm{MAS}_{ij} + \beta_{20}\mathrm{PAV}_{ij} + \beta_{30}\mathrm{ETA}_{ij} + \mathrm{e}_{ij} \tag{10.3}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{CLN}_j + \gamma_{02}\mathrm{HRC}_j + \gamma_{03}\mathrm{MRK}_j + \gamma_{04}\mathrm{ADH}_j + u_{0j} \tag{10.4}$$

$$\mathrm{MAS}_{ij} = \alpha_{00} + \alpha_{10}\mathrm{ETA}_{ij} + \epsilon_{ij} \tag{10.5}$$

In Eq. (10.3), the intercept β_{0j} is a random effect that varies across departments, while the slopes β_{10}, β_{20}, and β_{30} are fixed. Cultures at the department level predict the intercept of innovativeness β_{0j}. Equation (10.5) allows the ETA to predict the MAS, in which the intercept α_{00} and the slope α_{10} are fixed. ETA was not allowed to predict PAV because ETA had no correlation with PAV, as shown in the correlation matrix (see Table 10.2).

The parameter estimates all converged adequately as the PSR values decreased smoothly over the iterations, reaching a value of 1.010, which is below the cut-off value of 1.05. The model showed a good fit to the data, as the PPP was 0.278 (above the cut-off value of .10), and the 95% C.I. for the difference between the observed and the replicated χ^2 values covered zero, with a lower bound of −16.494 and an upper bound of 32.025. The DIC value was 1143.358.

As shown in Table 10.3, ETA was negatively associated with MAS ($\alpha_{10} = -0.149$) and INNOV ($\beta_{30} = -0.083$), while MAS was positively associated with INNOV ($\beta_{10} = 0.261$). As expected, PAV was negatively associated with INNOV ($\beta_{20} = -0.097$). Although the individual-level variables maintained a significant association with INNOV, a significant value of δ^2_{eij} may suggest that there remained a variance in INNOV that had not yet been explained.

Contrary to our expectations, none of the cultures explained the variance of the innovativeness's random intercept, even though a random effect of the intercept (δ^2_{u0j}) pointed to a significant variation in the intercept (γ_{00}) between departments.

Table 10.2 Means, standard deviations, and two-tailed correlations among the variables at the individual and department levels

		1	2	3	4	5	6	7	8	9
1. INNOV	*r*	1					−0.077	0.259	−0.098	−0.081
	BF_{01}						6.847	2.546	6.455	6.781
2. ETA	*r*	−0.157	1							
	BF_{01}	0.455								
3. MAS	r	**0.329**	−0.151	1						
	BF_{01}	0.000	0.620							
4. PAP	*r*	−0.023	0.067	0.067	1					
	BF_{01}	20.536	10.973	11.118						
5. PAV	*r*	**−0.173**	0.055	0.094	**0.667**	1				
	BF_{01}	0.192	13.827	5.588	0.000					
6. CLN	*r*						1	0.026	**−0.489**	**0.489**
	BF_{01}							7.435	0.105	0.104
7. HRC	*r*							1	−0.010	−0.303
	BF_{01}								7.500	1.668
8. MRK	*r*								1	0.180
	BF_{01}									4.488
9. ADH	*r*									1
	BF_{01}									
	M	3.75	3.73	3.95	2.22	2.67	3.02	2.87	2.49	2.94
	SD	0.565	0.861	0.693	0.830	0.948	0.322	0.243	0.430	0.295

Note: *INNOV* Innovativeness; *ETA* Entity Theory of Ability; *MAS* Mastery goal orientation; *PAP* Performance-Approach goal orientation; *PAV* Performance-Avoidance goal orientation; *CLN* Clan culture; *HRC* Hierarchy culture; *MRK* Market culture; *ADH* Adhocracy culture; *PC* Pearson correlation; BF_{01} Bayes factor (ratio of H_0/H_1 less than .333 indicates moderate evidence for H_1 and values less than .049 indicate strong evidence). Values below the diagonal are correlations at the individual level ($n = 315$); values above the diagonal are correlations at the department level ($n = 34$)

Table 10.3 Bayesian parameter estimates and credible intervals of the random intercept model

Path	Estimate (*SD*)	95% Credible Interval		Significance
		Lower	Upper	
Within-level				
MAS → INNOV (β_{10})	0.261 (0.044)	0.175	0.347	*
PAV → INNOV (β_{20})	−0.097 (0.032)	−0.159	−0.034	*
ETA → INNOV (β_{30})	−0.083 (0.035)	−0.153	−0.014	*
ETA → MAS (α_{10})	−0.149 (0.045)	−0.237	−0.062	*
Residual variances				
INNOV ($\delta^2{}_{\text{e}ij}$)	0.244 (0.021)	0.208	0.289	*
MAS ($\delta^2{}_{\varepsilon ij}$)	0.418 (0.034)	0.359	0.492	*
Between-level				
CLN → INNOV (γ_{01})	−0.380 (0.266)	−0.900	0.147	
HRC → INNOV (γ_{02})	0.308 (0.231)	−0.139	0.773	
MRK → INNOV (γ_{03})	−0.290 (0.173)	−0.625	0.056	
ADH → INNOV (γ_{04})	0.244 (0.259)	−0.263	0.759	
Intercepts				
INNOV (γ_{00})	3.758 (0.048)	3.662	3.853	*
Residual variances				
INNOV ($\delta^2{}_{u0j}$)	0.040 (0.021)	0.014	0.095	*

A significant overall fixed intercept γ_{00}, which is the expected value of INNOV when all predictors are on their means, suggested that the intercept was significantly different from zero.

10.3.7.2 Random Slope Model

We allowed the slopes of the relationships between the psychological variables and innovativeness to vary across departments in a random slopes model. The slope of the ETA on MAS was permitted to vary as well. The rest of the model remained as it was in the random intercept model to allow a comparison of the two models using the DIC value.

$$\text{INNOV}_{ij} = \beta_{0j} + \beta_{1j}\text{MAS}_{ij} + \beta_{2j}\text{PAV}_{ij} + \beta_{3j}\text{ETA}_{ij} + \text{e}_{ij} \tag{10.6}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{CLN}_j + \gamma_{02}\text{HRC}_j + \gamma_{03}\text{MRK}_j + \gamma_{04}\text{ADH}_j + u_{0j} \tag{10.7}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{10.8}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \tag{10.9}$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \tag{10.10}$$

$$MAS_{ij} = \alpha_{00} + \alpha_{1j}ETA_{ij} + \epsilon_{ij} \quad (10.11)$$

$$\alpha_{1j} = \lambda_{10} + \epsilon_{1j} \quad (10.12)$$

The slopes β_{1j}, β_{2j}, and β_{3j} were random effects that varied across departments in this model (see Eq. 10.6). The slopes β_{1j}, β_{2j}, and β_{3j} were functions of fixed intercepts (γ_{10}, γ_{20}, and γ_{30}) and random variances (u_{1j}, u_{2j}, and u_{3j}), while no variables were assigned to predict those slopes, as shown in Eqs. (10.8), (10.9), and (10.10). Similarly, the slope α_{1j} was a function of the fixed intercept λ_{10} and a random part ϵ_{1j}, as shown in Eq. (10.12).

Good MCMC convergence was manifested by (a) a steady decrement in the PSR values to values close to 1 for the last few tens of thousands of iterations, (b) tight horizontal bands for the parameter estimation in the trace plots, and (c) low dependence in the chain in the autocorrelation plots (Hox, 2010). The DIC value was 1122.479, which is clearly lower (desired difference should be larger than 7 units, see Cain & Zhang, 2019) than in the previous model (1143.358–1122.479 = 20.879), allowing the slopes to vary across departments, leading to a better fit to the data.

In this model, only MAS and PAV appeared to have significant positive ($\gamma_{10} = 0.266$) and negative effects ($\gamma_{20} = -0.089$), respectively, on INNOV (see Table 10.4). The variances of the slopes (u_{1j}, u_{2j}, u_{3j}) were significant, indicating significant variations between departments in the relationships between the psychological variables and innovativeness.

10.3.8 Conclusions

As demonstrated above, the analysis examined the influence of the entity theory of ability, mastery goals, and performance-avoidance goal orientations on innovativeness. In the random intercept model, we assumed that the psychological factors (ETA, MAS, PAV, and INNOV) influenced each other with fixed relationships, which were not allowed to vary across departments. In this model, ETA showed a significantly negative effect on MAS and INNOV. MAS and PAV showed positive and negative effects, respectively, on INNOV. In the random slopes model, when we allowed the relationships among the psychological factors to vary across departments, all relationships retained their significance, except for the influence of ETA on other factors (MAS and INNOV). Thus, the results partially supported Hypothesis 1 but failed to support Hypothesis 2. Our findings challenge the results reported in many previous studies that confirmed the relationship between ETA and MAS (Aldahdouh et al., 2018; Chen & Pajares, 2010; De Castella & Byrne, 2015). It is worth mentioning that these studies did not consider the hierarchical structure of the data and thus reported results that were similar to our results in the random intercept model. De Castella and Byrne (2015), for example, reported a significant relationship between the entity theory of intelligence and MAS by sampling 680 Australian

Table 10.4 Bayesian parameter estimates and credible intervals of the random slopes model

Path	Estimate (*SD*)	95% Credible Interval		Significance
		Lower	Upper	
Within-level				
Residual variances				
INNOV ($\delta^2{}_{eij}$)	0.218 (0.020)	0.183	0.262	*
MAS ($\delta^2{}_{\varepsilon ij}$)	0.388 (0.033)	0.331	0.459	*
Between-level				
CLN → INNOV (γ_{01})	−0.375 (0.263)	−0.887	0.146	
HRC → INNOV (γ_{02})	0.314 (0.229)	−0.130	0.773	
MRK → INNOV (γ_{03})	−0.284 (0.171)	−0.615	0.056	
ADH → INNOV (γ_{04})	0.247 (0.256)	−0.263	0.746	
Intercepts				
INNOV (γ_{00})	3.772 (0.050)	3.675	3.872	*
Means				
MAS → INNOV slope (γ_{10})	0.266 (0.062)	0.143	0.390	*
PAV → INNOV slope (γ_{20})	−0.089 (0.044)	−0.177	−0.002	*
ETA → INNOV slope (γ_{30})	−0.072 (0.043)	−0.158	0.013	
ETA → MAS slope (λ_{10})	−0.124 (0.068)	−0.256	0.010	
Variances				
MAS → INNOV slope (u_{1j})	0.044 (0.035)	0.006	0.138	*
PAV → INNOV slope (u_{2j})	0.023 (0.016)	0.005	0.065	*
ETA → INNOV slope (u_{3j})	0.012 (0.014)	0.001	0.051	*
ETA → MAS slope (ϵ_{1j})	0.059 (0.037)	0.014	0.158	*
Residual variances				
INNOV ($\delta^2{}_{u0j}$)	0.042 (0.021)	0.016	0.098	*

students from five different high schools while overlooking the fact that the sample had a hierarchical structure (students nested within classes, and classes nested within schools). Neglecting the hierarchical structure of the data might produce misleading results (Hox, 2010). This study sheds light on the importance of accounting for the group level while analysing the effects of individual factors.

Besides the informative findings of this study, we emphasise the methodological approach as an example of utilising Bayesian statistics in researching professional development studies. The issue of sample size occurred in the present study at the group level. The recommendation was to have at least 50 groups at the group level (Maas & Hox, 2005). Our data, however, consisted of only 34 groups. Thus, it seemed challenging to proceed with multilevel modeling using traditional estimators, such as maximum likelihood. Further, we assumed that this challenge would present itself in many studies in the research field of PLD. Adding more groups, such as those in the current study, is beyond the researcher's control. Therefore, we resorted to the Bayesian approach, which proved effective in handling the analysis of small sample sizes.

10.4 Discussion

This chapter described the potential of Bayesian methods[2] in professional learning and development research. Bayesian approach is strongly recommended if a PLD researcher is forced to operate with a very small or small sample (e.g., the ratio of sample size to the number of estimated parameters is equal to or less than 5) *and* he/she has access to prior information related to the research question (e.g., expert opinions or previous comparable studies). Although the Bayesian approach has been applauded in many research fields as *the* small sample size method, a responsible researcher carefully considers the issues related to the best practices (e.g., the "When-to-Worry-and-How-to-Avoid-the-Misuse-of-Bayesian-Statistics" checklist by Depaoli & van de Schoot, 2017) to avoid a "contamination" of the method, as we have witnessed with misunderstandings related to the frequentist statistics "*p*-values" and "confidence intervals" during the past 100 years (see, e.g., Kruschke & Liddell, 2018). It should be stressed that the Bayesian approach does not make miracles with small samples when relevant prior knowledge is not present (as uniform or non-informative priors may have a stronger than expected effect on the posterior probability; see Smid & Winter, 2020). A recent review has shown that, quite often, applied researchers still use the "non-informative" default priors of statistical programs (van de Schoot et al., 2017).

When the sample size is not an issue (i.e., it is moderate or large), the results of frequentist and Bayesian analyses will most likely lead to the same conclusions (e.g., Wakefield, 2013). The question, then, is how a researcher wishes to communicate the results to the audience. A frequentist would display point estimates (such as Pearson correlation coefficient r), p-values (of null hypothesis significance tests), confidence intervals (CI), and effect sizes (e.g., Cohen, 1988). A Bayesian would show posteriori point estimates (e.g., posteriori median) but also related posteriori distributions alongside credible intervals (C.I.) and Bayes factor values (BF). As Bayesian analyses are based on observed data (instead of sampling distributions of imaginary data), they provide answers that may be easier to communicate to the audience. For example, the result of the one-sided correlation test between cognitive and behavioural cultural intelligence (see Fig. 10.1, first column of B1b row) could be communicated as follows: *Based on existing research, a one-sided Bayesian correlation analysis with uniform prior setting was conducted to investigate whether self-assessed cognitive and behavioural cultural intelligence are positively correlated in a sample of 59 academics. The results showed a correlation of .49 with a posterior median of .48 between the two summative scales. Credible interval of*

[2] Although our discussion has focused on frequentist and Bayesian paradigms, there are also other paradigms, including the likelihood paradigm (e.g., Blume, 2011). Likelihood and Bayesian paradigms are similar concerning the use of likelihood function and likelihood ratio with only observed data (in accordance with likelihood principle), but the likelihoodists would not use prior information in their calculations. Frequentists would also calculate likelihood functions but would not follow the likelihood principle (i.e., they would use both observed and unobserved data in the analysis).

posteriori distribution showed that there is 95% probability that a true correlation in the population lies between .26 and .66. This is a plausible assumption, as the presence of a positive correlation given the current data is 698 times more likely than a zero or negative correlation.

Other motives of using Bayesian statistics include its flexibility in handling complex and demanding data structures (Smid et al., 2020), robustness in analysing data that contains both linear and non-linear correlations among the variables (Nokelainen & Ruohotie, 2009), and its capability to include missing data in the analysis by assigning it distributional properties (Gill & Witko, 2013). When we wish to impute missing data, the most recommended method of multiple imputation quite often utilises Bayesian statistics to predict multiple complete datasets out of the original incomplete data (Erler et al., 2021).

In addition to issues presented above, we expect to see further growth in the application of Bayesian statistics in PLD research in the next years, as there are technological supportive factors that encourage researchers in this direction. *First*, many mainstream software packages (e.g., *SAS*, *Stata*, *SPSS*) have started to integrate a growing number of Bayesian analysis features into their interfaces. For example, since its version 25 released in 2017, *IBM SPSS* has offered "Bayesian versions" of the t-test, Pearson correlation, linear regression, and analysis of variance. Moreover, as Bayesian computational methods have become increasingly available in popular structural equation modeling software packages, such as *Mplus* (Muthén & Muthén, 2017), researchers no longer need to be (most commonly *R*) programmers to apply Bayesian analysis. Rather, the current software and packages have simplified the Bayesian computations in different kinds of analyses, such as Bayes factor (Kass & Raftery, 1995), Bayesian structural equation modeling (Kaplan & Depaoli, 2012), and more recently in Bayesian multilevel analysis (Mai & Zhang, 2018), which was on the focus in the analysis example of the current chapter.

Second, the rapid development of available software packages (and related tutorials, see https://cran.r-project.org/web/views/Bayesian.html) and textbooks (e.g., Albert, 2009) in the free *R* statistical programming environment ensure that PLD researchers have access to all complexity levels of applying Bayesian statistics in their research work. Currently, the *R* environment offers over 20,000 software packages through the *Comprehensive R Archive Network* (CRAN, https://cran.r-project.org). The *R* environment also contains packages that link to other seminal Bayesian software (e.g., *WinBUGS*, *JAGS,* and *Stan*). Among the most "easy to use" Bayesian packages in the *R* environment for multilevel modeling are *brms* (Bürkner, 2017) and *JointAI* (Erler et al., 2021). *brms* is based on *Stan* (Stan Development Team, 2021), and *JointAI* is based on *JAGS* (https://mcmc-jags.sourceforge.io). The most popular package for structural equation modeling is *lavaan* (Rosseel, 2012), accompanied by its Bayesian counterpart, *blavaan* (Merkle et al., 2021). Further, *jamovi* (The jamovi project, 2021) and *JASP* (The JASP Team et al., 2022) serve as excellent examples of standalone software that provide a graphical "SPSS-like" drag and drop interface for both frequentist and Bayesian analyses.

References

Albert, J. (2009). *Bayesian computation with R*. Springer.

Aldahdouh, T. Z. (2020). *Individual innovativeness in higher education* [Doctoral dissertation, Tampere University]. https://trepo.tuni.fi/bitstream/handle/10024/120354/978-952-03-1539-9.pdf?sequence=2&isAllowed=y

Aldahdouh, T., Nokelainen, P., & Korhonen, V. (2018). Innovativeness of staff in higher education: Do implicit theories and goal orientations matter? *International Journal of Higher Education, 7*(2), 43–57. https://doi.org/10.5430/ijhe.v7n2p43

Aldahdouh, T. Z., Korhonen, V., & Nokelainen, P. (2019). What contributes to individual innovativeness? A multilevel perspective. *International Journal of Innovation Studies, 3*(2), 23–39. https://doi.org/10.1016/j.ijis.2019.06.001

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.

Bernardo, J., & Smith, A. (2000). *Bayesian theory*. Wiley.

Biemann, T., Cole, M. S., & Voelpel, S. (2012). Within-group agreement: On the use (and misuse) of r_{WG} and $r_{WG(J)}$ in leadership research and some best practice guidelines. *The Leadership Quarterly, 23*(1), 66–80.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass.

Blume, J. D. (2011). Likelihood and its evidential framework. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Philosophy of statistics* (pp. 493–511). Elsevier.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Bürkner, P. C., & Charpentier, E. (2020). Modeling monotonic effects of ordinal predictors in Bayesian regression models. *The British Journal of Mathematical and Statistical Psychology, 73*(3), 420–451.

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(1), 39–50.

Cameron, K. S., & Quinn, R. E. (2006). *Diagnosing and changing organizational culture based on competing values framework* (revised edition). Jossey-Bass.

Chen, J., & Pajares, F. (2010). Implicit theories of ability of Grade 6 science students: Relation to epistemological beliefs and academic motivation and achievement in science. *Contemporary Educational Psychology, 35*(1), 75–87. https://doi.org/10.1016/j.cedpsych.2009.10.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Congdon, P. (2001). *Bayesian statistical modeling*. Wiley.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529–569.

De Castella, K., & Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *European Journal of Psychology of Education, 30*(3), 245–267. https://doi.org/10.1007/s10212-015-0244-y

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics. *Psychological Methods, 22*(2), 240–261.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138.

Erler, N. S., Rizopoulos, D., & Lesaffre, E. M. E. H. (2021). JointAI: Joint analysis and imputation of incomplete data in R. *Journal of Statistical Software, 100*(20), 1–56.

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Chapman and Hall/CRC.

Fleiss, J. L. (1986). *Design and analysis of clinical experiments*. Willey.

Franzin, A., Sambo, F., & di Camillo, B. (2017). bnstruct: an R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics, 33*(8), 1250–1252.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*, 587–606.

Gill, J. (2007). *Bayesian methods. A social and behavioral sciences approach* (2nd ed.). Chapman and Hall/CRC.

Gill, J., & Witko, C. (2013). Bayesian analytical methods: A methodological prescription for public administration. *Journal of Public Administration Research & Theory, 23*(2), 457–494.

Gökçearslan, Ş., Karademir, T., & Korucu, A. T. (2017). Preservice teachers' level of web pedagogical content knowledge: Assessment by individual innovativeness. *Journal of Educational Computing Research, 55*(1), 70–94.

Goller, M., & Paloniemi, S. (2022). Agency: Taking stock of workplace learning research. In C. Harteis et al. (Eds.), *Research Approaches on workplace learning. Professional and practice-based learning*. Springer.

Hastings, K. J. (1997). *Probability and statistics*. Addison-Wesley.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of P-hacking in science. *PLoS Biol, 13*(3), e1002106.

Hoekstra, R., Morey, R. D., Roude, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review, 21*, 1157–1164.

Hox, J. (2010). *Multilevel analysis. Techniques and applications*. Routledge.

Hurt, H. T., Joseph, K., & Cook, C. D. (1977). Scales for the measurement of innovativeness. *Human Communication Research, 4*(1), 58–65.

Jong, J. P. J., & Hartog, D. N. (2007). How leaders influence employees' innovative behaviour. *European Journal of Innovation Management, 10*(1), 41–64.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In *Handbook of structural equation modeling* (pp. 650–673). The Guilford Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

König, C., & van de Schoot, R. (2018). Bayesian statistics in educational research: A look at the current state of affairs. *Educational Review, 70*(4), 486–509.

Kruschke, J. K. (2015). *Doing Bayesian data analysis. A tutorial with R, JAGS, and Stan*. Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*, 178–206.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852.

Lecat, A., Beausaert, S., & Raemdonck, I. (2018). On the relation between teachers' (in)formal learning and innovative working behavior: The mediating role of employability. *Vocations and Learning, 11*, 529–554.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*(4), 653–686.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421–1436.

Lindley, D. V. (1971). *Making decisions*. Wiley.

Maas, C. J., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Journal of Research Methods for the Behavioral and Social Sciences, 1*, 86–92.

Mai, Y., & Zhang, Z. (2018). Software packages for Bayesian multilevel modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 1–9.

Mayo, D. G., & Spanos, A. (2011). Error statistics. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Philosophy of statistics* (pp. 153–198). Elsevier.

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics, 22*(3), 1142–1160.

Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian structural equation modeling in Stan. *Journal of Statistical Software, 100*(6), 1–22.

Messmann, G., & Mulder, R. H. (2017). Proactive employees: The relationship between work-related reflection and innovative work behaviour. In M. Goller & S. Paloniemi (Eds.), *Professional and Practice-based Learning* (Vol. 20, pp. 141–159). Springer.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). *Manual for the patterns of adaptive learning scales*. The University of Michigan.

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313–335.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Myllymäki, P., Silander, T., Tirri, H., & Uronen, P. (2002). B-Course: A web-based tool for Bayesian and causal data analysis. *International Journal on Artificial Intelligence Tools, 11*(3), 369–387.

Nokelainen, P., & Ruohotie, P. (2009). Non-linear modeling of growth prerequisites in a Finnish Polytechnic Institution of Higher Education. *Journal of Workplace Learning, 21*(1), 36–57.

Park, J. S., & Kim, H. S. (2010). Impacts of individual innovativeness on the acceptance of IT-based innovations in health care fields. *Healthcare Informatics Research, 16*(4), 290–298. https://doi.org/10.4258/hir.2010.16.4.290.

Pinheiro, J., & Bates, D. (2000). *Mixed effects models in S and S-Plus*. Springer.

Puza, B. (2015). *Bayesian methods for statistical analysis*. Australian National University Press.

Pylväs, L., & Nokelainen, P. (2020). Academics' perceptions of intercultural competence and professional development after international mobility. *International Journal of Intercultural Relations, 80*, 336–348.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rogers, E. M. (1962). *Diffusion of innovations*. Press of Glencoe.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Samaniego, F. (2010). *A comparison of the Bayesian and frequentist approaches to estimation*. Springer.

Smid, S. C., & Winter, S. D. (2020). Dangers of the defaults: A tutorial on the impact of default priors when using Bayesian SEM with small samples. *Frontiers in Psychology, 11*(611), 963.

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling, 27*(1), 131–161.

Stan Development Team. 2021. *Stan user guide, 2.18*. Retrieved from https://mc-stan.org

Stegmueller, D. (2013). How many countries for multilevel modeling? a comparison of frequentist and Bayesian approaches. *American Journal of Political Science, 57*(3), 748–761.

The jamovi project. (2021). *jamovi (Version 1.6)* [Computer Software]. Retrieved from https://www.jamovi.org

The JASP Team, Dropmann, D., Selker, R., & Love, J. (2022). *jsq: Bayesian methods. R package version 1.1.0*.

Tierney, W. G., & Lanford, M. (2016). Conceptualizing innovation in higher education. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (pp. 1–40). Springer.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*(3), 842–860.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology. *Psychological Methods, 22*(2), 217–239.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharsky, S., Ly, A., Marsman, M., Matzke, D., Narendra Gupta, A., Komarlu, R., Sarafoglou, A., Stefan, A., . . . Wagenmakers, E. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*(3), 813–826.

van Dyne, L., Ang, S., & Koh, C. (2009). Cultural intelligence: Measurement and scale development. In M. A. Moodian (Ed.), *Contemporary leadership and intercultural competence: Exploring the cross-cultural dynamics within organizations* (pp. 233–254). Sage.

van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review, 25*, 143–154.

Wakefield, J. (2013). *Bayesian and frequentist regression methods* (Springer Series in Statistics). Springer.

Petri Nokelainen (PhD) is a Professor (Engineering Pedagogy) in Tampere University, Finland. His research focuses on teaching and learning in engineering higher education and workplace learning. He is the Chief Editor of Finnish Journal of Vocational and Professional Education, Board Member of VETNET (European Educational Research Association), and Editorial Board Member of several international scientific journals.

Tahani Z. Aldahdouh (PhD) is a postdoctoral researcher at Tampere University, Finland. Her main interests include innovativeness in higher education, professional growth, technology acceptance and quantitative analysis. She is a member of Higher Education in Transition (HET) and Professional Growth and Learning (PGL) research groups in Tampere University.

Alaa A. Aldahdouh (PhD) holds a PhD in educational technology from University of Minho, Portugal. He has a B.S in computer engineering with cumulative experience in software programming. His interests are in studying human and machine learning. His recent works focus on investigating new educational theories and their potential role to enhance students' learning processes.

Chapter 11
A Primer to Latent Profile and Latent Class Analysis

Johannes Bauer

Abstract This chapter gives an applied introduction to latent profile and latent class analysis (LPA/LCA). LPA/LCA are model-based methods for clustering individuals in unobserved groups. Their primary goals are probing whether and, if so, how many latent classes can be identified in the data and estimating their proportional size and response profiles. Moreover, latent class membership can serve as a predictor or outcome for external variables. Substantively, LPA/LCA adopt a person-centred approach that is useful for analysing individual differences in learning prerequisites, processes, or outcomes. This chapter provides a conceptual overview of LPA/LCA, a nuts-and-bolts discussion of the steps and decisions involved in their application, and illustrative examples using available data and the R statistical environment.

Keywords Latent variable models · Mixture models · Latent class analysis · Latent profile analysis

11.1 Introduction

Suppose you are the research adviser of a board of a chamber of commerce and industry responsible for final examinations in an advanced training master's degree programme on business management. One of the exams is a standardised test that provides scores ranging from 0 to 50. The board wishes to understand participants' achievements in recent test applications and asks you to analyse the available data ($N = 1000$). You are surprised to find that the data do not follow a normal distribution, as you would have expected, but a bi-modal one, as illustrated in Fig. 11.1a. Since you wonder whether something is wrong with the test you discuss

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-08518-5_11].

J. Bauer (✉)
University of Erfurt, Erfurt, Germany

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_11

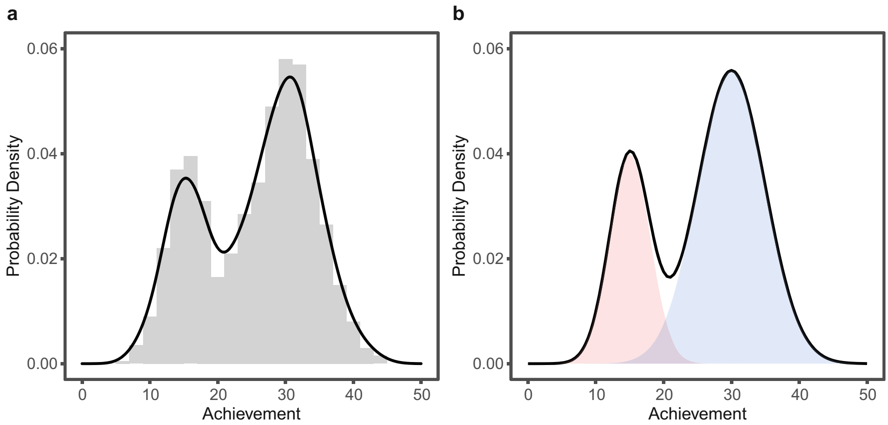

Fig. 11.1 Illustrative univariate finite normal mixture distribution: (**a**) empirically observed distribution ($N = 1000$); (**b**) true population distributions: two within-class distributions [red: $N(\mu = 15; \sigma = 3)$; blue: $N(\mu = 30; \sigma = 5)$] and their joint mixture distribution (black line)

this finding with a colleague. They mention that the distribution looks like the data came from two separate groups with different achievement levels. Reportedly, some participants try passing the exam by rote rehearsal whereas others prepare more thoroughly by applying deeper learning strategies. Given the nature of the test tasks, it is reasonable that rote rehearsal leads to diminished performance. Unfortunately, you do not have data on participants' learning strategies. All you have is the achievement data. Nevertheless, you wonder whether it would be possible to identify the unobserved groups given the data.

Let us put the latter question aside for a moment and suppose we have an insight into the actual population characteristics, as illustrated in Fig. 11.1b. There are two participant subpopulations: the low-achieving (supposedly superficial) learners marked red, and the high achieving (supposedly deep) learners marked blue. Because these subpopulations are unobserved in our empirical data, they are hidden from us and called *latent classes*. Achievement differences follow a normal distribution with class-specific (conditional) means and standard deviations within each latent class. The bi-modal cross-class distribution is a marginal combination of these two within-class distributions and, therefore, called a *mixture distribution*.

So, how can we recover the latent classes shown in Fig. 11.1b based on the observed achievement data? This is where *mixture models* come in. Mixture models are a family of statistical procedures for uncovering hidden groups. As Oberski (2016) put it, "Mixture modeling is the art of unscrambling eggs" (p. 275). This chapter introduces latent profile analysis (LPA) and latent class analysis (LCA), mixture models for cross-sectional data. The primary difference between them is that LPA applies to continuous response variables, whereas LCA applies to categorical ones. As we will see, an LPA for the example data in Fig. 11.1a gives us information on (i) the number of latent classes, (ii) model fit and the accuracy of the classification, (iii) the proportional size of each class, (iv) the within-class means and variances, and (v) probability values for each person's latent class membership. Of course, since we have not observed class membership, the classification is inevitably uncertain. A significant advantage of LPA/LCA over traditional cluster analysis is that this classification uncertainty is an explicit part of the statistical model (Rost, 2003; Vermunt & Magidson, 2002).

The following section provides readers with a conceptual background of LPA/LCA, including a discussion of their potential for research on professional learning and development. Next, I provide didactic examples of how to implement LPA and LCA. The chapter will close with some cautionary considerations, a note on software, and recommendations for further reading.

11.2 Conceptual Overview

11.2.1 Family Relations

As mentioned, LPA/LCA belong to the broad family of mixture models (Hancock & Samuelsen, 2008; Mclachlan & Peel, 2000; Masyn, 2013; Sterba, 2013). The common purpose of all mixture models is to provide a probabilistic classification of persons to latent classes based on a statistical model. Each class is one of the K categories of a discrete latent variable c. Hence, unlike traditional factor analysis or IRT modelling, LPA/LCA models latent variables that are categorical instead of continuous. All mixture models combine a within-class and a between-class model (Sterba, 2013). The within-class model defines the population parameters for persons in a given class k who constitute a relatively homogeneous subgroup regarding these parameters. The between-class model defines each persons' probability of being sampled from the subpopulation underlying class k. Thus, mixture models cluster persons who are similar in terms of the within-class parameters and yield a classification probability for each class per person. Persons then can be assigned to a latent class based on their highest membership probability (*modal class assignment*). This probability-based (soft) clustering is a significant conceptual advantage of mixture models over the deterministic (hard) assignment applied in traditional cluster analysis because it explicitly accounts for the mentioned classification uncertainty (Mair, 2018; Rost, 2003; Vermunt & Magidson, 2002).

Differences among various kinds of mixture models concern the parameters in the focus of the classification (i.e., conditional means/item probabilities vs regression parameters) and whether the data are cross-sectional or longitudinal (Table 11.1). Moreover, the observed items' level of scale can be considered, e.g., to distinguish LPA and LCA. However, the latter distinction is somewhat artificial since both LPA and LCA are just special cases of the general mixture model. It is even possible to combine mixed scale indicators in the same analysis (Vermunt & Magidson, 2002; see Mair, 2018, for an illustrative application).

As shown in Table 11.1, latent classes in LPA/LCA are characterised by qualitatively different response profiles on the analysed items (Rost, 2003; Vermunt &

Table 11.1 Overview of selected mixture models

	Classification by	
Design	*Means (continuous data)/ Item probabilities (categorical data)*	*Regression parameters*
Cross-sectional	Latent profile analysis Latent class analysis	Regression mixture models
Longitudinal	Repeated measures latent class analysis Latent transition analysis	Growth mixture models

Magidson, 2002). If the items are continuous, the profiles are patterns of class-specific means. If they are categorical, the profiles are defined by the items' class-specific endorsement probabilities (i.e., for dichotomous items, the probability of scoring 1 vs 0). This basic idea extends to the longitudinal case. For example, repeated measures LCA models classes of mean/probability-trajectories across multiple measurement occasions; latent transition analysis models latent classes for each time point (*latent statuses*) as well as transitions between them (Collins & Lanza, 2010). Though not elaborated in this chapter, mixture models for regression coefficients are noteworthy because they allow uncovering differential treatment effects in experimental or intervention research (Jung & Wickrama, 2008; Oberski, 2016).

As a final note, mixture models can be used both in an exploratory, hypothesis-generating mode of data analysis and in a confirmatory, hypothesis-testing one (Finch & Bronk, 2011). The former occurs more frequently in practice and applies when a researcher has only tentative or no prior assumptions about the latent classes. However, it should be emphasised that even confirmatory LPA/LCA cannot test hypotheses about a specific number of classes. As elaborated in Sect. 11.2.3, this number is not an estimable model parameter. Testable hypotheses concern only the class sizes and their response patterns on the items.

11.2.2 A Person-Centred Conceptual Perspective

Before introducing the statistical details of LPA/LCA, let's consider what added value mixture models provide from a substantive point of view for studying professional learning and development. Mixture models have unique advantages in studying *individual differences* in learning (Bauer et al., 2018; Collins & Lanza, 2010; Hancock & Samuelsen, 2008; Masyn, 2013). Such individual differences may relate to the premises (e.g., learner or contextual factors), processes (e.g., learning activities, social interactions), and outcomes (e.g., knowledge, performance, practices) of professional and work-related learning (cf. Gruber & Harteis, 2018; Tynjälä, 2013). Conceptually, mixture models adopt a person-centred approach that is helpful when a researcher aims to analyse categorical individual differences in classifying persons according to typical patterns of their characteristics or behaviour (Bergman & Magnusson, 1997). This person-centred perspective is complementary to the traditional variable-centred one that aims at identifying relationships between variables. For example, taking a variable-centred view, researchers might be interested in the degree to which some predictors are relevant for explaining employees' engagement in work-related learning activities. From a person-centred perspective, they may ask what typical patterns of engagement in learning activities occur, how prevalent these patterns are, and what background variables characterise people demonstrating a specific pattern. Hence the two perspectives are complementary and should not be dichotomised falsely (Masyn, 2013). Synergies can arise from taking both perspectives on the same research topic (Marsh et al., 2009). For example, Bauer and Mulder (2013) studied predictors of and individual differences in nurses' engagement in learning activities after errors at work. In a variable-centred analysis, they found that nurses' perceptions of errors and their social context predicted this engagement. A subsequent LPA complemented these findings by identifying four typical reactions to errors and learning engagement patterns, with a substantial proportion of nurses (>40%) showing dysfunctional behaviour.

More generally speaking, LPA/LCA can be used whenever there are theoretical reasons to believe that a latent construct is categorical. As Collins and Lanza (2010) argue, many variables of substantive interest can be understood both in terms of continua and categories. They employ the example of alcohol consumption which may be defined continuously as the volume of alcohol consumed in a given period. At the same time, it makes substantive sense to distinguish categories of alcohol consumption, such as non-drinking, occasional drinking, social drinking, binge drinking, or being an alcohol addict. Applied to the professional learning and development context, ability in a given professional task may be defined quantitatively (e.g., the ratio of correctly diagnosed x-ray images). Nevertheless, it may be substantively meaningful to categorise groups based on their differential achievement (e.g., novices, intermediates, and experts). Similarly, referring to our introductory example, test achievement is arguably a quantitative variable, but it can be meaningful to distinguish categories of learners showing mastery vs non-mastery. Extending these examples to a multivariate scenario, ability in a given task may be defined by accuracy and response time, both of which are quantitative. However, categories of different combinations of speed and accuracy (e.g., fast/inaccurate, slow/moderately accurate) may represent relevant qualitative differences in the patterns of the professional learning process (Hickendorff et al., 2018). All in all, readers will find that LPA/LCA (and other mixture models) apply to a broad spectrum of variables relevant to professional and work-related learning.

11.2.3 The LPA and LCA Models

The discussion so far clarified the general purpose of mixture models, with a particular eye on LPA/LCA. Readers should also understand why and for what research questions these techniques are useful. The section below will elaborate on the statistics of LPA/LCA before turning to the steps and decisions involved in their application.

Latent Profile Analysis In LPA, we typically assume a normal distribution within each latent class, though other distributions are possible, such as Poisson for count data (Mair, 2018; Vermunt & Magidson, 2002). The normality assumption applies only to these subpopulation distributions, not their joint distribution, and it is untestable. Each latent class (1, …, K) has a density $f_k(x_i|\mu_k, \sigma_k^2)$, where x_i is an observed value for person i on a continuous indicator m, and μ_k and σ_k^2 are its mean and variance in class k. The joint mixture model is given by

$$f(x_i) = \sum\nolimits_{k=1}^{K} \pi_k f_k(x_i|\mu_k, \sigma_k^2) \tag{11.1}$$

where π_k is the relative class size, or equivalently, the *class probability* (also called *mixture weight*). That is, each person's density in the mixture distribution is a sum of the class-specific normal densities weighted by the class probabilities. For example, consider persons with a test-score of 21 in Fig. 11.1b. Their mixture density (i.e., the black curve) is the sum of the two class-specific densities (i.e., the red and blue curves, respectively) weighted by the respective class probabilities. In the multivariate case, where $\boldsymbol{x_i}$ is a vector of responses on a set of M observed items, the model remains essentially the same but each class has its specific mean vector $\boldsymbol{\mu_k}$ and variance-covariance matrix $\boldsymbol{\Sigma_k}$.

Of course, before the analysis, we neither know the class probabilities nor the class-specific means and (co)variances. Unless further restrictions are imposed, these are the model parameters to be estimated. Because the number of free parameters proliferates with the number of latent classes and indicators, it makes sense to specify further restrictions to increase model parsimony and estimation stability (Vermunt & Magidson, 2002). Frequently, one imposes a local independence restriction meaning that the indicators are constrained to be uncorrelated *within* each latent class.[1] Additionally, a *cross-class* equality restriction on the indicator variances can be imposed delivering homogeneity of the variance-covariance matrices across classes (Vermunt & Magidson, 2002). Together these two restrictions ensure that the classes differ only in their mean profiles, and not in the form of the variance-covariance matrix, which facilitates interpretation (Peugh & Fan, 2013). Several other models with different restrictions are available (Masyn, 2013; Pastor et al., 2007; Scrucca et al., 2016). Finally, an essential point in Eq. 11.1 is that the number of classes is *not* an estimated model parameter. Instead, the number of classes must be decided by the analyst on the grounds of statistical and substantive criteria (see Sect. 11.2.4). Once a target model has been selected, its estimated parameters can be interpreted.

[1] That is, residual correlations among the indicators are zero given class membership. Local independence is a default assumption in many latent variable models, such as factor analysis or IRT models, but can be relaxed.

Latent Class Analysis The LCA model works similarly though it applies to categorical items. LCA starts with the basic idea of local independence; that is, it seeks to establish latent classes such that responses on a set of categorical items are independent within classes (Rost, 2003). Suppose we have a vector of responses $\boldsymbol{u_i}$ to a set of M categorical items. Under the local independence assumption, the LCA model provides the probability for a given response pattern on these items by

$$P(u_i) = \sum\nolimits_{k=1}^{K} \pi_k \prod\nolimits_{m=1}^{M} P(u_{mi}|c_i = k). \tag{11.2}$$

The probability of a specific combination of item responses depends on the class probabilities and the product of the item response probabilities given class membership. These are the model parameters to be estimated. For items with J categories, the item probabilities are the J-1 thresholds between these categories. Again, further restrictions may be imposed on the model. Note that the local independence assumption can be evaluated and partially relaxed in LCA (Asparouhov & Muthén, 2015; Masyn, 2013).

11.2.4 Steps in LPA/LCA

Conducting an LPA/LCA is an iterative, stepwise process similar to exploratory factor analysis. Four steps can be distinguished, the last of which is optional: (i) specifying the model, (ii) determining the number of latent classes, (iii) interpreting the target solution substantively, and (iv) including predictors or outcomes of latent class membership. In the following, I briefly elaborate on each of these steps.

Step 1: Model specification. In specifying the model, researchers need to make a theory-based selection of observed indicators, decide how these indicators enter the model, and what restrictions to impose on the model parameters. Regarding selecting indicators, one should consider that LPA/LCA are measurement models (Wang & Wang, 2020). So, there should be a rationale that the indicators measure a joint theory-based construct (e.g., different self-concept profiles). In practice, LPA/LCA is also frequently applied to classify persons over a range of conceptually related but distinct constructs (e.g., Bauer & Mulder, 2013; Gillet et al., 2020). In that case, the resulting profiles are mere summaries of persons with similar answer patterns and should not be interpreted as realisations of a latent trait (see Sect. 11.4.1 on the danger of reification). In either case, it makes only sense to apply LPA/LCA if the set of indicators is inter-correlated. Therefore, screening indicator associations before the analysis is warranted.

A further consideration in LPA is whether to include indicators on the item level or in aggregated form and apply prior transformations. Aggregation is frequently done by computing composite mean scores, factor scores, or (preferably) plausible values of the (sub-)scales to be analysed (e.g., Bauer & Prenzel, 2021; Hong et al., 2020; Kunst et al., 2018; Mair, 2018).[2] This approach considerably reduces the number of parameters to be estimated and better approximates continuous indicators. An arguably optimal strategy in this regard is second-order LPA, in which the latent class variable is a second-order variable with continuous first-order factors serving as indicators (for an application, see Bauer et al., 2018). With this specification, the indicators used for the latent classification are latent factors that are genuinely continuous and corrected for measurement error. Note that this approach requires scaling the first-order factors to have a meaningful mean structure (Little et al., 2006).

Concerning parameter restrictions, researchers may wish to start with the local independence model since this is the traditional basis of LCA and potentially consider cross-class equality constraints on the indicator variances. Any other restrictions should be made on substantive grounds (see Rost, 2006, for interesting applications). Attention should be paid to the default restrictions implemented by specific software packages. For example, Mplus (Muthén & Muthén, 1998–2017) by default imposes local independence and homogeneity across classes. An alternative strategy implemented routinely in the **mclust** package (Scrucca et al., 2016) is estimating a variety of models with different restrictions and choosing the best-fitting one (including a decision on the number of latent classes; see Step 2). This data-driven strategy alleviates researchers from a priori considerations about restrictions. However, interpreting what the restrictions applied by the selected model mean substantively may not always be straightforward. For further details, see Sect. 11.3.2.

Step 2: Class Enumeration. Deciding on the number of classes may be the most challenging step in the whole analysis. The goal is to yield a solution that balances model parsimony and fit and delivers substantively interpretable classes. For this purpose, one estimates[3] a series of models from 1 to K latent classes[4] and compares them on indices of *relative model fit* and *classification diagnostics*. Next, the candidate model(s) are considered substantively.

The most commonly used fit indices are the *Bayesian Information Criterion* (BIC), its sample-size-adjusted variant (SABIC), and the *Akaike Information Criterion* (AIC). With these and any other information criteria, lower values indicate a better model fit. Ideally, a minimum value in the series of candidate models points at the best-fitting solution. In practice, fit indices may continue to improve as more latent classes are extracted. In such circumstances, an elbow plot – similar to the

[2] Using composite scores of categorical items as indicators will turn an LCA into an LPA.

[3] Discussing details of estimation is beyond the scope of this chapter; for a tractable introduction, see Masyn (2013).

[4] The maximum number of probed classes frequently hinges on practical issues, such the occurrence of convergence problems or other issues as more classes are extracted (e.g., occurrence of small class sizes; see step 3). If possible, researchers should consider what maximum number of classes may be of theoretical interest.

scree-plot in exploratory factor analysis – can be helpful to identify at which step the improvement in model fit plateaus. Next to information criteria, models with adjacent latent classes can be compared by a *likelihood ratio test* (LRT) available in several versions. Some simulation studies indicate the *bootstrapped LRT* (BLRT) to be quite effective in the class enumeration (Nylund et al., 2007). A statistically significant result of the LRT supports the *k*-class model over *k*-1 classes. In theory, one chooses the model after which the LRT is statistically non-significant for the first time. Again, in practice, the LRT may continue to stay significant for the whole series of probed models, especially with large sample sizes.

Classification diagnostics can additionally inform the class enumeration process. For each latent class, *average posterior class probabilities (AvePP$_k$)* provide a measure of classification accuracy. As the name suggests, they are based on averaging for class k the model-estimated (i.e., posterior) latent class membership probabilities for individuals with their highest membership probability in this class. Values $\geq .70$ have been proposed as desirable (Masyn, 2013). The information about classification accuracy is condensed in the *entropy* measure that ranges between 0 and 1 and can be interpreted similarly to a reliability coefficient for the classification (values $\geq$.80 are desirable, $\geq$.60 a suggested minimum; Asparouhov & Muthén, 2014). Classification accuracy may be helpful to decide between solutions that have a similar fit (Rost, 2006), but note that entropy is not a good sole or primary model selection criterion (Masyn, 2013).

As a final word of caution, though there is extensive research on fit indices in mixture models (e.g., Nylund et al., 2007; Peugh & Fan, 2013; Tofighi & Enders, 2008), results have not yet converged, and there is no single best index. Applied researchers frequently combine an array of indices for class enumeration but – probably more often than not – their results may prove to be inconsistent. Therefore, there is general agreement in the methodological literature that model selection needs to be guided at least as much by the substantive interpretability of the class solution as by fit indices. If multiple solutions with similar fit occur, one should inspect all of them and decide on substantive grounds.

Step 3. Substantive interpretation of the target model(s). The interpretation process in LPA/LCA is similar to traditional cluster analysis or exploratory factor analysis. For this purpose, one inspects the class-specific mean/probability profiles across the indicators and differences in these profiles across classes. Depending on model specification, class-specific (co)variances are also significant. Based on this interpretation, a definition and label must be attached to each class. Finally, the class sizes are of interest to distinguish common or normative profiles from more exceptional ones. In general, no solution should be retained that is not well interpretable, regardless of model fit. Some authors consider a solution useless in which not all classes are interpretable (Wang & Wang, 2020). However, Mair (2018) notes that the occurrence of a single uninterpretable junk class is not problematic since it may simply contain the persons that did not fit in any other class.

Though interpretation is primarily a theoretical issue, some methodological considerations apply (Collins & Lanza, 2010; Masyn, 2013; Rost, 2006): First, the class profiles should be distinct enough. There should be low variability within classes and a high one between them (Collins & Lanza, 2010). In this context, it may be helpful to compare solutions for adjacent classes to determine from which classes in a prior solution the members of a new class are derived (Rost, 2006). Sometimes a class in a previous solution is simply split up into two classes, and the researcher must consider the value added by this differentiation. Moreover, finding that the structure of profiles varies significantly across different solutions may indicate that LPA/LCA is inappropriate for the data (Rost, 2006).

Second, it should be determined whether the indicators can be arranged so that the class profiles are ordered (i.e., do not cross each other in a profile plot). Such a solution – particularly if the profiles are more or less parallel – would indicate that the classes differ by degree rather than type (Rost, 2006; Wang & Wang, 2020). In that case, the classification is still meaningful. Still, a model focusing on quantitative differences in the latent variable (i.e., a factor or IRT model) would provide a more parsimonious account of the data.

Finally, solutions with several small classes (say $\leq 5\%$) may indicate that too many classes have been extracted (Nylund et al., 2007). Though such small classes can be of substantive value, they should be critically examined to determine if they have a distinct and interpretable profile.

Step 4: Include predictors and distal outcomes of most likely latent class membership. Having established a latent class solution, a logical next step is to analyse predictors or dependent variables of latent class membership (Vermunt & Magidson, 2002). This is called LPA/LCA with *covariates* and *distal outcomes*. Such analyses are useful to (i) enrich the description of the latent classes based on background information (e.g., social background variables or experimental condition), (ii) analyse the construct-validity of the latent class solution, or (iii) test hypotheses for answering substantive research questions. The latter two cases lead to a blend of exploratory and confirmatory analyses (Masyn, 2013) because the hypotheses can only be stated after the classification. However, even if the classification is intrinsically exploratory, the later hypotheses about the classes' relation to covariates and distal outcomes can still be tested in a confirmatory mode.

Methods for analysing covariates and distal outcomes in mixture models are still an active area of research. It may be tempting to simply extract individuals' modal class assignment from the results and use this as a manifest variable in further analyses such as an analysis of variance (ANOVA). However, it is well known that this procedure leads to biased results because it disregards the classification uncertainty and, thus, is inappropriate (Bakk & Kuha, 2020). A variety of alternative approaches has been developed. Recent overviews recommend using the *two-step method* or one of the *bias-corrected three-step methods* (BCH and ML) for handling covariates; (manual) BCH is advisable for continuous and categorical distal outcomes (Asparouhov & Muthén, 2021; Bakk & Kuha, 2021; Nylund-Gibson et al., 2019). Unfortunately, these approaches are currently only available in commercial software (Latent Gold, Mplus). Illustrating the application of LPA/LCA with covariates and distal outcomes is beyond the scope of this chapter. Readers will find didactic examples in the literature cited above and Wang and Wang (2020).

11.3 Implementation of LPA and LCA

This section provides three didactic examples of LPA and LCA. We will use simulated data and the packages **mclust** (Scrucca et al., 2016), **tidyLPA** (Rosenberg et al., 2018), and **poLCA** (Linzer & Lewis, 2011) in the R statistical environment (R Core Team, 2021) so that readers can reproduce all analyses. Though the data are simulated, Examples 2 and 3 are based on published studies that used LPA (Gillet et al., 2020) and LCA (Richter et al., 2013). The presentation below is restricted to the most critical parts of the in- and output. Complete code and data files are available in the supplemental online material at https://osf.io/x3ahn.

11.3.1 Example 1: Univariate LPA of Student Achievement

Turning back to the initial example from Fig. 11.1, we will use LPA to recover the two hidden groups in the mixture distribution of achievement in the exam. Since there is only one variable, our only choice in the model specification is whether we allow the class-specific variances to vary or not. Since our empirical data suggest a different spread in the presumed clusters (Fig. 11.1a), it seems reasonable to freely estimate the variances within each class. For class enumeration, we estimate models with $k = 1$ to 4 latent classes. Based on Fig. 11.1, we expect a solution with two latent classes to fit the data best. However, it is essential to compare this model with other solutions. A maximum of four classes were chosen to assess because solutions with more classes would be quite inconsistent with our assumption of superficial vs deep learning strategies shaping the data distribution. That being said, this choice of maximum classes to be probed is somewhat arbitrary, as is frequently the case.

The R code for estimating the models is pasted below. The argument `modelName = "V"` specifies varying variances across classes. If we omit this argument, **mclust** will estimate models with and without an equality constraint, and users can compare their fit to the data. The subsequent commands call for an elbow plot of BIC (see online supplement) and the BLRT.

```
library(mclust)
m1 <- Mclust(data = mix, modelName = "V", G = 1:4)
plot(m1, what = "BIC")
LRT <- mclustBootstrapLRT(mix, modelName = "V", maxG = 3)
```

Unfortunately, **mclust** delivers only a few fit indices. If we desire a wider variety of indices, we can compute them manually based on the information stored in the output object `m1`. Otherwise, we may get enhanced output by using the functions `estimate_profiles()` and `get_fit()` from the **tidyLPA** package, which provides a set of wrapper functions for **mclust**. Both options are demonstrated in the online supplement. Table 11.2 presents a selection of fit indices, all of which indicate that the two-class model fits the data best. This is supported by the BLRT showing

Table 11.2 Selected fit indices for the univariate LPA of student achievement

Classes	LL	AIC	BIC	CAIC	SABIC	BLRT	*p*
1	−3519.63	−7043.26	−7053.08	−7055.08	−7046.73	–	–
2	−3399.82	**−6809.64**	**−6834.18**	**−6839.18**	**−6818.30**	239.63	.001
3	−3400.23	−6816.46	−6855.72	−6863.72	−6830.31	−0.82	**.946**
4	−3398.41	−6872.80	−6837.86	−6883.80	−6818.82	3.65	.135

Note: *LL* = Log-Likelihood; note that in **mclust**, higher values on information criteria indicate a better fit; *boldface* = highest value/first non-significant BLRT

that the two-class model fits the data significantly better than the one-class model, whereas moving to three classes does not provide a further statistically significant improvement of model fit.

For classification diagnostics, we can use the uncertainty values and modal class assignments stored in the **mclust** output object to calculate average posterior class probabilities which are $AvePP_1 = .95$ for Class 1 and $AvePP_2 = .98$ for Class 2 (see online supplement). Entropy values can be obtained by `get_fit()` and show that the two-class model has high classification accuracy ($E = .89$). Thus, having identified a best-fitting model, we can inspect its estimated parameters in the output inserted below.

```
summary(m1, parameters = TRUE)
## -------------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## -------------------------------------------------------
##
## Mclust V (univariate, unequal variance) model with 2 components:
##
##   log-likelihood    n df       BIC       ICL
##        -3399.819 1000  5 -6834.176 -6903.184
##
## Clustering table:
##   1   2
## 309 691
##
## Mixing probabilities:
##         1         2
## 0.3090278 0.6909722
##
## Means:
##        1        2
## 15.23148 30.14721
##
## Variances:
##         1         2
##  9.633501 23.584189
```

Table 11.3 True class versus most likely class membership from LPA

	LPA		
True	1	2	Total
A	292	8	300
B	17	683	700
Total	309	691	1000

The first part of the output reiterates information about the model fit. The clustering table tells us that $n = 309$ persons had their modal class membership in Class 1 and $n = 691$ in Class 2. The estimated class sizes in the population (*mixing probabilities*) are 30.9% for Class 1 and 69.1% for Class 2. The most important part of the output is the sections on class-specific means and variances. Class 1 has an average achievement of $M = 15.2$ ($SD = 3.1$), and Class 2 $M = 30.1$ ($SD = 4.9$). Given these values, the two classes consist of *low-achievers* and *high-achievers*, respectively.

Our LPA effectively recovered the class-specific population parameters (cf. Fig. 11.1b). Of course, LPA cannot deliver evidence that there are two latent classes *because* some participants applied different learning strategies than others. To investigate this, we might conduct a new study collecting additional data on participants' learning strategies and checking whether they predict latent class membership. We could also collect additional data on some external criterion (say later occupational success) and use the latent class variable as a predictor.

Finally, with our simulated data, it is possible to compare class assignments based on the LPA results with individuals' actual class membership. As Table 11.3 shows, there is a correct classification rate of 97.5%. Therefore, the LPA was quite effective in recovering the unobserved groups.

This example of a univariate LPA demonstrated general procedures that transfer to the multivariate case. Though univariate LPA may rarely be of substantive interest in applied research, it can be a useful methodological tool. For example, in a recent online study, we used it to identify a cluster of participants with an unreasonably short response time that we discarded as junk participations (Reiser et al., in prep.). LPA arguably provides a more principled approach for this purpose than setting an arbitrary cut off time. Moreover, univariate LPA may be an alternative to standard grouping procedures, such as the median split (Rost, 2006). For this purpose, equality restrictions may be imposed on the class sizes to yield equally sized groups.

11.3.2 Example 2: Multivariate LPA of Work Recovery

In most applied cases, researchers prefer to use LPA on multiple indicators. The illustrative example presented in this section is based upon Study 1 in Gillet et al. (2020), who investigated profiles of work recovery (and related predictors) in a sample of $N = 415$ employees from various occupations. Factor scores on three psychological mechanisms of work recovery served as indicators: (i) *psychological* *detachment*, the ability to mentally step back from work after hours; (ii) *rumination*, a tendency to repetitively or continuously be preoccupied with distressful aspects after work; and (iii) *overcommitment*, a generic tendency to excessively strive for achievement and recognition at work. Psychological detachment is a functional mechanism for the work recovery process; rumination and overcommitment are dysfunctional. Population data for the simulation ($N = 500$) were chosen consistent with results from Gillet et al. (2020). Hence, within sampling error, our analysis will lead to roughly identical conclusions.

The code below instructs **mclust** to specify and fit a series of models with $k = 1$ to 8 latent classes.[5] As mentioned above, if we specify no particular model restrictions, the software estimates a set of 14 models that impose specific model constraints based on geometric features of Σ_k (i.e., volume, shape, and orientation; Scrucca et al., 2016). The substantive implications of these features may be hard to interpret for applied researchers, though. Important models, also discussed in other literature on LPA (Masyn, 2013; Pastor et al., 2007; Rosenberg et al., 2018), are in order of increasing model complexity:

- EEI: variances equal across classes, covariances fixed at zero
- VVI: class-varying variances, covariances fixed at zero
- EEE: variances and covariances equal across classes
- VVV: class-varying variances and covariances

EEI and VVI are local independence models, and EEE and VVV relax this assumption.

```
m1 <- Mclust(wr, G = 1:8)
summary(m1)
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust VVI (diagonal, varying volume and shape) model with
4 components:
##
##   log-likelihood   n df       BIC       ICL
##         -1450.855 500 27 -3069.505 -3104.645
##
## Clustering table:
##   1   2   3   4
## 58 159 225  58
```

The summary output suggests that the VVI model with four latent classes fits the data best. To inspect the fit more closely, we can call for an elbow plot comparing all **mclust** models (see online supplement).

[5]In this and the following example, the maximum number of latent classes to be extracted was chosen for practical reasons (i.e., estimation time when readers reproduce the analyses).

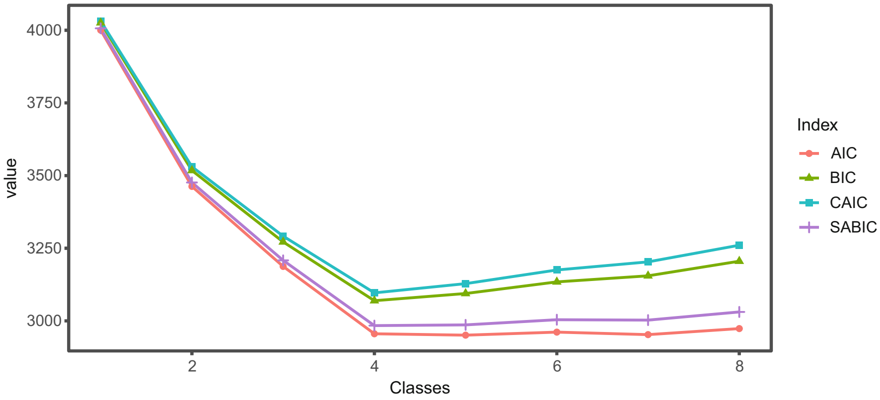

Fig. 11.2 Elbow plot of selected fit indices for the multivariate LPA

Of course, as an alternative to this data-driven strategy, we could have specified a priori parameter restrictions on Σ_k based on substantive considerations. For example, we could have specified local independence and equal cross-class variances by setting the argument `modelName = "EEI"`. To test the tenability of these restrictions, next, we could compare the fit of the EEI model to a less restrictive one such as VVI.

As demonstrated in the univariate example, it may be desirable to base the class enumeration decision on a broader set of indices. Again, this is achieved conveniently using **tidyLPA**. The function `compare_solutions()` from this package delivers a combined recommendation based on several fit indices (see online supplement). Moreover, it may be helpful to plot selected indices (Fig. 11.2). As visible from Fig. 11.2, the four-class solution fits best according to BIC and CAIC. AIC and SABIC seem less conclusive, but parsimony would speak for the four-class solution, too, unless a model with more classes would have advantages in terms of substantive interpretability. The BLRT inserted in the output below adds further support to the four-class solution by getting non-significant for comparing five versus four classes.

```
set.seed(123456)
LRT <- mclustBootstrapLRT(wr, modelName = "VVI", maxG = 5)
LRT
## ---------------------------------------------------------------
## Bootstrap sequential LRT for the number of mixture components
## ---------------------------------------------------------------
## Model        = VVI
## Replications = 999
##              LRTS bootstrap p-value
## 1 vs 2  551.686999         0.001
## 2 vs 3  289.270784         0.001
## 3 vs 4  245.745002         0.001
## 4 vs 5   18.770582         0.051
## 5 vs 6    3.532327         0.937
```

Next, we inspect classification diagnostics. The listing below delivers *AvePP* values, all of which are satisfactory. Entropy is high, too ($E = .93$; see online supplement). **mclust** offers additional diagnostic plots based on dimensionality reduction that inspect the boundaries between the classes to see whether they are well separated (see online supplement).

```
round(aggregate(x = 1 - m1$uncertainty,
  by = list(m1$classification),
  FUN = "mean"), 2)
##   Group.1    x
## 1       1 0.96
## 2       2 0.96
## 3       3 0.97
## 4       4 0.98
```

Having decided on the number of classes, we can refit the target model and inspect the results. Below is a shortened output.

```
m1.4 <- Mclust(wr, modelName = "VVI", G = 4)
summary(m1.4, parameters = TRUE)
(...)
## Clustering table:
##  1   2   3   4
## 58 159 225  58
##
## Mixing probabilities:
##         1         2         3         4
## 0.1138546 0.3152577 0.4546402 0.1162474
##
## Means:
##              [,1]       [,2]       [,3]      [,4]
## overcom -1.540950  0.5203160 -0.4246870  1.636629
## pdetach  1.417878 -0.4843335  0.3818300 -1.548249
## rumin   -1.264305  0.4850927 -0.3714648  1.185387
##
## Variances:
## [,,1]
##            overcom     pdetach      rumin
## overcom  0.1200567  0.00000000  0.0000000
## pdetach  0.0000000  0.05935828  0.0000000
## rumin    0.0000000  0.00000000  0.4626966
## [,,2]
##            overcom    pdetach      rumin
## overcom  0.1068036  0.0000000  0.0000000
## pdetach  0.0000000  0.1490478  0.0000000
## rumin    0.0000000  0.0000000  0.1430751
(...)
```

The mixing probabilities indicate that class sizes vary between 11% and 45% of the sample. The class-specific mean profiles can be found under "Means". A plot facilitates their interpretation (Fig. 11.3). As concluded in Gillet et al. (2020), the

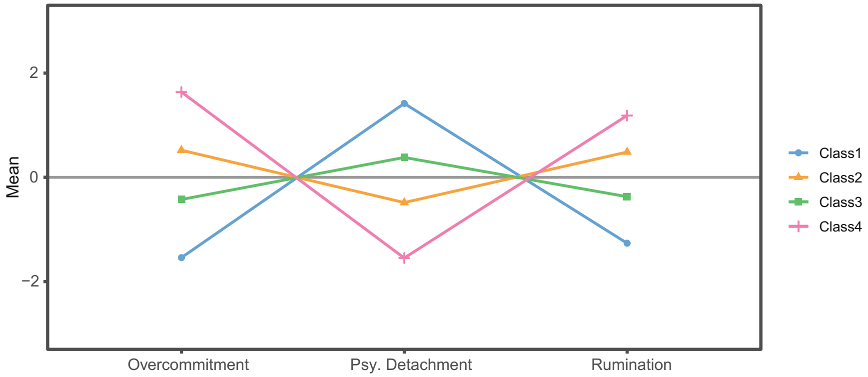

Fig. 11.3 Mean profiles of four latent classes of work recovery. (Simulated data based on Gillet et al., 2020)

classes can be interpreted to demonstrate *high* (Class 1, 11%), *moderately high* (Class 3, 45%), *moderately low* (Class 2, 32%), and *low* (Class 4, 11%) ability to achieve work recovery. Note that this implies that the identified latent classes differ primarily in terms of degree rather than by type. Indeed, the profiles would be perfectly ordered had we recoded psychological detachment so that higher values for all three indicators meant lower ability to achieve work recovery. As discussed above, such an ordered classification is meaningful by pointing at clusters of persons who show similar levels of the investigated trait(s). Nevertheless, these quantitative differences could be modelled more parsimoniously using a continuous latent variable model (Rost, 2006; Wang & Wang, 2020). Some might question whether LPA is the best modelling choice in the present case.

The last part of the output shows the class-specific variance-covariance matrices. As implied by the VVI model, the class-specific covariances are fixed at zero (i.e., local independence), and the indicator variances are allowed to vary. For example, the variance of rumination is higher in Class 1 than in Class 2. Such information may be necessary for understanding the nature of the latent classes.

Finally, a comparison with individuals' actual population class membership shows that about 97% of our sample have been classified correctly by the LPA (see online supplement).

11.3.3 Example 3: LCA of Teachers' Participation in Professional Development

This section gives an applied example of LCA for categorical variables using the **poLCA** package. It draws upon a study by Richter et al. (2013), who investigated individual differences in language teachers' ($N = 2076$) participation in professional development (PD) programmes. In their LCA, the authors analysed eight dichotomised items indicating whether or not (yes/no) teachers had participated in PD activities on each of the following topics during the last 5 years: (i) content knowledge, (ii) pedagogical content knowledge, (iii) teaching and learning methods, (iv) teaching heterogeneous students, (v) classroom management and discipline, (vi) individualised learning support, (vii) counselling students and parents, and (viii) knowledge about kids and youth. As in Example 2, we will use simulated data ($N = 2076$) with population parameters based on the outcomes of the original study.

For model specification, **poLCA** requires users to provide a model formula. The first part of the code below specifies an LCA model with no covariates and stores it in an R object. Next, we estimate a series of models with $k = 1$ to 7 latent classes. Unlike **mclust**, each model must be estimated in a separate function call. The argument `nrep` specifies the number of times to estimate each model using different sets of random starting values for the estimation algorithm. This is important to avoid converging solutions at local instead of global maxima of the likelihood function. Local maxima are a plague in mixture modelling (Masyn, 2013). We did not need to address this problem in the previous LPA examples because **mclust** employs a different approach to avoid local maxima (Scrucca & Raferty, 2015). In the present LCA, we estimate each model with at least ten sets of random starting values. Larger numbers are typically required for complex models, especially in serious applications. Here, we increase the number of random starts for the models with 6 and 7 latent classes and increase the maximum number of iterations (default = 1000) to attain convergence of the estimation algorithm.

```
library(poLCA)
f <- cbind(CK, PCK, TLM, THS, CMD, ILS, CSP, KKY) ~ 1
set.seed(934857)
m1 <- poLCA(f, pd[ ,-9], nclass = 1, nrep = 10)
m2 <- poLCA(f, pd[ ,-9], nclass = 2, nrep = 10)
m3 <- poLCA(f, pd[ ,-9], nclass = 3, nrep = 10)
m4 <- poLCA(f, pd[ ,-9], nclass = 4, nrep = 10)
m5 <- poLCA(f, pd[ ,-9], nclass = 5, nrep = 10)
m6 <- poLCA(f, pd[ ,-9], nclass = 6, nrep = 50, maxiter = 5000)
m7 <- poLCA(f, pd[ ,-9], nclass = 7, nrep = 50, maxiter = 5000)
```

To assess model fit, **poLCA** delivers the information criteria AIC and BIC. As in the previous examples, other fit indices can be calculated manually based on the model results (see online supplement). The elbow plot in Fig. 11.4 shows that BIC, CAIC, and SABIC point to a best-fitting solution with five latent classes, whereas AIC continues to improve as more classes are extracted. Unfortunately, the BLRT is not available in **poLCA**. Therefore, based on the current information, we may preliminarily decide on the five-class model and inspect classification diagnostics.

The listing below first calculates *AvePP,* all of which are >.70. Regarding entropy, note that **poLCA** does not estimate *relative* entropy as is typically used in other software. The function inserted below provides this index (cf. Collins & Lanza, 2010). As we can obtain from the output, entropy for the five-class model is

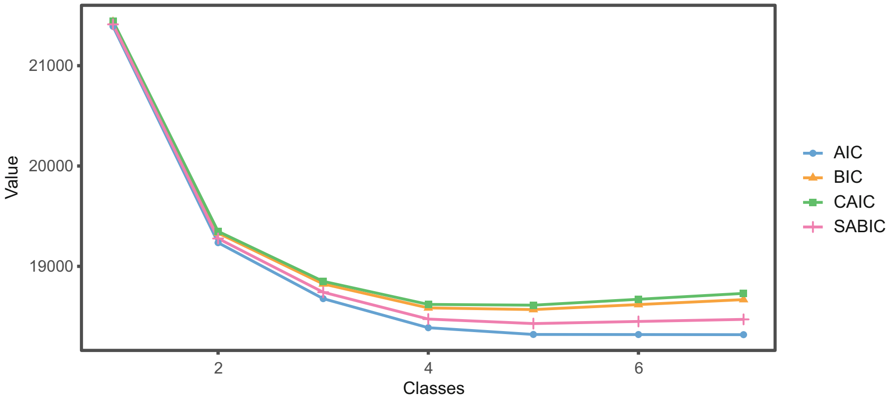

Fig. 11.4 Elbow plot of selected fit indices for the LCA

somewhat below the recommended value of .80 but acceptable (i.e., >.60). From the series of values, we can see that initially, entropy decreases slightly as more classes are extracted. As Collins and Lanza (2010) note, this is frequently the case and an artefact of chance. Recall from the discussion above that entropy is not a good indicator for class enumeration.

```
round(aggregate(x = m5$posterior, by = list(m5$predclass), FUN =
"mean"), 2)
##   Group.1   V1   V2   V3   V4   V5
## 1       1 0.75 0.06 0.09 0.09 0.01
## 2       2 0.01 0.88 0.02 0.03 0.06
## 3       3 0.03 0.03 0.89 0.04 0.00
## 4       4 0.06 0.06 0.05 0.78 0.05
## 5       5 0.00 0.13 0.00 0.03 0.84
lca_entropy <- function(x) {
1 - ((sum(-x$posterior*log(x$posterior)))/(nrow(x$posterior)*log
(ncol(x$posterior))))
}
lca_models <- list(m2, m3, m4, m5, m6, m7)
entropy <- sapply(lca_models, FUN = lca_entropy)
round(entropy, 2)
## [1] 0.83 0.78 0.76 0.76 NaN 0.79
```

Having decided on the five-class model, we can inspect the model parameters. Below is a shortened output.

```
m5r
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $CK
```

```
##           Pr(1)  Pr(2)
## class 1: 0.0217 0.9783
## class 2: 0.0355 0.9645
## class 3: 0.6953 0.3047
## class 4: 1.0000 0.0000
## class 5: 0.8745 0.1255
##
(...)
##
## Estimated class population shares
## 0.2908 0.3073 0.143 0.0718 0.1872
##
## Predicted class memberships (by modal posterior prob.)
## 0.3179 0.2813 0.1392 0.0737 0.1879
(...)
```

In the first part of the output, **poLCA** gives class-specific response probabilities per indicator and category. For example, teachers in Class 1 have a 98% probability of having attended PD focusing on content knowledge, whereas teachers in Class 5 only have a 13% probability. Next, the output contains the latent class sizes in two forms: as estimated class population shares and based on modal class assignment. Both types of values are quite close as they should be. For class interpretation, **poLCA** provides a 3D-barplot of the conditional item response probabilities (see online supplement). However, this plot is hard to read, especially with larger numbers of classes, indicators, and answer categories. Since we analysed dichotomous variables, a profile plot of the conditional item probabilities is better suited for interpretation (Fig. 11.5).

Following Richter et al. (2013), the classes can be interpreted as *high on all activities* (Class 1, 29.1%), *high on activities focusing (pedagogical) content knowledge and teaching methods* (Class 2, 30.7%), *high on activities focusing general pedagogical topics* (Class 3, 14.3%), *high on activities focusing teaching methods* (Class 4, 7.2%), and *low on all activities* (Class 5, 18.7%). Note that the class sizes estimated from our simulated data deviate somewhat from Richter et al. (2013) due to sampling error.

Two more points are noteworthy. First, the order of classes in any mixture model is arbitrary and, thus, may change over a series of models, software used etc. That is, Class 1 from one solution may be Class *k* in another. The order of classes can be adjusted by providing respective starting values to the estimation algorithm. In the above output, classes are reordered to match results reported in Richter et al. (2013) (see online supplement). Second, a closer inspection of the output reveals that Class 4 has boundary probabilities on the indicator CK (1 and 0 for *no* and *yes*, respectively). Finding many such boundary estimates may be a warning sign that the model solution is invalid or that too many classes have been extracted (Geiser, 2013). Even a single boundary estimate should be a cause for caution. In general, boundary values may be legitimate estimates, but they can also indicate model misspecification. In the current example, the respective population value for the data simulation was $p = .98$, which already is close to the upper bound of probability. So, within sampling error, the present result is a reasonable estimate.

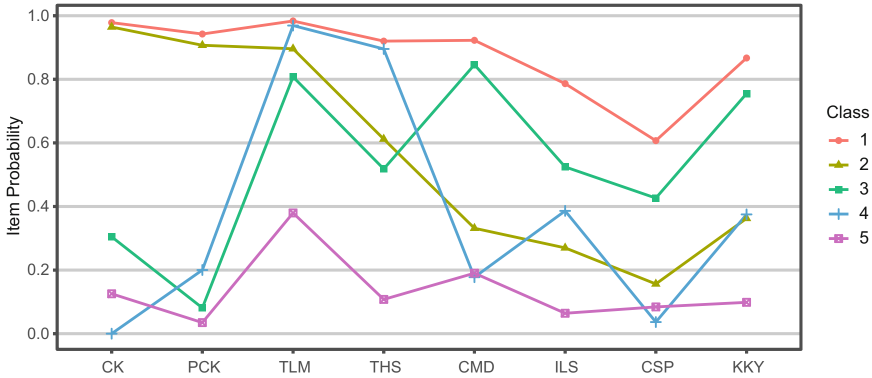

Fig. 11.5 Conditional item probabilities of five latent classes of participation in professional development activities. (Simulated data based on Richter et al., 2013); *CK* = content knowledge, *PCK* = pedagogical content knowledge, *TLM* = teaching and learning methods, *THS* = teaching heterogeneous students, *CMD* = classroom management and discipline, *ILS* = individualised learning support, *CSP* = counselling students and parents, and *KKY* = knowledge about kids and youth

Finally, a check of how well LCA recovered individuals' actual class membership shows only an 85% correct classification rate, with misclassification particularly apparent for Class 4 (see online supplement). This reduced accuracy may result from the classes not being well separated from the outset, as the suboptimal entropy value hints. Hence, this example may serve as a caveat that even in solutions that seem reasonable in terms of fit and interpretation, a non-trivial degree of misclassification can be prevalent.

11.4 Conclusions

This chapter provided an applied introduction to LPA/LCA and their implementation in statistical software using didactic examples based on actual research. As demonstrated, LPA/LCA are effective analytic techniques for identifying unobserved groups and are flexible to use on different kinds of data. Their primary benefit for research on professional learning and development lies in their complementary perspective on studying and understanding individual differences in learning. As the examples have shown, LPA/LCA allow identification of groups characterised by similar patterns of preconditions (e.g., coping with occupational stress), process aspects (e.g., engagement in professional learning activities), and outcomes of learning (e.g., performance). In a similar vein, though not elaborated in this chapter, mixture models for longitudinal data can enhance the understanding of individual differences in professional change and development trajectories. Next to these substantive benefits, readers will have seen that LPA/LCA have clear conceptual and methodological advantages over traditional clustering approaches. Modern software and methodological advances have greatly facilitated these analytic techniques for applied researchers.

11.4.1 Some Words of Caution

Despite the great potential of mixture models, some cautionary words are in order. First, as elaborated in this chapter, some aspects of mixture modelling are inevitably exploratory. This should humble researchers who may draw overly firm conclusions from them. In the same vein, readers will have recognised that mixture modelling is a complex process that requires a multitude of decisions and, thus, provides a high degree of freedom for the researcher. The analytic results can be quite sensitive to these decisions. Therefore, it is crucial to develop a solid and principled plan ahead of the analyses, including criteria for the various decisions to be made. Otherwise, there is a substantial danger of getting lost in the analytic garden of forking paths.

Second, several fallacies in the interpretation of mixture models must be avoided. First, as Loken and Molenar (2008) note, "successfully applying a mixture model does not immediately confirm the appropriateness of the model" (p. 278). Finding a relatively best-fitting model within a series of candidate models neither implies that the model is statistically valid in an absolute sense or that it is a substantively correct representation of the phenomenon in the population. Second, results from a mixture model cannot provide sufficient evidence to confirm that the construct of interest is categorical (Loken & Molenar, 2008; Masyn, 2013). Above, I emphasised that, conceptually, person-centred and variable-centred analyses examine two sides of the same coin. The statistical counterpart of this is that – though factor and mixture models are not wholly equivalent – "a *k*-group mixture (or latent profile) model can be redescribed as factor model with *k*-1 factors that are identical up to first- and second-order moments" (Loken & Molenar, 2008, p. 293). Third, as with other latent variable models, the reification-fallacy should be avoided. Reification occurs when abstract concepts represented by latent variables are taken too literally and interpreted as if they had an actual, physical existence (Agresti, 2019; Bauer, 2007). Successfully identifying latent classes and attaching labels to them does not imply that these classes exist in the real world. In many cases, latent classes should be seen as statistical summaries that cluster persons with similar response patterns and, thus, can be helpful interpretative tools. However, such analytic summaries are insufficient evidence to make the case that a specific typology has a concrete reality or that an individual is of one particular naturally occurring type (Bauer, 2007; Masyn, 2013).

11.4.2 Notes on Software and Further Reading

There is a wealth of alternatives to the software introduced in this chapter. Many R packages relevant for mixture modelling are listed on the Cran Task Views *Psychometric Models and Methods* and *Clusters* (https://cran.r-project.org/web/views). As for proprietary software, Mplus (Muthén & Muthén, 1998–2017) and Latent Gold (Vermunt & Magidson, 2016) are easy to use, can estimate a broad array of mixture models, and implement modern approaches to handling covariates and distal outcomes. An applied introduction to mixture modelling using Mplus is available, for example, in Wang and Wang (2020). For Latent Gold, instructive user guides are freely available on the software's website. These and other software options are reviewed in Haughton et al. (2009), Hickendorff et al. (2018), and Uebersax (2012).

The following selected sources may be of interest for delving more deeply into mixture models. Highly readable introductions can be found in Collins and Lanza (2010), Masyn (2013), and Vermunt and Magidson (2002). Systematic, though more technically challenging, overviews are available in Hancock and Samuelsen (2008), Hancock et al. (2019), McLachlan and Peel (2000), Muthén (2008), and Sterba (2013). The edited volumes by Hagenaars and McCutcheon (2002) and Rost and Langeheine (1997) contain high-quality illustrative applications (mostly of LCA) from a broad range of disciplinary and substantive perspectives.

References

Agresti, A. (2019). *An introduction to categorical data analysis*. Wiley.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 329–341. https://doi.org/10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2015). Residual associations in latent class and latent transition analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(2), 169–177. https://doi.org/10.1080/10705511.2014.935844

Asparouhov, T., & Muthén B. (2021). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary second model* (Mplus Web Notes 21, Version 11). https://www.statmodel.com/MixtureModeling.shtml

Bakk, Z., & Kuha, J. (2021). Relating latent class membership to external variables: An overview. *British Journal of Mathematical and Statistical Psychology, 74*(2), 340–362. https://doi.org/10.1111/bmsp.12227

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*(4), 757–786. https://doi.org/10.1080/00273170701710338

Bauer, J., & Mulder, H. R. (2013). Engagement in learning after errors at work: Enabling conditions and types of engagement. *Journal of Education and Work, 26*(1), 99–119. https://doi.org/10.1080/13639080.2011.573776

Bauer, J., & Prenzel, M. (2021). For what educational goals do preservice teachers feel responsible? On teachers' ethos as professional values. In F. Oser, K. Heinrichs, J. Bauer, & T. Lovat (Eds.), *The international handbook of teacher ethos: Strengthening teachers, supporting learners* (pp. 173–195). Springer. https://doi.org/10.1007/978-3-030-73644-6_12

Bauer, J., Gartmeier, M., Wiesbeck, A. B., Moeller, G. E., Karsten, G., Fischer, M. R., & Prenzel, M. (2018). Differential learning gains in professional conversation training: A latent profile analysis of competence acquisition in teacher-parent and physician-patient communication. *Learning and Individual Differences, 61*(1), 1–10. https://doi.org/10.1016/j.lindif.2017.11.002

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology, 9*(2), 291–319. https://doi.org/10.1017/S095457949700206X

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Wiley.

Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(1), 132–151. https://doi.org/10.1080/10705511.2011.532732

Geiser, C. (2013). *Data analysis with Mplus*. Guilford.

Gillet, N., Morin, A. J. S., Mokounkolo, R., Réveillère, C., & Fouquereau, E. (2020). A person-centered perspective on the factors associated with the work recovery process. *Anxiety, Stress, & Coping*, 1–26. https://doi.org/10.1080/10615806.2020.1866174

Gruber, H., & Harteis, C. (2018). *Individual and social influences on professional learning*. Springer.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge University Press.

Hancock, G. R., & Samuelsen, K. M. (Eds.). (2008). *Advances in latent variable mixture models*. IAP.

Hancock, G. R., Harring, J. R., & Macready, G. B. (Eds.). (2019). *Advances in latent class analysis*. IAP.

Haughton, D., Legrand, P., & Woolford, S. (2009). Review of three latent class cluster analysis packages: Latent Gold, poLCA, and MCLUST. *The American Statistician, 63*(1), 81–91. https://doi.org/10.1198/tast.2009.0016

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences, 66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology, 112*(7), 1409–1430. https://doi.org/10.1037/edu0000445

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Kunst, E. M., van Woerkom, M., & Poell, R. F. (2018). Teachers' goal orientation profiles and participation in professional development activities. *Vocations and Learning, 11*(1), 91–111. https://doi.org/10.1007/s12186-017-9182-y

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42*(10), 1–29. https://doi.org/10.18637/jss.v042.i10

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Loken, E., & Molenaar, P. (2008). Categories or continua? The correspondence between mixture models and factor models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 277–297). IAP.

Mair, P. (2018). *Modern psychometrics with R*. Springer. https://doi.org/10.1007/978-3-319-93177-7

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*(2), 191–225. http://dx.doi.org/10.1080/10705510902751010

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 551–611). Oxford University Press.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Wiley.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). IAP.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(6), 967–985. https://doi.org/10.1080/10705511.2019.1590146

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 275–287). Springer. https://doi.org/10.1007/978-3-319-26633-6_12

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*(1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 616–639. https://doi.org/10.1080/10705511.2013.824780

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Reiser, S., Schacht, L., Thomm, E., Schick, K., Berberat, P. O., Gartmeier, M., & Bauer, J. (in prep.). *On the validity of assessing medical students' communication competence in physician-patient encounters by a video-based situational judgment test.*

Richter, D., Engelbert, M., Weirich, S., & Pant, H. A. (2013). Differentielle Teilnahme an Lehrerfortbildungen und deren Zusammenhang mit professionsbezogenen Merkmalen von Lehrkräften [Differential use of professional development programs and its relationship to professional characteristics of teachers]. *Zeitschrift für Pädagogische Psychologie, 27*(3), 193–207. https://doi.org/10.1024/1010-0652/a000104

Rosenberg, J., et al. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software, 3*(30), 978. https://doi.org/10.21105/joss.00978

Rost, J. (2003). Latent class analysis. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (Vol. 1, pp. 539–543). Sage.

Rost, J. (2006). Latent-class-analyse. In F. Petermann & M. Eid (Eds.), *Handbuch der psychologischen Diagnostik* (pp. 275–287). Hogrefe.

Rost, J., & Langeheine, R. (Eds.). (1997). Applications of latent trait and latent class models in the social sciences. Waxmann.

Scrucca, L., & Raftery, A. E. (2015). Improved initialisation of model-based clustering using Gaussian hierarchical partitions. *Advances in Data Analysis and Classification, 9*(4), 447–460. https://doi.org/10.1007/s11634-015-0220-z

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016, August). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research, 48*(6), 775–815. https://doi.org/10.1080/00273171.2013.827564

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). IAP.

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*(1), 11–36. https://doi.org/10.1007/s12186-012-9091-z

Uebersax, J. (2012). *LCA software*. http://john-uebersax.com/stat/soft.htm

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, advanced, and syntax*. Statistical Innovations Inc.

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). Wiley.

Johannes Bauer (PhD) is full professor of educational research and methodology at the University of Erfurt, Germany. His research addresses the development of professional competences in higher education and during working life. Current research topics include (pre-service) teachers' competences to appraise, argue with and use research evidence in professional and everyday situations; understanding and debunking educational misconceptions and myths; and communication between experts and laypersons with a specific focus on training and assessing communication competence in medical education using digital tools.

Chapter 12
PLS-Based Structural Equation Modelling: An Alternative Approach to Estimating Complex Relationships Between Unobserved Constructs

Michael Goller and Frederic Hilkenmeier

Abstract A traditional approach to test complex relationships between different unobserved constructs included in theoretical models is to apply covariance-based structural equation modelling (CB-SEM). This chapter aims at introducing an alternative approach to estimating structural equation models that has not yet widely been used in research on professional learning and development or in research on learning in general: Partial-least squares structural equation modelling (PLS-SEM). PLS-SEM is based on ordinary least square regression analysis and uses an iterative algorithm to find parameter estimates. This estimation approach has several advantages including fewer statistical assumptions. In addition, PLS-SEM allows for the incorporation of both lower order and higher order formative constructs as well as for estimating rather complex models, which is not always possible with CB-SEM. The conceptual explanation of this particular SEM technique will be illustrated using a replication study of a published research study, focussing on the influence of learner factors and learning context on different professional learning activities.

Keywords PLS · Partial least square · SEM · Informal learning · Workplace learning · Professional development · Learning culture

M. Goller (✉)
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: michael.goller@upb.de

F. Hilkenmeier
Hochschule Fresenius, Psychology School, Hamburg, Germany
e-mail: frederic.hilkenmeier@hs-fresenius.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_12

12.1 Introduction

Similar to research in other fields within social and behavioural sciences, research on professional learning and development is often interested in explaining complex relationships between constructs that cannot be observed directly. Typical examples are studies that aim to explain how different characteristics of the work environment as well as different employee attributes themselves explain learning-related behaviours (e.g., Cangialosi et al., 2020; Goller, 2017; Leicher et al., 2013), learning transfer (e.g., Cheng & Ho, 2001; Gegenfurtner, 2013) or learning motivations and intentions (Kyndt et al., 2011; Maurer et al., 2003). All those studies have in common that they translate theoretical ideas into research models in which constructs are linked with each other in a set of complex relationships (e.g., direct relationships, mediation, moderation).

These constructs are theoretical in nature and provide a degree of abstraction that permits us to generalise relationships (Bollen, 2002). Since they cannot be observed directly, constructs need to be operationalised using a set of distinctive manifest indicators (Hyland, 1981). In statistical terms, the entirety of relationships among these constructs (or latent variables, as they are often called after being represented in a dataset) is usually described as a structural model and the set of correspondence rules between the unobserved constructs and their observed indicators is referred to as the measurement model (Hair et al., 2018). In the following, these theoretical, unobserved, latent constructs or variables will simply be referred to as "constructs", and the observed, manifest variables will simply be referred to as "indicators".

The typical method for estimating and testing relationships between constructs is to apply covariance-based structural equation modelling (CB-SEM; Chin, 1995; Hair et al., 2016). CB-SEM is a combination of factor analysis and multiple regression to simultaneously estimate parameters of both the specified measurement and the structural model. Besides giving insights on how the different constructs in the model relate with each other, CB-SEM allows researchers also to determine how well a specified measurement and structural model fits the observed data and therefore gives information about the adequacy of the theoretical assumptions using global goodness-of-fit criteria like χ^2 significance test, *CFI*, *RMSEA* or *SRMR*. A range of software packages exists for using CB-SEM in a multitude of application contexts (e.g., R lavaan, Mplus). Unfortunately, CB-SEM requires data to be multivariate normally distributed (at least when using the popular maximum likelihood estimator) with independent observations, demands relatively large sample sizes, and is not easily able to estimate complex relationships like moderating effects (Chin & Newsted, 1999; Hair et al., 2016). These requirements sometimes pose problems for scholars interested in professional learning and development, as data in behavioural and social sciences are often not multivariate normally distributed which can cause additional statistical issues (although correction methods and approaches for estimating robust standard errors exist; Kline, 2015; Satorra & Bentler, 1994).

Structural models based on constructs, however, cannot be estimated and investigated using CB-SEM only. An alternative approach is to use the composite-based partial least square structural equation modeling (PLS-SEM; Esposito Vinzi et al., 2010; Hair et al., 2016). PLS-SEM is based on classical ordinary least squares (OLS) regression for estimating parameters of the structural as well as the measurement model. In comparison to CB-SEM, the main advantage of PLS-SEM is that it does not require data to stem from a (multivariate) normal distribution. In addition, PLS-SEM allows for specifying rather complex models, including moderators as well as formatively operationalised constructs (see Chin & Newsted, 1999; Hair et al., 2016; see Ridgon, 2016, for a more in-depth view). However, PLS-SEM also comes with certain disadvantages that will be discussed later.

Although PLS-SEM is approximately as old as CB-SEM (see, e.g., Wold, 1975 vs. Jöreskog, 1970), it has only been used by a few scholars in the field and it is not often taught in methods courses. As such, it is less well known within the scientific community. The aim of this contribution is to introduce PLS-SEM and to explain its potential use in research on professional learning and development. For this purpose, Sect. 12.2 gives a short introduction to the general topic of structural equation modelling. Section 12.3 then describes PLS-SEM as an approach for estimating and investigating structural equation models. It is followed by a description of a real-word study undertaken by the chapter authors to illustrate PLS-SEM in Sect. 12.4. The aim of this study was to replicate the findings of a published study (Hilkenmeier et al., 2021). The chapter ends with a conclusion presenting suggestions concerning the use of PLS-SEM.

12.2 Structural Equation Modelling in a Nutshell

12.2.1 Relationships Between Constructs and Indicators: The Measurement Model

Since theoretical constructs cannot be observed directly, they cannot be assessed straightforwardly (Jöreskog & Sörbom, 1979) but must be operationalised based on theoretical assumptions. Such assumptions might include ideas of particular manifestations of these constructs or certain operations that are indeed observable and can be measured. For instance, someone might be interested in job involvement as part of work-related attitudes; a suitable approach to operationalise this construct could be to come up with a set of items describing how strongly someone is involved in their job. A sample item could be "The most important things which happen to me involve my job" (Maurer et al., 2003), which then can be reacted to by the study participants on a Likert-type rating scale from *Strongly disagree* to *Strongly agree*. If it is assumed that the construct causally determines how study participants respond to the observable variables, then the operationalisation follows a *reflective measurement* mode and the indicators are referred to as "effect indicators" (Bollen & Ting, 2000). The theoretical idea here is that each effect indicator stems from the same content domain and measures the same thing. That is why it is expected that all effect indicators measuring one construct are strongly related (correlated) to each other.

If it is instead assumed that the observed variables measure distinct aspects of the construct and that the observed variables form or shape the construct, then a *formative measurement* mode is used and the indicators are referred to as "causal indicators" (Bollen & Ting, 2000; Diamantopoulos & Winklhofer, 2001; see Sarstedt et al., 2016 for a more in-depth view). Causal indicators are not necessarily related to each other. For instance, the construct job satisfaction might be measured by asking respondents about their satisfaction with their salary, their working hours, their colleagues and so on, which might be interrelated or not. In structural equation modelling, the relationship between constructs and their respective indicators is referred to as the measurement model (see "A" and "B" in Fig. 12.1). What measurement mode should be used between construct and indicators is a theoretical question. However, confirmatory tetrad analysis might help to support these considerations on an empirical basis (Bollen & Ting, 2000; modified for PLS by Gudergan et al., 2008).

Both reflective and formative constructs can either be operationalised as *first-order* or *second-order* constructs. First-order constructs measure constructs directly with measurable manifest indicators. Second-order constructs instead use other first-order constructs as indicators (Jarvis et al., 2003). For instance, researchers might be interested in constructs such as learning culture at work that are theoretically

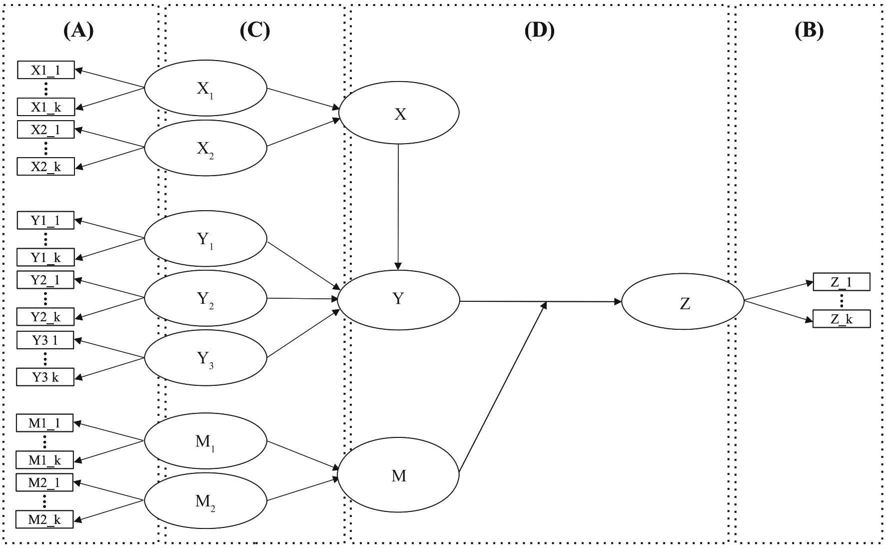

Fig. 12.1 Example of a structural equation model. (**A**) Reflective measurement model of the exogenous first-order constructs; (**B**) Reflective measurement model of the endogenous first-order construct; (**C**) Formative measurement model of the endogenous second-order constructs; (**D**) Structural model

assumed to be composed of different facets like supervisor support, coworker support, or learning conditions (e.g., Hilkenmeier et al., 2021). A second-order construct might now be used to operationalise this construct so that it becomes a composite of its different facets (formative measurement mode) that are themselves measured as reflective first-order constructs. This second-order construct can now be used as a single variable within a structural equation model (see “C” in Fig. 12.1).

12.2.2 Relationship Types Specified Between Constructs: The Structural Model

While operationalisation of the constructs is necessary to empirically investigate the phenomena in question, researchers are usually more interested in the complex relationships between different first- or second-order constructs. The entirety of these relationships is called the structural model (see “D” in Fig. 12.1). Such relationships describe, on a theoretical level, how constructs affect each other either on a correlative or a causal foundation. Four main relationship types can be distinguished: (a) two constructs *correlate* with each other ($X \leftrightarrow Y$), (b) two constructs are in a *direct relationship* when one construct directly affects another ($X \rightarrow Y$), (c) within a *mediated relationship* two constructs are related via another third construct ($X \rightarrow Y \rightarrow Z$), and (d) a *moderated relationship* is characterised by a direct relationship between two constructs in which the strength and direction of this relationship is affected by a third construct, a so-called moderator (M). These relationship types can be used in combination within a structural model to represent the assumptions of the underlying theory (see Fig. 12.1 for an example).

12.2.3 Estimating Structural Equation Models

In the context of SEM, the theoretical ideas about the different relationships between the constructs in question as well as the measurement modes of the individual constructs are translated into a conceptual research model (often graphically depicted as in Fig. 12.1). This research model represents the relationships between the constructs (structural model) as well as the relationships between the constructs and their indicators (measurement model). Using different estimating techniques, this representation of the theory can now be used to obtain empirical information about the degree of assumed relationships between different constructs (via regression coefficients) as well as the constructs and their indicators (via factor loadings as well as weights, see below). Besides the advantage of incorporating an explicit measurement model into the parameter estimation, which is not possible in the world of generalised linear models (i.e., regressions based on simple mean scores of the underlying manifest indicators), another strongly relevant benefit of using

SEM emerges: A construct can simultaneously be specified as a dependent (endogenous) as well as an independent (exogenous) variable in the model (Hair et al., 2018). Again, this is something not possible using generalised linear models where a variable must be specified as either one.

As briefly described in the introduction, two different statistical methods exist to specify and estimate structural equation models: Covariance-based SEM (CB-SEM) and partial least square structural equation modeling (PLS-SEM). CB-SEM is often seen as the standard approach to SEM due to its ability to estimate regression coefficients that are corrected for potential measurement errors included in the indicators used for operationalising the constructs. In other words, it estimates the relationship between the constructs as if the underlying indicators were 100% reliable, thus leading to stronger relationships between the constructs. In addition, CB-SEM can provide information that allows for assessing how well the whole model fits the available empirical data (i.e., global model fit). Its focus explicitly lies on model testing.

In comparison, PLS-SEM uses a set of multiple ordinary least square regressions on all available data for parameter estimation. The main advantages of this estimation procedure in comparison to CB-SEM are as follows: (a) data do not need to stem from a (multivariate) normal distribution, (b) constructs can be operationalised using formative measurement models, (c) constructs can easily be operationalised as higher order variables, (d) structural models can include moderator relationships, and (e) a higher level of statistical power is available to detect existing relationships (Chin & Newsted, 1999). In addition, some scholars suggest that PLS-SEM requires smaller sample sizes for model estimation (e.g., Chin & Newsted, 1999; Hair et al., 2016). However, other authors question this characteristic (e.g., Rigdon, 2016). The benefits of PLS-SEM come with at least three relevant downsides. First, PLS-SEM is not able to correct for measurement errors included in the manifest indicators. This leads typically to an overestimation of parameters of the measurement model and an underestimation of the parameters of the structural model (the so-called PLS-bias; Hair et al., 2016). In other words, PLS-SEM estimates the relationships between constructs in a more conservative way than does CB-SEM. Second, no accepted information criteria exist that can be used to assess the global model fit (see Sect. 12.3.2). That is why PLS-SEM cannot be used to test research models from a strict perspective. Third, PLS-SEM might capitalise on correlations that do not exist in the population but are non-zero due to sample variability (chance correlations), thereby producing less reliable parameter estimates than CB-SEM (Rönkkö, 2014). This issue seems to affect small models with weakly related constructs more strongly than larger models or models with strong relationships between the respective constructs (Rai et al., 2013). Table 12.1 summarises the differences between the two estimation approaches.

Although PLS-SEM comes with relevant disadvantages, it also has benefits that could make it attractive for researchers in general as well as those interested in professional learning and development. In particular, its capacity to test complex research models provides such researchers with a powerful method for investigating their theoretical ideas (e.g., studies that aim at investigating models that contain

Table 12.1 Features of CB-SEM and PLS-SEM

Features	CB-SEM	PLS-SEM
Assumptions:	Multivariate normal distribution and independent observations.	Similar to OLS regression (e.g., spherical errors, normally distributed errors).
Variable types:	Typically only reflective indicators.	Reflective and formative indicators possible (also: Different types of higher order constructs).
Measurement errors:	Corrects for measurement errors.	Does not correct for measurement errors.
Model complexity:	Limited.	Large complexity possible (e.g., moderated relationships).
Global model fit:	Existing (χ^2, *CFI*, *RMSEA*, *SRMR* etc.).	Existing (*GoF*) but deprecated.

Note: Based on Chin and Newsted (1999) and Hair et al. (2016)

moderated relationships based on data that are not normally distributed). The next section will give a more detailed account of how PLS-SEM works and how it can be used in research.

12.3 Partial Least Square SEM

12.3.1 PLS Algorithm

PLS-SEM applies a special algorithm to estimate model parameters. This algorithm will be briefly introduced in this section. Readers who are not interested in the technical details on how path coefficients are estimated can skip those descriptions and jump directly to Sect. 12.3.2.

PLS-SEM uses a set of simple OLS regressions to estimate factor loadings (λ; for reflectively operationalised constructs; quantification on how strongly an item is related to a construct), outer weights (*w*; for formatively operationalised constructs; value that determines with what weight an item forms a construct), as well as regression coefficients (β; strength of the relationships between the constructs specified in the structural model). In order to obtain estimates for these parameters, PLS-SEM employs an iterative algorithm that includes three distinct stages (see especially Sanchez, 2013, but also Henseler et al., 2012 and Lohmöller, 1989):

1. Within the *first stage*, the algorithm estimates the scores of the constructs. For this purpose, in *Step i*, attention is paid only to the measurement model where values of outer weights are assigned to each individual indicator. Most commonly, in the first round, all outer weights are set to "1". These weights are then used to calculate a *provisional* factor score for each construct as the weighted linear sum of its indicators. In this first round of the iterative process, the provisional factor scores are identical to the summed scores of the indicators, since each

indicator has the same outer weight. To conclude this first step, the *provisional* factor scores are z-transformed. It should be noted here that those *provisional* factor scores are, at this time, entirely determined by their indicators.

In S*tep ii*, attention is paid to the structural model. As discussed, the structural model specifies the theoretical assumed relationships between the constructs. These relationships are now *provisionally* estimated using the z-transformed *provisional* factor scores of *Step i* in simple regression analyses. This means that each construct is separately regressed on any other construct in the structural model with which it is connected. Since the *provisional* factor scores used here are simply the z-transformed summed scores of the indicators, these *provisional* regression coefficients are identical to those one would get using simple regression analysis by hand. As of now, *provisional* factor scores are still a linear combination of their indicators (see *Step i*).

In S*tep iii*, each factor score is now re-estimated. How exactly this re-estimation takes place depends on the so-called weighting scheme (centroid, factor, or path); however, the basic idea is similar among the three weighting schemes. For instance, in the factor weighting scheme, a factor score is estimated using the *provisional* regression coefficients β of S*tep ii.* This means, if construct A is connected to only one other construct B, the re-estimated factor score for construct A is calculated as the provisional regression coefficient between A and B times the *provisional* factor score of B. If construct C is connected to construct D and construct E, the re-estimated factor score for construct C is calculated as the provisional regression coefficient between C and D times the *provisional* factor score of D plus the provisional regression coefficient between C and E times the *provisional* factor score of E. To conclude S*tep iii,* these newly re-estimated factor scores are again z-transformed. It follows that the re-estimated factor scores are no longer linear combinations of their respective indicators but are now linear combinations of the constructs they are connected to ($\hat{y}$ in terms of linear regression).

In *Step iv*, the outer weights used in *Step i* are updated by considering the new estimate of the factor scores obtained in *Step iii*. For reflectively operationalised constructs this is done by estimating the new outer weights individually as the individual regression coefficients obtained by calculating an individual simple regression for each effect indicator on their *provisional* factor score from *Step iii*. For formatively operationalised constructs this is instead done by calculating one single multiple regression with the *provisional* factor score regressed on all its causal indicators at once. To conclude *Step iv*, these outer weights are z-transformed.

Using these new outer weights, *Steps i to iv* are repeated until the weights between the current and the last iteration do not considerably change anymore (i.e., the algorithm converges based on an a priori defined convergence criterion, e.g., $\sum|\Delta w| < 10^{-5}$). By now the final outer weights are obtained, as are the final factor scores–that is, the last *provisional* factor scores of *Step i* of the final iteration.

2. In the *second stage*, these final factor scores are used to estimate the true regression coefficients between the constructs by calculating different (multiple) regressions for each endogenous construct with its predicting exogenous construct(s).
3. In the *third and last stage*, the indicator loadings of the reflectively operationalised constructs are estimated. This is done by calculating the individual regression coefficients between the final factor scores of each construct and their respective effect indicators. The outer weights for formatively operationalised constructs are directly taken from the multiple regressions calculated in *Step iv* of the final iteration.

It should be clear from this short description of the algorithm that parameter estimation within the PLS-SEM approach is based upon a series of simple and multiple OLS regressions that focus on parts of the specified structural equation model (hence: *partial* least square SEM). Due to the fact that PLS-SEM uses only OLS-based regressions to obtain parameter estimates, its assumptions are essentially similar to those usually described in the context of linear regression: (a) absence of multicollinearity (i.e., no strong correlations between independent variables in the model), (b) homoscedasticity (i.e., the variance of the error terms in a regression does not depend on the values of the independent variables), (c) strict exogeneity (i.e., error terms of a regression are random and have a mean of zero), and (d) independence of errors (i.e., observations of the error term are uncorrelated with each other). Significance tests of obtained parameter estimates are typically based on nonparametric bootstrapping procedures within the PLS-SEM approach.

Both the conceptual and mathematical ideas behind PLS-SEM and CB-SEM differ significantly, and there is ongoing debate about the merits and fallacies of both methods (see, e.g., Rönkkö, 2014; Ridgon, 2016). However, simulation studies show that both approaches often result in relatively similar parameter estimates (e.g., Reinartz et al., 2009), with some authors suggesting that PLS-SEM performs better when characteristics of the underlying population are unknown (Sarstedt et al., 2016). This can be regarded as good news as the interpretation of the model estimates is not greatly affected by the choice for or against one of the methods.

12.3.2 Practical Issues Using PLS-SEM

Within this section a few recommendations concerning the practical application of PLS-SEM will be given. We hereby heavily draw on the excellent book written by Hair et al. (2016) as well as the extensive edited volumes by Esposito Vinzi et al. (2010) and Lathan and Noonan (2017) that describe the typical research process behind most projects employing PLS-SEM, as depicted in Fig. 12.2. For more details, readers are referred to these books.

After formulating a research interest and deriving a research question, a structural research model should be specified that reflects both theoretical assumptions as well

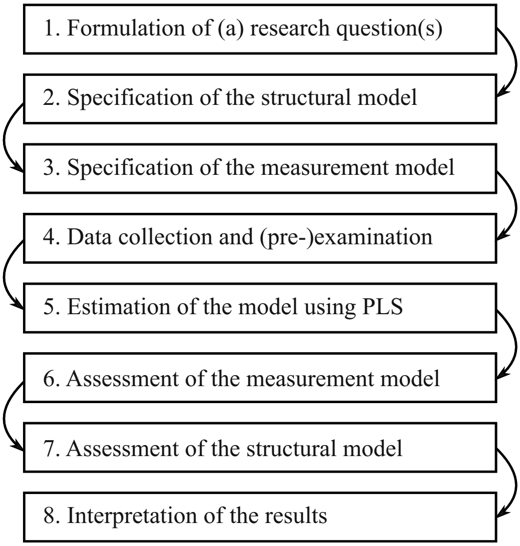

Fig. 12.2 Research process behind a project using PLS-SEM. (Adapted from Hair et al., 2016)

as the current state of research. In this step, it is necessary to identify all constructs of interest and to clearly specify the hypothesised relationships among those constructs. It is, however, necessary that causal loops in the structural model must strictly be avoided. In other words, it is not possible to specify an endogenous variable as (mediated) predictor of itself. Otherwise, the algorithm behind PLS-SEM is not able to converge. Besides the structural model, the measurement model needs to be specified. Decisions must be made about what manifest indicators are selected to operationalise the constructs and what measurement modes should be used in the measurement model (formative and/or reflective).

After model specification, decisions are made concerning the data collection procedure. It has been recommended to determine the concrete number of cases using power analysis (Hair et al., 2016) or even more elaborate calculations (e.g., Kock & Hadaya, 2018). However, other researchers advise a less rigorous approach by relying on rules of thumb, for example that the sample size should be larger than 10 times either the largest number of formative indicators in the measurement model or the largest number of structural paths directed at an endogenous variable (Barclay et al., 1995). As soon as the data are available they should be investigated for multicollinearity, outliers, and other issues that might cause problems when estimating linear regressions using OLS. Cases with large numbers of missing values (>15% or 30% of data points) might be removed from the dataset. The remaining missing values should be imputed to avoid case-wise deletion and therefore a significant sample reduction.

Estimating a model using PLS-SEM can be done with different software packages available on the market. Table 12.2 gives a short overview of a selection of those packages, including a short characterisation of each. SmartPLS is easy to use for beginners due to its graphical user interface. In addition, it still seems to offer the

Table 12.2 Software packages to estimate models using PLS-SEM

Software package	Characterisation
SmartPLS	Proprietary, graphical user interface, large set of statistical features to estimate PLS-SEM included.
plspm (R package)	Open source, script-based, embedded in the R environment, includes all standard features to estimate PLS-SEM including multigroup analysis and modelling of unobserved heterogeneity; requires R to run, not maintained any more.
semPLS (R package)	Open source, script-based, embedded in the R environment, includes all standard features to estimate PLS-SEM, requires R to run.
seminr (R package)	Open source, script-based, embedded in the R environment, includes all standard features to estimate PLS-SEM, requires R to run.
cSEM (R package)	Open source, script-based, embedded in the R environment, includes all standard features to estimate PLS-SEM, requires R to run.
plssem (Stata package)	Open source, script-based, embedded in the Stata environment, includes all standard features to estimate PLS-SEM including multigroup analysis, requires Stata (proprietary) to run.
XLSTAT-PLSPM (Excel plugin)	Proprietary, graphical user interface, includes all standard features to estimate PLS-SEM including multigroup analysis and modelling of unobserved heterogeneity, requires Microsoft Excel (proprietary) to run.

most statistical features to estimate PLS-SEM. However, R-based packages like plspm and semPLS are much more convenient in specifying large models and are generally more powerful due to their embeddedness in the R environment. In addition, they are open source and therefore free of charge.

Assessment of the measurement model can be undertaken inside the PLS-SEM framework as well as outside of it. Within the PLS-SEM framework the internal consistency of indicators operationalising reflective constructs can be investigated using Cronbach's α or measures of congeneric reliability (both should usually be above .70; Hair et al., 2018). In addition, indicator reliability should be checked by assessing significance and the size of indicator loadings which should be above .50. Discriminant validity between reflective constructs can be checked by using the Fornell-Larcker criterion (Fornell & Larcker, 1981; Hilkenmeier et al., 2020). A few authors additionally check the psychometric quality of reflective measurement models outside the PLS-SEM framework using confirmatory factor analysis (CFA; e.g., Goller, 2017; Goller et al., 2020). This approach has the advantage that the measurement model can be scrutinized more strongly based on fit indices assessing the global model fit. However, all these recommendations apply only for reflective measurement models. Formative measurement models should be checked for collinearity issues based, for instance, on the variance inflation factor (*VIF*) that should be below 5.0 (Hair et al., 2018). In addition, all outer weights of a formative construct should significantly differ from zero.

If the psychometric qualities of the measurement model are confirmed to be acceptable, the structural model can be assessed. In a first step, collinearity issues based on the factor scores of the constructs should be investigated using measures like *VIF*. If multicollinearity can be ruled out, the statistical significance (p values) as well as the size of the regression coefficients describing the assumed relationships between the constructs in the model can be assessed. In a second step, the coefficient of determination R^2 of each endogenous variable is inspected. A generalisable effect size characterisation of R^2 does not exist (Götz et al., 2010). However, R^2 of above .20 might be considered to be substantial in studies concerned with human behaviour (Hair et al., 2016). In a third step, partial f^2 effect sizes can be calculated to obtain additional information on whether single variables or a set of variables have substantive impact on an endogenous construct within the model. Based on its definition, partial f^2 provides information on whether the omitted predictor(s) have substantial impact on the endogenous construct in question (Hair et al., 2016). Values of .02, .15, and .35 can be characterised as small, medium, and large effect sizes, respectively (Cohen, 1988).

PLS-SEM is a suitable method for estimating structural equation models, especially when the specified model is rather complex (e.g., due to moderated relationships or higher order constructs in the model), formatively operationalised constructs are included, or the available data do not meet the distributional assumptions of CB-SEM. PLS-SEM should not be used if the sole empirical focus is on theory testing, since no accepted information criterion exists that allows the assessment of the global model adequacy like *CFI*, *RMSEA* or *SRMR* in CB-SEM. A goodness-of-fit index (*GoF*) has been proposed (Tenenhaus et al., 2004, 2005) but is highly debated (Henseler & Sarstedt, 2013). In fact, the current literature deprecates its use (Hair et al., 2016; Henseler & Sarstedt, 2013). Taken together, PLS-SEM is rather suited for research endeavours that aim at identifying those predictors that most strongly explain particular endogenous variables as well as for projects in which theoretical models need to be estimated that cannot be specified with CB-SEM. CB-SEM should be used instead if the main focus is on theory testing and the researchers can ensure that the corresponding algorithm is able to estimate the model in question.

12.4 Illustrating PLS-SEM with a Replication Study

As briefly described in the introduction, the empirical example of a PLS-SEM study provided here is a replication of the research project published in Hilkenmeier et al. (2021). Therefore, the discussion of the theoretical framework as well as the method is an abbreviated version of the original study. Interested readers can readily find a more sophisticated derivation of the hypotheses as well as a more detailed description of the instruments used in Hilkenmeier et al. Moreover, the description of the results is compressed as well. The idea behind this short illustrative example is to give readers a general idea on how to proceed when conducting a PLS estimation, rather than explaining and reporting each step in detail. If readers are interested in using PLS-SEM themselves, they should refer to the resources discussed in the first paragraph of Sect. 12.3.2.

12.4.1 Motivation and Theoretical Framework

The heart of the argument in Hilkenmeier et al. (2021) is that designated professional learning activities (commonly referred to as formal learning) and learning afforded directly by the workplace (commonly referred to as informal or workplace learning) are very different in nature and, therefore, participation in these different learning activities (i.e., formal and informal learning) should be influenced by different antecedents. On the one hand, the core characteristic of designated professional learning activities is that they are explicitly planned for knowledge or skill development and take place in an a priori structured context using pedagogical means. As discussed by Butler and Brooker (1998) or Vaughan (2008), the intention behind such structured education is to enable incumbents to develop a deeper understanding of work-related issues beyond the immediate demands of the workplace. Thus, the expectation (and intention) when participating in designated learning activities like workshops, seminars, pre-planned instructions at the workplace or structured job-rotation, to name just a few, is indeed to learn something new. It is therefore not surprising that participation in these forms of professional development is usually intentional and consciously planned (e.g., Kyndt & Baert, 2013). And thus, following Ajzen's (1991) theory of planned behaviour, participation highly depends on employees' expected benefits from the learning activity, their attitudes towards learning, or their self-efficacy in regard to learning (Maurer et al., 2003)–that is, their learning-related beliefs.

On the other hand, learning afforded directly by the workplace largely arises as a by-product of everyday experiences at work and is related to the employees' current situation and, hence, is more spontaneous. For instance, when employees need to tackle unfolding and pressing tasks or problems, "there is a clear work-based goal" (Tynjälä, 2013, p. 18). Thus, when employees experiment with new strategies, discuss their problem with colleagues, or read codified information that might help with the situation at hand, there is a focus on "production" (Butler & Brooker, 1998), and the acquired knowledge and skills are most likely highly contextualised (Vaughan, 2008). In fact, employees often lack awareness of their own learning (Eraut, 2004). In essence, the learning process itself, its objectives, content, and means of acquisition, strongly depend on situational conditions. How often employees engage in (non-routine) situations that bring forward these situational conditions and therefore offer opportunities for workplace learning mainly depends on two factors (Hilkenmeier et al., 2021). First, employees' engagement in situations with high learning potential should strongly depend on their work-related attitudes–that is, how they identify themselves with their work and how engaged they are in their work (e.g., Goller, 2017). And second, employees' engagement in workplace learning activities should depend on the workplace itself (i.e., if the workplace affords any opportunities for learning)–that is, if the organisational culture is conducive to learning and knowledge sharing. Such a facilitative learning culture consists of interpersonal support (e.g., Facteau et al., 1995), institutionalised support (e.g., Cerasoli et al., 2017) or employees' empowerment (e.g., Janz & Prasarnphanich, 2003).

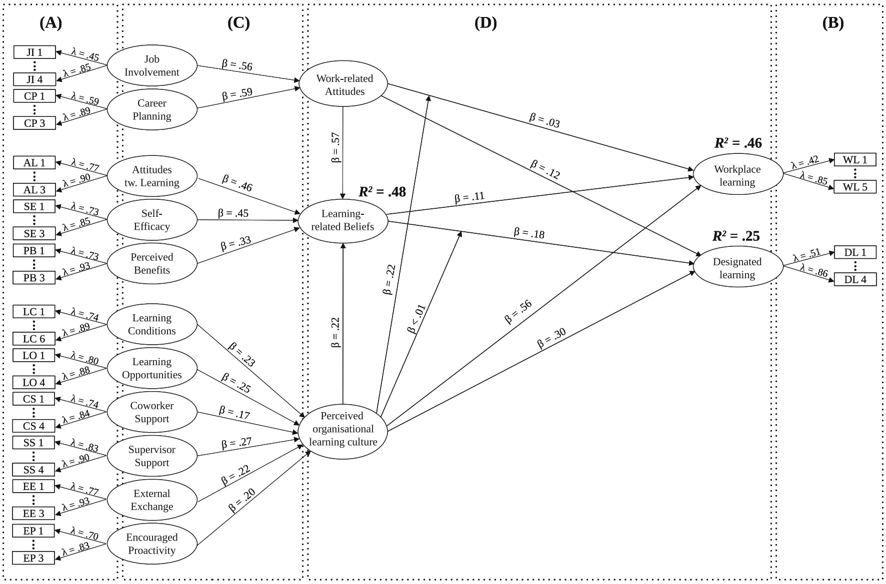

Fig. 12.3 PLS-SEM model and estimates of the replication study used in Sect. 12.4. (**A**) Reflective measurement model of the exogenous first-order constructs with minimum and maximum loadings for each construct; (**B**) Reflective measurement model of the endogenous first-order constructs with minimum and maximum loadings for each construct; (**C**) Formative measurement model of the endogenous second-order constructs with weights (all significant, two-tailed); (**D**) Structural model, all path coefficients <.22 are nonsignificant (two tailed) results

Indeed, Hilkenmeier et al. (2021) found empirical support for their hypotheses that learning-related beliefs determine employees' participation in designated learning activities, but not in workplace learning. Furthermore, their PLS-SEM model (see Fig. 12.3 for the general structure of the model; see Table 12.4 for path coefficients of the original study as well as the current replication) showed a strong direct, indirect, and moderating influence of organisational learning culture on both engagement in workplace learning and participation in designated learning opportunities to the point that partial effect sizes show barely any effects for the hypothesized individual antecedents learning-related beliefs and work-related attitudes (Partial $f^2 = .01$ to .02). In the face of recent discussions about the importance of human agency and individuals' proactivity for workplace learning and professional development (e.g., Goller, 2017; Goller & Paloniemi, 2017), the dominance of an organisational factor like learning culture is somewhat unexpected.

Hilkenmeier et al. (2021) used PLS-SEM for model estimation. As discussed above, the reliability of path coefficients estimated using this approach has been questioned due to the PLS algorithm capitalising on chance correlations that are idiosyncratic to the sample in a given study (see Sect. 12.2.3). Therefore, to further test whether the overall structure reported in Hilkenmeier et al. holds true, the present empirical study reruns the whole analysis on a new convenience sample of

99 different German employees. These new data were modelled in the same way as in Hilkenmeier et al.'s original study. To compare the path coefficients estimated using the original sample and the path coefficients estimated using this new replication, sample parametric multi-group analyses are used (Hair et al., 2016).

12.4.2 Method

The replication uses a convenience sample of 99 German employees in total. According to the rules of thumb discussed in Sect. 12.3.2, this sample size should be sufficient since the largest number of formative indicators of any given construct is six (organisational learning culture, see Fig. 12.3). However, power-analyses show that the chance to find an effect of $f^2 = 0.1$ is only 67.4%. This relatively low power should be kept in mind when interpreting the results.

As discussed in Sect. 12.2.1 and seen in Fig. 12.3, all independent constructs are operationalised as formative second-order ones with several reflective first-order constructs underneath. This measurement structure allows for integration of different first-order constructs, which are distinct from each other, into a theoretical meaningful and parsimonious aggregate (Diamantopoulos et al., 2008). All constructs used as independent variables were measured using multiple items (see Fig. 12.3) on a 5-point Likert scale at t_0 using established measurement instruments. The dependent construct "participation in designated learning opportunities" asked at t_0 for frequency of participation in different learning activities during the past year. The dependent construct "engagement in workplace learning" asked for intensity in engagement in different workplace learning activities during the given week. To increase reliability and to reach a stable sample of individuals' informal learning activities, engagement in workplace learning was repeatedly measured on a weekly basis from t_0 to t_5. Participants' responses to these repeated items were then averaged. As supported by confirmatory tetrad analysis (see Sect. 12.2.1), both dependent constructs were operationalised as reflective first-order constructs. For further details regarding the methods used, the interested reader is again referred to Hilkenmeier et al. (2021).

As described in Sect. 12.3.2, data analysis in the context of PLS-SEM is usually conducted in two separate stages: (a) evaluation of the measurement model, and (b) evaluation of the structural model. Since the present study is a replication study, it adds a third stage: (c) model comparison and predictive power.

Measurement Model (Stage 1) For reflective constructs, this means evaluating the internal consistency of the effect indicators, their loadings, as well as the discriminant validity between the reflective constructs. As can be seen in Table 12.3, composite reliabilities (*CR*), as a measure of internal consistency, are well above the established threshold of .70 for all reflective first-order constructs. Table 12.3 also reveals that the constructs' square root of the average variance extracted (*AVE*) is always higher than the highest correlation with any other construct. This means that each construct is more closely related to its own indicators than to any other construct within the study, thus establishing discriminant validity (Fornell & Larcker, 1981). Two of the 50 effect indicators displayed loadings below .50 on their respective constructs (job involvement indicator 1, $\lambda = .45$ and workplace learning indicator 1, $\lambda = .42$, also see Fig. 12.3). Since both constructs still exhibit *AVE*s larger than .50 and this study was planned as a replication study, these two items were retained.

The formative nature of the second-order constructs allows grouping of variables, which do not need to be correlated with each other and which cover different aspects of a construct to form a common construct. Therefore, it first must be ensured that all causal indicators (the respective first-order constructs in this case) significantly contribute to the formative construct in question. As presented in Fig. 12.3, all these outer weights indeed differ significantly from zero. Second, collinearity between the causal indicators needs to be checked to ensure stability of the estimated coefficients. As can be seen in Table 12.3, the maximal variance inflation factor (*VIF*) of all formative second-order constructs is 2.5 and therefore below well-accepted thresholds (Hair et al., 2018). It follows that, based on the presented analyses, both the reflective first-order constructs as well as the formative second-order constructs can be used to estimate the assumed relationships in the structural model.

Structural Model (Stage 2) Since the evaluation of the measurement model produced satisfactory results, the second stage of the data analysis focuses on the structural model. As described in Sect. 12.3.1, PLS-SEM uses a set of OLS-based regressions to obtain parameter estimates in this model part. Therefore, as a first step, it is important to again check for collinearity, this time among the constructs that are specified as predictors of other constructs in the structural model. In this replication study, the resulting *VIF* for work-related attitudes, learning-related beliefs, and organisational learning culture range from 1.3 to 1.9, which indicates that the path coefficients estimated by the PLS-SEM algorithm will not be affected by collinearity. Therefore, R^2 values of the endogenous variables, path coefficients, and partial f^2 effect sizes of the independent factors are inspected next. As can be seen in Fig. 12.3, the predicting constructs in the structural model together explain 25% of the variance in participation in designated (formal) learning, and 46% of the variance in engagement in (informal) workplace learning which, in Cohen's terms (1988), can both be classified as strong effects. As in Hilkenmeier et al. (2021), the organisational learning culture seems to be the strongest driver for both forms of learning activities, as indicated by the estimated path coefficients ($\beta = .30$ and $\beta = .56$, $p < .01$). In fact, it is the only significant driver for both forms of learning activities (which, of course, might at least be partially due to the relatively low power; see Sect. 12.4.2). This dominance of the organisational context is further corroborated by inspection of the partial effect sizes. Partial f^2 is calculated by estimating the full model with all predictors and then a reduced model without the respective predictor (or group of predictors) for which the partial f^2 shall be calculated. As seen in Fig. 12.3, the full model explains 46% of variance in engagement in workplace learning. Without the

Table 12.3 Reliabilities, Collinearity, and Intercorrelations among the First- and Second-Order Constructs

Construct	*Items*	*M*	*SD*	*CR*	1	2	3	4	5	6	7	8	9	10	11	12	13	A	B
(1) Job involvement	4	3.70	0.67	.83	*.79*														
(2) Career planning	3	3.96	0.75	.83	.52	*.75*													
(3) Attitudes towards learning	3	4.17	0.71	.88	.43	57	*.84*												
(4) Perceived benefits from learning	3	3.97	0.65	.87	.33	.34	.56	*.83*											
(5) Self-efficacy regarding learning	3	3.78	0.65	.83	.42	.65	.54	.27	*.79*										
(6) Learning conditions	6	3.55	0.93	.93	.37	.25	.41	.17	.40	*.84*									
(7) Learning offerings	4	3.12	1.09	.91	.09	.08	.25	.10	.21	.63	*.85*								
(8) Coworker support	4	3.58	0.76	.87	.28	.24	.35	.23	.23	.41	.36	*.79*							
(9) Supervisor support	4	3.51	1.03	.92	.37	.26	.42	.13	.26	.70	.62	.44	*.87*						
(10) External exchange	3	2.79	1.15	.91	.24	.10	.23	−.02	.16	.41	.41	.29	.44	*.87*					
(11) Encouraged proactivity	4	2.91	1.15	.86	.35	.31	.39	.12	.34	.37	.29	.22	.48	.54	*.78*				
(12) Designated learning opportunities	4	3.83	0.71	.76	.25	.30	.29	.22	.39	.34	.35	.13	.30	.32	.26	*.65*			
(13) Workplace learning	5	2.91	0.61	.84	.22	.23	.36	.16	.35	.44	.48	.38	.55	.42	.41	.41	*.73*		
	Constructs			*VIF*															
(A) Work-related attitudes	2			1.4	.86	.88	.57	.38	.61	.34	.12	.29	.37	.20	.38	.33	.27	–	
(B) Learning-related beliefs	3			1.9	.50	.66	.89	.71	.79	.43	.25	.34	.36	.18	.37	.39	.38	.66	–
(C) Organisational learning culture	6			2.5	.38	.29	.47	.16	.39	.81	.78	.58	.85	.69	.64	.43	.64	.39	.45

Note: *M* *M*ean of the composite score, *SD* Standard deviation of the composite score, *CR* composite reliability. Square root of the *AVE* on the diagonal. *VIF* max. variance inflation factor of all first-order constructs of this particular second-order construct. All $r \geq .20$ differ significantly from zero with $p < .05$ (two-tailed test) individual factors learning-related beliefs and work-related attitudes, the organisational learning culture still explains 42% of the variance in engagement in workplace learning, leaving only a small partial effect for individual antecedents ($\Delta R^2 = 4$ percentage points, partial $f^2 = .07$). In accordance, a reduced model without the organisational learning culture explains only about 15% in variance in engagement in workplace learning, leaving a large effect for the organisational factor ($\Delta R^2 = 31$ percentage points, partial $f^2 = .57$). The same is true for participation in designated learning opportunities. Here, the full model explains 25% of the variance in the dataset, the reduced model without the individual factors still 19% (partial $f^2 = .08$, small effect), and a reduced model with only the individual factors and without the organisational context can explain 16%, leaving a small-to-medium effect for organisational learning culture (partial $f^2 = .12$).

Model Comparison and Predictive Power (Stage 3) To quantitatively corroborate the reliability of the path coefficients reported in Hilkenmeier et al. (2021), the invariance of the models estimated in the original sample and the replication sample was inspected. To that end, strict factorial invariance was established within the measurement model using covariance-based confirmatory factor analysis (Dimitrov, 2010; see Sect. 12.3.2 for checking psychometric quality of reflective measurement models outside the PLS-SEM framework to scrutinise it based on global fit indices). Next, multi-group analyses for the structural model were conducted. Multi-group analyses test for differences between identical models estimated for different groups of respondents (Chin & Dibbern, 2010; Hair et al., 2016), in this case the estimated path coefficients of Hilkenmeier et al. (2021) with 229 participants, and the estimated path coefficients reported here with 99 participants. As can be seen in Table 12.4, none of the path coefficients in the structural model significantly differ between the original and the replication study, indicating reliability of the path coefficients and thus validity of the overall structure of the proposed model.

In addition, an analysis of the predictive power following the procedure outline by Shmueli et al. (2016) revealed that the parameter estimates of the original study by Hilkenmeier et al. (2021) are strong enough to predict 11% of the variance in participation in designated learning activities, and 32% of the variance in engagement in workplace learning using the data of the 99 participants of the replication study. In other words, the prediction effects are, according to Cohen (1988), still medium-to-large and large, respectively.

12.4.3 Discussion

As set out at the beginning of Sect. 12.4, the aim of the present replication study was to test the reliability of the path coefficients of Hilkenmeier et al. (2021) and thus the validity of the overall structure of the model estimated there. The key finding of Hilkenmeier et al. was the somewhat unexpected strong dominance of the organisational context, namely, the organisational learning culture, on both

Table 12.4 Results of the parametric multi-group analyses comparing the original and the replicated model in regard to the structural model

	Original sample (n = 229)		Replication sample (n = 99)		Sample comparison		
	β_o	SE_r	β_r	SE_o	$\lvert\beta_o - \beta_r\rvert$	t^*	p
Work-related attitudes → learning-related beliefs	.54	0.05	.57	0.08	0.03	0.39	.69
Organisational learning culture → learning-related beliefs	.17	0.05	.22	0.10	0.05	0.48	.63
Learning-related beliefs → designated learning opportunities	.17	0.07	.18	0.11	0.01	0.05	.96
Learning-related beliefs → workplace learning	−.11	0.07	.11	0.11	0.22	1.65	.10
Work-related attitudes → designated learning opportunities	−.03	0.08	.12	0.13	0.16	1.05	.30
Work-related attitudes → workplace learning	.03	0.08	.03	0.11	<0.01	0.01	.99
Organisational learning culture → designated learning opportunities	.32	0.07	.30	0.11	0.02	0.17	.86
Organisational learning culture → workplace learning	.38	0.06	.56	0.08	0.18	1.65	.10
Learning-related beliefs * Org. learning culture → Designated learning opportunities	.03	0.08	.22	0.14	0.19	1.18	.24
Work-related attitudes * Org. learning culture → Workplace learning	.19	0.08	<.01	0.11	0.19	1.36	.18

Note: β_o and β_r are standardised path coefficients of the original and replication sample, respectively; SE_o and SE_r depict the bootstrapped standard errors of β_o and β_r, respectively. * $df = 229 + 99 - 2 = 326$

participation in designated learning opportunities (formal learning) as well as engagement in workplace learning (informal learning). This key finding emerged in the replication sample as well: path coefficients as well as partial effect sizes clearly show the relative importance of organisational learning culture on both kinds of professional learning activities. Moreover, direct comparisons of the path coefficients of the original sample and the replication sample via multi-group analyses show no significant differences between the estimates. This robustness of the overall model structure found in Hilkenmeier et al. is further corroborated by its predictive power.

Taken together, the present study assesses the robustness of the PLS-SEM estimates and therefore presents evidence that the findings of Hilkenmeier et al. (2021) are not due to chance correlations (see Sect. 12.2.3) but are in fact reliable. In other words, the replication strongly indicates that the reported effects are not sample specific but seem to exist in the underlying population, too (see Zwaan et al., 2018). However, the replication does not provide further evidence for the generalisability of the results in the sense that it expands on the findings of Hilkenmeier et al., for instance by using different procedures or operationalisations. The selection of the constructs used to represent work-related attitudes in particular, although theoretically motivated, should be reconsidered in further studies to better reflect employees' agency and proactivity and therefore maybe give these individual antecedents "a better shot" at predicting learning behaviour and asserting themselves next to contextual antecedents.

12.5 Conclusion

PLS-SEM is a highly versatile approach for estimating structural equation models. In fact, it is well suited to analysing a set of complex relationships between different constructs including mediation and moderation. In comparison with a covariance-based approach towards structural equation modelling, PLS-SEM does not require its data to stem from a multivariate normal distribution and has, in general, a larger power to detect existing relationships between the constructs specified. Due to its underlying principles, PLS-SEM is also able to include constructs that might be operationalised in not only a reflective but also a formative way. A range of different software alternatives exists that allow for comfortably specifying and testing structural equation models using PLS-SEM. It follows that PLS-SEM could indeed be a viable proposition for researchers interested in professional learning and development who want to investigate relationships between constructs (as illustrated with the replication study of Hilkenmeier et al., 2021).

PLS-SEM should be particularly used in situations (a) where data cannot be assumed to stem from a certain distribution (which is often the case in the context of professional learning and development as both population and sample sizes are rather small and restricted), (b) when it is aimed to test models that include moderators, formative constructs or even higher order constructs (which might become useful in the context of professional learning and development as it allows complex theories to be investigated), and (c) when researchers want to identify drivers (i.e., certain independent variables) that most strongly explain particular phenomena (i.e., certain dependent variables; this allows both researchers and practitioners to find individual and organisational characteristics that can be further developed to foster professional learning and development). At the same time, however, potential users of PLS-SEM should be aware that this approach does not allow for strict model testing, as no accepted global model fit indices are provided. If the focus within a research project lies on testing existing theories, then CB-SEM should be used as long as researchers can make sure that the corresponding algorithm is able to estimate the research model in question.

As book chapters are limited in size, more advanced topics like the modelling of unobserved heterogeneity within the data similar to mixture models in CB-SEM (see Bauer, 2022; see in the PLS context: Esposito Vinzi et al., 2008; Ringle et al., 2013) have not been covered here. For a more extensive overview on PLS-SEM readers might want to consult existing textbook literature (e.g., Hair et al., 2016, 2018), relevant handbooks on PLS-SEM (Esposito Vinzi et al., 2010; Lathan & Noonan, 2017) or even published studies that use PLS-SEM as the primary research method (e.g., Goller, 2017; Goller et al., 2020).

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 243–268). Springer.

Barclay, D. W., Higgins, C. A., & Thompson, R. (1995). The partial least squares approach to causal modeling: Personal computer adoption and use as illustration. *Technology Studies, 2*, 285–309.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Bollen, K. A., & Ting, K.-F. (2000). A tetrad test for causal indicators. *Psychological Methods, 5*(1), 3–22. https://doi.org/10.1037/1082-989X.5.1.3

Butler, J., & Brooker, R. (1998). The learning context within technical and further education colleges as perceived by apprentices and their workplace supervisors. *Journal of Vocational Education & Training, 50*(1), 79–96.

Cangialosi, N., Odoardi, C., & Battistelli, A. (2020). Learning climate and innovative work behavior, the mediating role of the learning potential of the workplace. *Vocations and Learning, 13*(2), 263–280. https://doi.org/10.1007/s12186-019-09235-y

Cerasoli, C. P., Alliger, G. M., Donsbach, J. S., Mathieu, J. E., Tannenbaum, S. I., & Orvis, K. A. (2017). Antecedents and outcomes of informal learning behaviors: A meta-analysis. *Journal of Business and Psychology, 33*, 203–230.

Cheng, E. W. L., & Ho, D. C. K. (2001). The influence of job and career attitudes on learning motivation and transfer. *Career Development International, 6*(1), 20–28. https://doi.org/10.1108/13620430110381007

Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies, 2*, 315–319.

Chin, W. W., & Dibbern, J. (2010). An introduction to a permutation based procedure for multigroup PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In V. Esposito Vinzi, W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 171–193). Springer. https://doi.org/10.1007/978-3-540-32827-8_8

Chin, W. W., & Newsted, P. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203–1218.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273. https://doi.org/10.1080/158037042000225245

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.). (2010). *Handbook of partial least squares: Concepts, methods and applications*. Springer.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., & Tenenhaus, M. (2008). REBUS-PLS: A response-based procedure for detecting unit segments in PLS path modelling. *Applied Stochastic Models in Business and Industry, 24*, 439–458.

Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21*, 1–25.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39. https://doi.org/10.2307/3151312

Gegenfurtner, A. (2013). Dimensions of motivation to transfer: A longitudinal analysis of their influence on retention, transfer, and attitude change. *Vocations and Learning, 6*(2), 187–205. https://doi.org/10.1007/s12186-012-9084-y

Goller, M. (2017). *Human agency at work: An active approach towards expertise development*. Springer VS.

Goller, M., Harteis, C., Gijbels, D., & Donche, V. (2020). Engineering students' learning during internships: Exploring the explanatory power of the job demands-control-support model. *Journal of Engineering Education, 109*(2), 307–324. https://doi.org/10.1002/jee.20308

Goller, M., & Paloniemi, S. (Eds.). (2017). *Agency at work: An agentic perspective on professional learning and development*. Springer.

Götz, O., Liehr-Gobbers, K., & Krafft, M. (2010). Evaluation of structural equation models using the partial least squares (PLS) approach. In V. Esposito Vinzi (Ed.), *Handbook of partial least squares: Concepts, methods and applications* (pp. 691–711). Springer.

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research, 61*(12), 1238–1249. https://doi.org/10.1016/j.jbusres.2008.01.012

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equations modeling (PLS-SEM)* (2nd ed.). SAGE.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2012). Using partial least squares path modeling in international advertising research: Basic concepts and recent issues. In S. Okazaki (Ed.), *Handbook of research in international advertising* (pp. 252–276). Edward Elgar.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics, 28*(2), 565–580. https://doi.org/10.1007/s00180-012-0317-1

Hilkenmeier, F., Bohndick, C., Bohndick, T., & Hilkenmeier, J. (2020). Assessing distinctiveness in multidimensional instruments without access to raw data–a manifest Fornell-Larcker criterion. *Frontiers in Psychology, 11*, 1–9. https://doi.org/10.3389/fpsyg.2020.00223

Hilkenmeier, F., Goller, M. & Schaper, N. (2021). The differential influence of learner factors and learning context on different professional learning activities. *Vocations and Learning, 14*(3), 411–438. https://doi.org/10.1007/s12186-021-09266-4

Hyland, M. (1981). *Introduction to theoretical psychology*. Macmillan Education UK. https://doi.org/10.1007/978-1-349-16464-6

Janz, B. D., & Prasarnphanich, P. (2003). Understanding the antecedents of effective knowledge management: The importance of a knowledge-centered culture. *Decision Sciences, 34*, 351–384.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199–218. https://doi.org/10.1086/376806

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*(2), 121–145. https://doi.org/10.1111/j.2044-8317.1970.tb00439.x

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Abt Books.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261.

Kyndt, E., & Baert, H. (2013). Antecedents of employees' involvement in work-related learning: A systematic review. *Review of Educational Research, 83*, 273–313. https://doi.org/10.3102/0034654313478021

Kyndt, E., Govaerts, N., Dochy, F., & Baert, H. (2011). The learning intention of low-qualified employees: A key for participation in lifelong learning and continuous training. *Vocations and Learning, 4*(3), 211–229. https://doi.org/10.1007/s12186-011-9058-5

Latan, H., & Noonan, R. (Eds.). (2017). *Partial least squares path modeling*. Springer.

Leicher, V., Mulder, R. H., & Bauer, J. (2013). Learning from errors at work: A replication study in elder care nursing. *Vocations and Learning, 6*(2), 207–220. https://doi.org/10.1007/s12186-012-9090-0

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Physica-Verlag HD.

Maurer, T. J., Weiss, E. M., & Barbeite, F. G. (2003). A model of involvement in work-related learning and development activity: The effects of individual, situational, motivational, and age variables. *Journal of Applied Psychology, 88*(4), 707–724. https://doi.org/10.1037/0021-9010.88.4.707

Rai, A., Goodhue, D. L., Henseler, J., & Thompson, R. (2013). To PLS or not to PLS: That is the question. *AMCIS 2013 Proceedings*. Retrieved from http://aisel.aisnet.org/amcis2013/Panels/PanelSubmissions/2

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing, 26*(4), 332–344. https://doi.org/10.1016/j.ijresmar.2009.08.001

Ridgon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal, 34*(6), 598–605. https://doi.org/10.1016/j.emj.2016.05.006

Ringle, C. M., Sarstedt, M., Schlittgen, R., & Taylor, C. R. (2013). PLS path modeling and evolutionary segmentation. *Journal of Business Research, 66*, 1318–1324.

Rönkko, M. (2014). The effects of chance correlation on partial least squares path modeling. *Organizational Research Methods, 17*(2), 164–181. https://doi.org/10.1177/1094428114525667

Sanchez, G. (2013). *PLS path modeling with R*. http://gastonsanchez.com/PLS_Path_Modeling_with_R.pdf

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010. https://doi.org/10.1016/j.jbusres.2016.06.007

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications to development research* (pp. 399–419). SAGE.

Shmueli, G., Ray, S., Valasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049

Tenenhaus, M., Amato, S., & Vinzi, V. E. (2004). A global goodness-of-fit index for PLS structural equation modeling. *Proceedings of the XLII SIS Scientific Meeting*, 739–742.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis, 48*(1), 159–205. https://doi.org/10.1016/j.csda.2004.03.005
Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*, 11–36. https://doi.org/10.1007/s12186-012-9091-z
Vaughan, K. (2008). *Workplace learning: A literature review*. NZCER Press.
Wold, H. O. A. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307–357). Academic.
Zwaan, R., Etz, A., Lucas, R., & Donnellan, M. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*, E120. https://doi.org/10.1017/S0140525X17001972

Michael Goller (PhD) is a research associate at the Institute of Educational Science at Paderborn University. Between October 2018 and March 2020, he served as visiting professor at the University of Bamberg and between April and September 2021 as a visiting professor at the University of Jena. His current research interests comprise expertise development and workplace learning in general, including in the particular context of digitalisation, as well as learning and development of students at the boundary of institutionalised and less institutionalised learning environments such as internships or particular learning arrangements in VET.

Frederic Hilkenmeier (PhD) studied psychology at the University of Hamburg and obtained his PhD from the Paderborn University in 2012. He is currently working as a lecturer and researcher in the psychology department at Fresenius University of Applied Sciences. In his classes, he mostly teaches psychological research methods, whereas his research deals with the impact of organizational context factors as well as individual motivational factors on learning and performance.

Chapter 13
Participant's Video Annotations as a Database to Measure Professional Development

Bianca Steffen and Maikki Pouta

Abstract Correct noticing and interpreting of important cues are among the few measurable criteria of professional performance. Elaborated expertise knowledge structures are opaque and, therefore, can hardly be measured. The rise of studies in this research field is due to new opportunities offered by modern technologies to grasp more subtle elements of professional development (e.g., tools to measure skin conductance response and eye tracking, and to take video-annotations). This chapter aims to present video annotations as a database for analysing professional learning and development. Video annotation tools (e.g., ELAN) can be used to operationalise participant's noticing of important cues in a video in the form of annotations. These annotations can be analysed to answer quantitative, qualitative, or mixed-methods research questions. Hence, video annotation technique (VAT) is also a suitable technique for data triangulation. This chapter emphasises that the videos are only a vignette to evoke participants' knowledge, unlike in the other methods, which rely on the video technique.

This chapter focuses on the use of video annotations as a data gathering technique. The first part describes participants' video annotations as material for analysis and the theoretical and practical benefits. In the second part, an example study will illustrate the application of video annotations as a data-gathering technique and research data in a methodological triangulation.

Keywords Video annotation · Data gathering · Methodological triangulation

B. Steffen (✉)
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: bianca.steffen@uni-paderborn.de

M. Pouta
Faculty of Education, University of Turku, Turku, Finland
e-mail: maikki.pouta@utu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_13

13.1 Introduction

Catching professionals' opaque and complex knowledge structures is a challenge for research. To explore professionals' development and expertise, videos have been widely used in research (Bates, 2014; Endlsey, 1988). Videos provide important contextual material for the situation to be explored. Originally used as a tool of analysis, video annotation technique (VAT) provides interesting opportunities for gathering data about a participant's process of thought and knowledge use triggered by the video (Rich & Trip, 2011). VAT describes the overall technique when video annotations, participants' written notes, are collected as data with video annotation tools i.e. software.

To differentiate VAT from video-based research methods, it is important to stress that the video only provides a stimulus for the participants to evoke their process of thought and knowledge use; consequently, their annotations are the database for further analysis (Pérez-Torregrosa et al., 2017). Video annotation technique offers opportunities to operationalise and measure participants' perspectives, visions, and thoughts connected to the concrete scene within a video. As a flexible data gathering technique, annotations suit several research settings and data gathering procedures and benefit methodological triangulation Furthermore, new technologies enable expanding video annotations with, for example, tools to measure skin conductance response or eye tracking.

This chapter focuses on the use of video annotations as a data gathering technique. The chapter first describes participant's video annotations as material for analysis, focusing on the conceptual perspective and the chances it offers to overcome challenges in workplace learning research. The theoretical benefits and practical implementations will be described and discussed following the three main steps of the research process. In the second part, an example study will illustrate the application of video annotations as a data gathering technique and research data in a methodological triangulation of teachers' professional vision.

13.2 Video Annotation

One focus in workplace learning research is to understand professional's learning process. Two aims are being followed in this research tradition: first, researchers are interested in understanding professional learning and development; second, findings can be made accessible to novices to support their professional development by helping them understand expert's perceptions of situations, their knowledge use, and their decision-making. The main empirical challenge in workplace learning research is the tacit nature of learning and knowledge. Professional development mainly occurs on the job, often informal and opaque, and in various forms (Billett et al., 2018; Goller & Paloniemi, 2017; Gruber, 1999). It follows that the complex knowledge structures due to these learning processes can hardly be measured, as participants struggle with their recollection and ability to abstractly express their line of thought (Harteis, 2014). This creates a challenge for data gathering with, for example, interviews or questionnaires, which many methodological approaches try to overcome by making formerly obscure knowledge structures accessible and quantifiable. Video annotations are a promising database for overcoming these obstacles (Pérez-Torregrosa et al., 2017) as participants do not have to consciously recall their knowledge but describe their knowledge or argue for decisions based on their knowledge.

One main factor of successful professional performance is visual recognition and cue-based decision-making within complex situations. Experts are known to be more knowledge-driven and focused on their observations (Wolff et al., 2016) and have continuity and diversity in their interpretation (Wolff et al., 2015). Experienced professionals focus on the current situation and perceive "the elements in the environment, within a volume of time and space, [comprehend] their meaning and [project] their status in the near future" (Endsley, 1988, p. 97). This focus, especially the meaning it has to professionals, can be measured by video annotations, which enable connecting participant's thoughts and impressions with the scene in the video on time. Video annotation has been used for teaching professionals and evoking reflection for learners in workplace settings (McFadden et al., 2014; Pérez-Torregrosa et al., 2017; Stancliffe, 2019).

The increasing ubiquity of videos and ease of producing videos have induced a rise in add-on functions and further uses, such as video annotation tools. Video annotation tools were originally developed to allow researchers to apply video annotation technique and analyse videos by annotating scenes of interest, such as group discussions, social interactions, or complex social situations (e.g., classroom interaction; Rich & Trip, 2011). The main benefit of annotating is to easily mark and categorise situations on the video. This chapter presents video annotations as a data gathering technique. In such settings, the participant watches a specific video (e.g., a specific workplace situation or a video of their own performance) and uses the software to annotate the video following the researcher's instructions. Based on this research's aims, the participant can be instructed to annotate, for example, important cues, relevant noticed information, and assumptions of how the situation might develop. Various forms for annotations, such as open-ended, a set of prepared phrases to be matched with the video or closed items to indicate, for example, agreement, can be used. Video annotations could be helpful in measuring the knowledge structures of a participant and how that knowledge is applied to the information shown in the videos. The participant annotates thoughts and impressions into the programme. Afterwards, researchers analyse the quantity or quality of these annotations.

Video annotation tools are usually software programmes, but some can also be used as browser based. Video annotation software is compared in many studies along different criteria, such as programming language, the target group, the interface, options for collaboration, open-source, charge, option for multi-media, and the form of annotations (Gaur et al., 2018; Rich & Trip, 2011). In annotation software, the user can annotate, ergo mark timepoints, and make notes about specific parts of the material. Often, annotation software uses multiple layers of notes on the same time point. This enables the use of several instructions and structured annotation procedures in data gathering and categorising during the analysis phase. A detailed description of one software tool will be elaborated in the “Practical Application” section.

13.3 Video Annotation as a Data Collecting Method

Besides all the obstacles connected with the practical implication of gathering participant’s annotations, many practical possibilities also arise from the innovative character of using video annotations as a database. The following section elaborates on the chances and benefits of the application of video annotations as a database for workplace learning research. Hereinafter, the practical and theoretical advantages of this technique are illustrated.

13.3.1 Comparing Video Annotation with Other Methods of Data Collection

Video annotation is a flexible data-gathering procedure that can be leashed to provide suitable data for different research settings. The tiers can be used to guide participants’ attention in the annotation process. With suitable modifications, the tier data provided by VAT can be either more survey-like and controlled or raw text from open data. Therefore, VAT is also very suitable as part of methodological triangulation when attached to data gathering methods (by Pouta et al., 2020). The benefits of VAT compared to other datasets are next presented to illustrate the possibilities of VAT in triangulation or as its own data-gathering technique.

Video annotation is better suited to overcoming existing empirical challenges of more traditional research methods, such as *think-aloud, interviews* (Lemmetty & Collin, Chap. 18, in this book), *or* cued retrospective reporting (CRR) (Paloniemi et al., Chap. 5, in this book). Advantages of VAT are similar to CRR as they offer a stimulus, and the participants do not have to abstractly recall situations or images, but they can connect their visual perception with the concepts of their profession (Schmidt et al., 1990). Hence, CRR combines concurrent reporting with retrospective recall (van Gog et al., 2005) similar to VAT. In VAT participants answers are ready in text form and no transcriptions are needed like in interview methods. With its efficiency VAT comes close to survey method as data collecting technique as in VAT several participants can watch the same video simultaneously but individually. As described before, in VAT it is possible to create timepoints ready for the participants to the software and let them choose from several tiers and items that suit the presented part of the video (Max Planck Institute, 2021). Hence it enables more straightforward qualitative analysis or even quantitative data analysis (e.g. Seidel et al., 2011). In contrast to *field observation,* VAT narrows workplace situations down to the information presented in the video, which enables keeping the amount of data appropriate by framing the data collection in line with the research questions. Additionally, VAT is a dynamic technique of data collection, where participants actively share their thoughts, knowledge, or reflections regarding a video and therefore it is motivating technique to include as data collection technique in professional development programs. As described, VAT suits for research projects that aim to add efficiency of survey method to data collection without losing possibilities of the qualitative in-depth analysis that rich interview or observation data offers.

Modern technical devices allow researchers to measure a participant's visual perceptions, such as attention, noticing, and pattern recognition. Eye-tracking devices measure participants' eye movement without additional information to interpret this data regarding professional learning and knowledge. The same applies to the *measurement of the skin conductance response*. Both methods also collect a huge amount of complex data that is challenging to interpret. These more technical-based techniques can be supplemented with think-aloud interviews to gain more qualitative information about the expert's knowledge structure and what they saw and noticed (Chang et al., 2018; Menshikova et al., 2014; Pouta et al., 2020; Rodeghero et al., 2015). These methods can also be attached with VAT, as the eye-tracking and skin conductance response methods provide information about the annotation task situation, for example, when exploring professionals' effects stimulated by video and their interpretations as annotations. Eye-movement videos can also be used as a basis for annotations since the eye-movement data is always attached to the video data. Modelling of experts' eye movements has recently been a more studied field of research and used in higher education and in different domains (e.g., Gegenfurtner et al., 2017).

13.3.2 Video Annotation from the Perspective of Quality Criteria

Video annotations, as a database, also offer some theoretical advantages from the perspective of quality criteria for empirical social research as will be discussed in the following. In particular, the phase of data collection benefits the research quality in quantitative and qualitative research projects. At first, *economy* as a main criterion of overall research quality is met by this technique of data collection. If data collection takes place in a remote setting, it is more practicable and time saving. This criterion is also fulfilled because the annotations are the data for further analysis, and no audio or video record of the data gathering process, more specifically, a transcript of the record, is needed. Not documenting personal data such as voice and image also reduces specifications of data protection.

The following section discusses the advantage of video annotation for the main quality criteria of quantitative research. Focussing on data *validity* (Kuckartz, 2014; Völcker, 2019), video annotations allow for an immediate connection between cause (e.g., video stimulus) and reaction (e.g., the annotation made by participants). It follows that the stimulus can be reduced to theoretically relevant cues, and the participant's annotations could be more valid connected to the research interest. One main threat to validity is that gathering annotations from participants is mediated by their ability to use the technology (the hardware and the software) and to express their knowledge, line of thoughts, or reflection. A limitation of this thread could lie in closed items, which only ask the participant to express agreement on a scale or to choose from a prepared set of words or phrases. Pictograms or voice recordings can also be used as forms of annotations. *Reliability* (Kuckartz, 2014; Völcker, 2019) can also be increased by a default set of annotations that the participant is allowed to use (instead of free text). This criterion is also met because, in contrast to data collection in real workplace situations, videos offer a limited amount of information to be processed and therefore reduces confounders. VAT also creates a convenient platform for double coding, which aims to increase intercoder reliability, as the qualitative analysis of the annotations can be done to the same software by two researchers and then compared. The criterion of *objectivity* (Kuckartz, 2014; Völcker, 2019) can also be met more easily because, unlike in personal methods of data collection, there is no need for personal interaction between the researcher and participant. A remote setting avoids all interactions between researcher and participant. It follows that all obstacles to social interactions (e.g., self-portrayal and social desirability) can be reduced. The participants can annotate the video and answer additional items by themselves and remotely if the instructions are sufficient. This would also enable data collection in the complete absence of researchers. In a more reactive and interactional setting, the researcher could set up the technology and offer an introduction on how to fulfil the task but only answers clarifying questions. This also reduces the chance of gaining findings that are influenced by participant's tendency to give socially desirable answers.

For qualitative research, *transferability* can be met because the context (e.g., time and place) in which the data are gathered can be described, which allows for a higher applicability with other participants (Kuckartz, 2014; Völcker, 2019). Video annotations are also highly comparable because all participants can be presented with the same stimulus (when the situation is controlled and does not include recall of participant's experiences), whereas, in interviews, the questions, intonation, and follow-up questions might differ. The strong theoretical framing of both the video vignette and the impulse, instruction, and task induces higher *consistency* (Kuckartz, 2014; Völcker, 2019) of the collected data. This means that the logical connection between the data, themes, and the process of data collection can be secured and documented. The *intersubjective comprehensibility* (Kuckartz, 2014; Völcker, 2019) is highly met because of the strong link between the video and the participant's reactions (e.g., annotation). The collected data is a direct record of participants' expressions, and no transcript moderates the reaction. Documentation of the process is also facilitated because the impulses are unaltered, unlike in more interactive formats in which wordings and phrases may differ.

After laying out these theoretical grounded arguments favouring participant's video annotations as material for analysis within empirical research projects, the following chapter elaborates on the practical application of this technique and the challenges related to it.

13.4 Practical Application of Video Annotation

Social science research looks back on a long tradition of working with videos (Bates, 2014; Haw & Hadfield, 2011; Jewitt, 2012). Regarding video annotations as a database for analysis, the researcher provides the videos as vignettes, which the participants interpret. Selection of suitable videos for the annotation is essential and should match the research aim, practical possibilities of the data gathering procedure, and a participant's vigilance (Jewitt, 2012). Videos offer multiple possibilities for different kinds of setting – video could show routine tasks (e.g. when studying teachers: a pupil's presentation), situations that are more complex (e.g. classroom situations), or exceptionally challenging tasks (e.g., school shooting). It is also possible to use videos as shared material or as a group task. The perspective of the video (is the video stable or mobile) is crucial since it frames and limits a participant's attention (Rich & Trip, 2011). Participant-annotator relationship with the video scene setting requires consideration, as it affects the reflective approach of the task. For a more detailed description of video-based interaction analysis regarding workplace learning and professional development, see Filliettaz et al. in this handbook. The ethical use of videos and information security throughout the research process requires consideration when using highly sensitive data, such as video data, where people can be recognised. This is especially important when considering the use of already-existing video data as a basis for annotations. Especially in browser-based data gathering settings, the information security of the gathered data alongside the presented video should be highly emphasised and approved by the institution's ethical board.

This chapter introduces Max Planck Institute's ELAN[1] as an example of a tool to be used for video annotation technique because it is open source, user friendly, and offers several relevant features, which enable flexible use of the tool in different kinds of data gathering settings. According to a criteria for video annotation tools, ELAN's most relevant features are that (a) up to four files can be uploaded, (b) multiple tiers can be used, (c) it builds on broadly used media frameworks (like Windows Media Player, QuickTime or VLC), (d) it is available for Windows, macOS, and Linux, (e) it is open source written in Java programming language, and

[1] More detailed information about the software as well as the user manual can be found on the homepage: https://www.mpi.nl/corpus/html/elan/

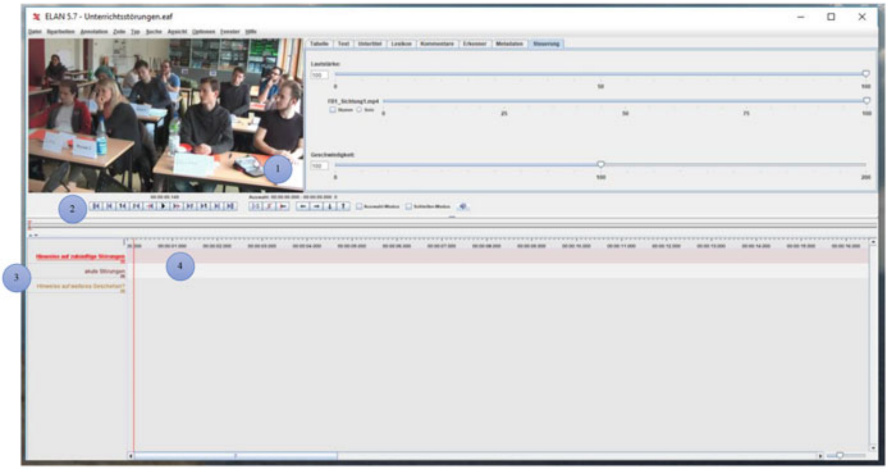

Fig. 13.1 User View of ELAN Annotation Software Window. (Own screenshot with participants' permission)

(f) it is a free of cost software for an unlimited number of annotations. Other noteworthy programmes are described by Cruciani et al. (2011), Hosack (2010), and Salisbury et al. (2016).

Figure 13.1 shows a screenshot of the ELAN software. On the left corner of the screen plays the video of the situation to be coded (1). In ELAN, it is also possible to add another video (e.g., from another perspective), and these can be played simultaneously. The buttons for start, stop, and rewind are below the video (2). Tiers for annotation are in the bottom part (3). These are user-generated and can be set up in advance by the researcher based on the research aims. Therefore, the tier descriptions (written on the left side with different colours) can include, for example, questions where the participants should answer, or descriptions they should pick from different tiers. The video's time stamp is recorded in the annotations (4). It follows that the annotations are aligned with the time of the video and can easily be matched with the corresponding scene. The layout can be modified, and other tools can be attached regarding the research question and the sample.

The research process includes many factors to be considered regarding the particular research settings and research question(s) alongside the practical implementation possibilities of the data gathering procedure. These considerations are described next, aligning with the steps of the research structure. *Planning* includes all steps that need to be done in advance of data collection and all decisions that have to be made in advance. The whole process should be considered, including the analysis. Regarding VAT, an intense planning and piloting of the data gathering is advised because some pitfalls can be avoided. In the phase of *data collection*, the main tasks are the setup of the technical equipment, the introduction to the programme, and the instruction for the annotation and the flawless execution of the actual data collection. By now, it should become obvious how fail-safe and dependable the planning phase was conducted. *Analysis* of participant's video annotations should be considered throughout the whole process. Implications will be kept general due to the multiple options of common data analysis techniques in which the annotations can be analysed, such as other databases.

13.4.1 Planning

The decision for a research method follows the formulation of the research question (s). As a data-gathering technique, video annotations offer flexible possibilities for research settings; therefore, the following questions should be considered when structuring the data-gathering procedure for a research project.

- Which programme should be used for the annotation? Software differs by features, which affect the simplicity and user-friendliness of the programme alongside the data analysing possibilities. When choosing the software, data protection security should also be considered and investigated, especially with browser-supported programmes. For detailed overviews and comparisons of video annotation software, see Jewitt (2012), Gaur et al. (2018), and Rich and Trip (2011).
- What kinds of instructions do the participants need? The level of detail should be considered related to the research aim and adaptivity of the software for the users (McFadden et al., 2014). Also, the form of the instructions should be considered (e.g., written instructions or demonstration video).
- What training do the participants need? It should be considered based on the accuracy of the needed annotations and participant groups' needs if the hands-on training should be implemented before the data gathering or if, for example, a detailed demonstration video of the annotation procedure is enough to guide the participants in the annotation process (McFadden et al., 2014).
- How much time do the participants have for annotating? Experience show the nature of the research should be considered (e.g., time-pressured setting) alongside the practical limitations of the data gathering procedure and participants' vigilance.
- How often should the participants watch the video? Are they allowed to rewind scenes? Watching the video only once to get data about the participant's first impression vs. watching the entire video or specific scenes several times to let participants process information more elaborately.
- What kinds of tiers should be used? The tiers can be used to guide participants' attention in the annotation process. Hence, the research questions should be considered when forming tier descriptions. However, the tiers should be described in an annotation-friendly and accessible way for the participants. Therefore, the data for particular research questions can also be gathered with several tier descriptions or questions. It is also possible to create more standardised settings for video analysis with survey-kind of ready categories as

tiers from which the participants select when annotating instead of free writing (Gaur et al., 2018; Rich & Trip, 2011).

- What form of annotation should be used? The software offers different opportunities for annotating: such as free text, prepared phrases, or metric data (Cruciani et al., 2011; Gaur et al., 2018; Hosack, 2010). The analysis method should be considered when deciding on the form for annotations.
- Which timepoints should be annotated? Depending on the research design, it is also possible to set the relevant episodes and timepoints for the video data instead of instructing the participants to mark the timepoints themselves as part of annotating. In this case, the annotator selects one or several tiers or categories that the point of the video presents in their opinion (Rich & Trip, 2011).
- How should the data be analysed? Video annotations create possibilities for many kinds of analysis settings regarding qualitative, quantitative, and mixed-method designs (Gaur et al., 2018). The type of annotation and formulation of the tier descriptions and instructions affect the analysis process; therefore, these should be formed with the idea of the analysis procedure.
- What kind of video should be watched? It makes a difference not only by the research settings but also regarding the complexity and focus of the annotation task, whether the video is of the participants themselves or of someone else (Seidel & Stürmer, 2014). For the criteria for the construction of vignettes, see Anselmann and Mulder in this handbook. The decision of an appropriate video is important because it is the crucial point of the data collection (Rich & Trip, 2011).
- How long should the video be? The length of the video should be considered as annotating multiplies the time used for video watching (Rich & Trip, 2011). Piloting the procedure is crucial for defining the suitable length of the video and deciding the length, considering the frames of the data gathering procedure in practice and participants' vigilance. The length of the video is also a main factor influencing the duration of one data collection session (besides the smooth or faulty operation of the hardware and software alongside the participant's technical and typing skills). The data collection should be as time-economic as possible on behalf of the participants.
- What should be considered in preparation? It is essential to pilot the data-gathering plan carefully to detect the possible practical and technical challenges of the procedure before the actual data gathering (Rich & Trip, 2011). In our projects taking time and making notes about the data-gathering procedure offer golden information about the settings that can be modified. Participants' experiences are also an important source of information about, for example, their vigilance, motivation, and understanding of the task. Furthermore, piloting tests the technical arrangements of the data gathering, such as training in the usage of annotation software, creating annotations, and saving the data. Also, experience shows the importance of testing the capacity of the equipment: how efficiently the computers used run the video in the software, how saving and backup of the data could be ensured during the data gathering, etc.

Some of the presented questions affect one another, so the researcher might go through some planning steps not chronologically but rather circularly until a promising and flawless procedure is set up. After going through an intense phase of planning and piloting the gathering and analysis of the data, data gathering can be initiated.

13.4.2 Data Collection

After the instructions and possible training phase for the participant, the actual data collection starts when the participant watches and annotates the video within the programme. During this phase, the researcher's presence depends on the data collection setting (whether the data is collected remotely or in a live situation; Gaur et al., 2018).

In a remote setting, the participants receive (written or recorded) instructions from the researcher on how to use the programme and what to annotate in the video. A browser-based software is advised to ensure accessibility to the annotating. One main threat to remote data gathering is that the participants cannot ask clarifying questions. With technical approaches, this is possible to overcome if the data gathering settings (number of participants, time, etc.) allow the researcher to be remotely reachable for technical difficulty via chat etc. High emphasis on clear instructions and automatic saving of browser-based software is essential for remote data gathering in our experience.

In a physical setting, the researcher sets up the technical device and instructs the participant on how to use the programme, how to make annotations, and what annotations to make. For simultaneous data collection, several technical devices with video annotation software (and probably as well headphones) are concurrently needed. Structuring the procedure is important both for successful data gathering and for researcher's nerves. In simultaneous data collection, using competent assistants would be needed to ensure the procedure fluency. Afterwards, saving the supplied data on all devices should also be ensured, and it is recommended to use automatic saving, backup, or reminders as much as possible. Furthermore, consecutive data collections need only one technical device with the needed programme installed, and problems can be solved directly because the researcher is in a bilateral setting. This way is more time consuming but less fault-prone.

13.4.3 Analysis

After data collection, the annotations made by the participants are the material the researcher has to analyse. One benefit of video annotation is that it offers relatively clean data to be analysed. As such, video annotations are beneficial compared to think-aloud material, and it can even be close to survey-kind straightforward answers if the tiers were defined as options where to choose from when annotating. However, researchers should be prepared to use some time for data cleaning regarding possible difficulties that may occur during data gathering.

It is possible to use the same annotation programme as a researcher to analyse the annotations of the participants (Gaur et al., 2018). For example, based on the qualitative analysis coding scheme, new tiers can be defined to categorise annotations. It is also possible to export the annotations to another programme. If the data from all participants are compiled in one file, time stamps should also be transferred. For qualitative analysis, annotations may have to be translated into metric values. The analysis used always depends on the approach suitable for the particular research questions and aims. However, video annotation data offer multiple possibilities for data analysis, such as quantitative, qualitative, and mixed methods designs (Rich & Trip, 2011).

Finally, the annotations made by the participants can be analysed using qualitative, quantitative, or mixed methods (Gaur et al., 2018). Qualitative methods may try to answer questions such as "How do subjects on different levels of professional development (e.g., novices, intermediates, and experts) interpret the same situations?" or "How theoretically founded is the subject's argumentation?" and could be analysed using thematic content analysis or objective hermeneutics. When analysing the data using categories, these categories can be defined by relevant scenes. They can be derived either deductively (e.g., from theoretical assumptions) or inductively (e.g., a high number of annotations could indicate the relevance of a scene). However, quantitative analysis would be appropriate for questions such as "How many critical incidents did the subjects notice?" or "How often do subjects argue with theoretical knowledge or experiential knowledge in specific situations?" using methods of descriptive statistics or comparison of variances or mean values.

13.4.4 Practical Possibilities

After laying out all the challenges for video annotation technique in the preceding chapter, the practical opportunities offered by this technique of data collection shall be elaborated here. Adding to the challenges presented above derived from the practical application of video annotation as a form of data collection, the main practical opportunities are presented here.

- Video annotations can be collected remotely because some tools are browser-based (participants do not have to install a programme) and can be embedded with (written or recorded) instructions (Gaur et al., 2018). If these are sufficient, the researcher must be present during data collection. Data can also be collected completely anonymously because the researcher does not get to know the participants in person. This is very economical for larger projects with a large sample.
- Researchers should decide on an easy-to-use tool for the specific sample (for an overview see Gaur et al., 2018). An easy interface is user friendly and will keep the obstacles low for participants who are not technically versed.

- Different types of annotations make it possible to work with participants who cannot elaborately formulate their answers or type them (e.g., speech-impaired or handicapped people, children). Icons and pictograms (e.g., emojis) could be used instead of written words to gather information about their perspective on a situation. Some tools (e.g., Video Traces, VoiceThread) have an option for recorded voice-over, meaning that the participants do not have to write their responses, but they can explain them verbally. Others offer the upload of images (e.g., VAST, Video Papers); additionally, others also accept drawings in the video. Researchers should carefully consider these options following the planned form of analysis. Rich and Trip (2011) concluded, "Thus the ease of adding comments may conflict with the ease of analysing them later." (Rich & Trip, 2011, p. 17). Written impulses and phrases for the annotation also allow for easy, comparable, and culturally sensitive translation (for an overview see Gaur et al., 2018).
- Video annotations also offer the chance for easy translation in international projects. Videos and impulses can be translated more easily, and the cultural context can be documented in more detail than in interviews.
- Some tools (e.g., ELAN) have the option of using several tiers and layers for annotation. It follows that the participants can fulfil different tasks or items for one video. Nevertheless, researchers should carefully consider the extent to which this is practicable and economic regarding time investment. As few annotations as needed should be collected considering the limited cognitive capacities of participants that may be overburdened. Videos offer only a partial perspective of reality. Some tools (e.g., ELAN) provide the option of concurrently showing several videos. Hence, the video vignettes can be more complex.
- There is no need for a transcript. The collected data are almost directly suited for different methods of qualitative (e.g., content analysis), quantitative (e.g., descriptive statistics), and mixed-method analysis. Several tools already offer statistical analysis (e.g., IRIS contains statistical analysis tools, and The Observer XT calculates interrater reliability).
- Furthermore, it can be combined with different means of data collection. Modern and technical methods (e.g., eye tracking, measure of skin conductance response) can be supplemented. Yonemoto (2012) used AR technology in expansion to video annotation. In the example study presented below, Pouta et al. (2020) used mobile eye-tracking techniques combined with cued retrospective reporting. However, traditional methods (e.g., observations, questionnaires) can also be used for data triangulation within one project.

13.5 Example Study

The study "Student Teachers' and Experienced Teachers' Professional Vision of Students' Understanding of the Rational Number Concept" by Pouta et al. (2020) uses methodological triangulation to explore prospective and experienced teachers' knowledge-based reasoning stimulated by the video of their own teaching. This already published study illustrates the usage of teacher's annotations on videos of their own teaching to investigate their professional vision in classroom situations. This chapter focuses on the methodological steps in which video annotations are advantageous compared to oral interviews.

13.5.1 Framework

The study gathers data about teachers' professional vision to explore differences by expertise in teacher's noticing, knowledge-based reasoning and instructing students' understanding of the rational number concept. The rational number concept is considered challenging for both learning and teaching because it demands the conceptual change towards novel ways of thinking (Merenluoto & Lehtinen, 2004) and active inhibition of natural number bias (van Dooren et al., 2015). Therefore, settings in which the rational number concept is taught are promising situations for examining teachers' professional vision and instruction quality. Van Es and Sherin (2002) describe professional vision as a system of selective attention and knowledge-based reasoning, whereas the latter includes description, explanation, and prediction (Seidel & Stürmer, 2014; Jacobs et al., 2010) of the teaching situation. One important factor that should be considered to predict successful deciding and acting within a challenging situation is knowledge and experience. Therefore, teachers with a broad basis of experiential knowledge will be more likely to make successful decisions. Professional vision is understood as a system bridging knowledge and practice (Goodwin, 1994).

Former studies in this field have focused on teachers' attending and interpretation classroom situations (e.g., van Es & Sherin, 2002). Most of them were conducted using more classical methods (e.g., interviews), whereas others also used rather modern techniques of data gathering, especially the application of eye-tracking techniques that are forging ahead in professional vision research field (Shvarts, 2018). The study investigated the differences in experienced teachers' and student teachers' professional vision with cued retrospective reporting (CRR) based on video that each participant recorded with mobile eye-tracking glasses of their own teaching situation. Pouta et al. (2020) try to close this research gap by answering the following questions:

1. How do experienced teachers and student teachers differ in noticing students' mathematical thinking?
2. How do experienced teachers and student teachers differ in interpreting students' understanding of fractions?

In the illustrated study, teachers' verbal data from cued retrospective reporting (CRR) was used to collect data for the analysis. In the following section, the conductance of the study is elaborated by using VAT as data collection technique instead of CRR that was used in the original study. The underlying theoretical assumptions and practical requirements are sufficiently similar for video annotations to be used as a form of data collection in this research setting.

13.5.2 Data Gathering

Pouta et al. (2020) compare the professional vision of four experienced teachers and four student teachers. The student teachers were in their final practical training period, while the experienced teachers had between 9 and 21 years' experience. In the first step, the participants were made familiar with the mobile eye-tracking glasses. Then they conducted one 45-min mathematical lesson about the rational number concept and recorded it from the first perspective with mobile eye-tracking glasses while they were teaching. In all participants gathered 5 h of video data. Last, the CRR was conducted right after the lesson using their own recording without eye movements. In CRR participants watched the video of their own and were instructed to think aloud two questions: (1) what they attended to, and (2) how they instructed the students based on their noticing. When using VAT instead of oral reports, these two categories would be in the bottom part as tiers for annotations – maybe adding a residual category for other interesting cues that the teachers would like to share but that do not fit into one of the prepared categories.

In the original study (Pouta et al., 2020), teachers were allowed to talk and pause the video by themselves if needed. When gathering written annotations, stopping the video by the participants would of course be obligatory to assure that the participants have sufficient time to express their thoughts in written words. In both settings, the researcher would not participate by commenting or asking questions. In the illustrated study, the CRR lasted for 54 min on average. This duration would be longer if the teachers would type their answers. Ethical clearance was obtained in this study, and it would also be obligatory using written annotations.

13.5.3 Data Analysis

In the original study, the oral data of the CRR was transcribed before the analysis, which is not needed when using VAT. One challenge in the study was to present the oral reports of CRR in line with the events of video. Therefore, the using VAT would be beneficial as the annotations are appointed to specific timepoint within the video.

The analysis of the CRR data was a multi-step process that structured the data by selecting the relevant episodes and analysing teachers noticing and their interpretations of the teaching situation of the episodes. The main database for analysis was the transcripts of CRR and the videos were used as a secondary source to offer contextual information on the teachers' references. First, episodes including references to mathematics-related content was defined and separated from episodes referring other content based on verbal data from CRR. Next, the episodes including references to students' understanding of fractions were identified and separated from episodes including all other notions of students' mathematical thinking. Last, teachers' verbalizations from CRR on the selected fraction-related teaching episodes were categorised on three different levels following van Es and Sherin (2008): descriptive (teacher describes students' actions), evaluative (teacher evaluates task difficulty), and interpretive (teacher notices specific fraction-related student understanding). 25% of the data set was selected for double coding by two researchers. The interrater reliability was good, between 79% and 86%, and differences were solved through discussion.

If the data was gathered with VAT the episodes to annotate would be readily marked to the data sparing researchers from creating episodes based on verbal and video data. Thus, the start and ending point of the episodes would be needed to check for technical errors of the annotators to keep the content and situation on video in line. Also, the same annotation software could be used for analysis as the categories of each analysing phase could be added as new tiers. Many annotation software also includes export of descriptive data and in addition inbuilt statistical features to explore the data quantitatively. Some annotation software, like Observer XT, includes tools for inter-rater reliability analysis.

13.5.4 Results

The results of Pouta et al. (2020) study indicated that no significant difference could be found in attending to mathematical and non-mathematical features between experienced teachers and student-teachers ($\chi 2$ (1, N = 152) = 1476). Similarly, there was no difference in experienced teachers' and student teachers' attending to other mathematical content-related fraction-related episodes ($\chi 2$ (1, N = 86) = 3285). Align with previous studies, both experienced teachers and student teachers focused on other features than students' mathematical thinking too on the teaching situation. Interestingly, the analysis of the interpretations revealed that student teachers' notions were mostly evaluative (63.6%) and interpretative (27.3%) whereas experienced teachers' notions were mostly descriptive (44.4%) and evaluative (41.7%) with a significant difference in the interpretation levels of experienced teachers and student teachers ($\chi 2$ (2, N = 69) = 10.93, $p < 0.01$). The results indicate that despite lack of teaching experience, student teachers were surprisingly capable of noticing students mathematical thinking and they were more in-depth with their interpretations than the experienced teachers. This could be because of student teachers' routine from their studies to verbalize their observation and knowledge. Also, based on the CRR data, it is possible that on some occasions experienced teachers' knowledge of individual students might cover their specific fraction-related interpretations.

In all, the study used an advanced triangulation of CRR and eye tracking, i.e. first-person perspective recording (Pouta et al., 2020). The low standardisation of teaching conditions and other factors as well as small sample size limits generalization.

However, the authenticity of the teaching situation is fostered by the minor interference during the data-gathering. If the study would have been conducted with VAT, the bigger sample size might have been possible if the participants could have done the annotations themselves e.g., with a browser-based annotation tool. Then, ethical considerations should be even more specific. Also, the moment for participants' annotations should be pre-set to prevent the differences in recall time.

13.6 Conclusions

As described in this chapter, using VAT can be suitable for varying research questions and interests as well as many kinds of research settings to overcome the shortages of more traditional research techniques (e.g., surveys, interviews, or observations). Hence, as part of methodological triangulation, VAT can offer possibilities for quantitative research methods aside of in-depth analysis of the qualitative data. As a research technique VAT is flexible because of its adjustability to different research designs and purposes. The video used as a basis of the annotations and definitions of the tiers in VAT frame the research design towards the aim of the study.

As described, VAT makes the data collection more efficient because it can be used simultaneously but independently for many participants. Furthermore, VAT can be considered a motivating data collection tool as the participants have an active role in it. Also, VAT's independence from time and place makes it easily employable in the context of workplace learning. Hence, VAT suits well for professional development programs (cf. e.g., Seidel et al., 2011). As the name implies, videos are in the heart of VAT. The stakes to work with video stimulus in the observation tasks are relatively low because participants and researchers are familiar with these media in their everyday (working) life. Therefore, researchers should feel encouraged to imply VAT to their repertoire of methods.

Of course, like any other research technique, VAT has some shortcomings. As a flexible tool the use of it also requires a lot of careful planning and clear instruction. Defining the task, the tier type, and the questions for them as well as preparing the video data and editing it for suitable use include many steps that can guide the study design towards research aims or away from it. These shortcomings, however, can be overcome with precise planning and careful conduction not to forget methodological triangulation. To help creating VAT research design and collect the data successfully, some practical tips for different phases of research process are elaborated in this chapter.

First and foremost, VAT's strength is in bringing efficacy to data collection of qualitative data, that is usually found to be time consuming to collect. As seen in our practical example based on Pouta et al. (2020), VAT offers a less time-consuming option for data collection as qualitative data could be gathered from several participants simultaneously instead of individual (CRR) interviews. Also, in structured interviews, the un-natural interaction caused by structured questions can make the situation pressured and uncomfortable for the participants despite researcher's attempts to ease the interaction. Furthermore, as VAT provides participants verbalizations straight as text and in the same software system as it can be analysed, the analysing phase of the research is also more efficient as there is no need for transcription or inserting data to the analysing system. Hence, the user-friendliness of the technique reaches not only to the participants but also to the researchers.

Participant's video annotations offer a new database to measure professional development. Scientists already know the programmes from their own research and can introduce it to possible participants. The advantages and challenges (in comparison to more traditional research methodologies) are discussed above and based on these we would like to encourage fellow researchers and students to adapt this more innovative and advanced form of data collection. Concerning the rise of mix-methods studies in social science in general and professional development in particular the ease to triangulate gathered data is a major advance for empirical research. Future research can make use of the omnipresence of videos when investigating professional vision and situational awareness with large samples and without having to interfere in the professional field.

References

Bates, C. (Ed.). (2014). *Video methods. Social science research in motion*. Routledge.

Billett, S., Harteis, C., & Gruber, H. (2018). Developing occupational expertise through work activities and interactions. In K. A. Ericsson, N. H. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook on expertise and expert performance* (11th ed., pp. 105–126). Cambridge University Press.

Chang, C.-F., Gröschner, A., Hall, N. C., Alles, M., & Seidel, T. (2018). Exploring teachers' emotions via nonverbal behavior during video-based teacher professional development. *AERA Open, 4*(4), 233285841881985. https://doi.org/10.1177/2332858418819851

Cruciani, F., Donnelly, M. P., Nugent, C. D., Parente, G., Paggetti, C., & Burns, W. (2011). DANTE: A video based annotation tool for smart environments. In G. Parr (Ed.), *Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering: Vol. 71. Sensor systems and software: Second international ICST conference; revised selected papers* (Vol. 57, pp. 179–188). Springer. https://doi.org/10.1007/978-3-642-23583-2_13

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society Annual Meeting, 32*(2), 97–101. https://doi.org/10.1177/154193128803200221

Gaur, E., Saxena, V., & Singh, S. K. (2018). Video annotation tools: A review. In *International conference on advances 10/12/2018 – 10/13/2018* (pp. 911–914). https://doi.org/10.1109/ICACCCN.2018.8748669

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). 1. Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education, 113*, 212–225.

Goller, M., & Paloniemi, S. (Eds.). (2017). *Agency at work: An agentic perspective on professional learning and development*. Springer. https://doi.org/10.1007/978-3-319-60943-0

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Gruber, H. (1999). *Erfahrung als Grundlage kompetenten Handelns* [Experience as a foundation for competence]. Huber.
Harteis, C. (2014). Investigating intuition under the perspective of expertise: Experiences from two experimental studies. In M. Sinclair (Ed.), *Handbook of intuition research methodology* (pp. 116–129). Edward Elgar.
Haw, K., & Hadfield, M. (2011). *Video in social science research: Functions and forms*. Routledge. https://doi.org/10.4324/9780203839119
Hosack, B. (2010). VideoANT: Extending online video annotation beyond content delivery. *TechTrends, 54*(3), 45–49. https://doi.org/10.1007/s11528-010-0402-7
Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202.
Jewitt, C. (2012). *An introduction to using video for research*. NCRM.
Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice & using software*. SAGE. https://doi.org/10.4135/9781446288719
Max Planck Institute for PsycholinguisticsNijmegen. (2021). *ELAN: Linguistic Annotator*. The Language Archives. Retrieved from www.mpi.nl›corpus›manuals›manual-elan
McFadden, J., Ellis, J., Anwar, T., & Roehrig, G. (2014). Beginning science teachers' use of a digital video annotation tool to promote reflective practices. *Journal of Science Education and Technology, 23*(3), 458–470. https://doi.org/10.1007/s10956-013-9476-2
Menshikova, G., Kovalev, A., Klimova, O., Chernorizov, A., & Leonov, S. (2014). Testing the vestibular function development in junior figure skaters using the eye tracking technique. *Procedia-Social and Behavioral Sciences, 146*, 252–258. https://doi.org/10.1016/j.sbspro.2014.08.123
Merenluoto, K., & Lehtinen, E. (2004). Number concept and conceptual change. Towards a systemic model of the processes of change. *Learning and Instruction, 14*(5), 519–534. https://doi.org/10.1016/j.learninstruc.2004.06.016
Pérez-Torregrosa, A. B., Díaz-Martín, C., & Ibáñez-Cubillas, P. (2017). The use of video annotation tools in teacher training. *Procedia-Social and Behavioral Sciences, 237*, 458–464. https://doi.org/10.1016/j.sbspro.2017.02.090
Pouta, M., Lehtinen, E., & Palonen, T. (2020). Student teachers' and experienced teachers' professional vision of students' understanding of the rational number concept. *Educational Psychology Review, 33*(1), 109–128. https://doi.org/10.1007/s10648-020-09536-y
Rich, P. J., & Trip, T. (2011). Ten essential questions educators should ask when using video annotation tools. *TechTrends, 55*(6), 16–24. https://doi.org/10.1007/s11528-011-0537-1
Rodeghero, P., Liu, C., McBurney, P. W., & McMillan, C. (2015). An eye-tracking study of java programmers and application to source code summarization. *IEEE Transactions on Software Engineering, 41*(11), 1038–1054. https://doi.org/10.1109/TSE.2015.2442238
Salisbury, E., Stein, S., & Ramchurn, S. (2016). CrowdAR: A live video annotation tool for rapid mapping. *Procedia Engineering, 159*, 89–93. https://doi.org/10.1016/j.proeng.2016.08.069
Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine : Journal of the Association of American Medical Colleges, 65*(10), 611–621. https://doi.org/10.1097/00001888-199010000-00001
Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal, 51*(4), 739–771. https://doi.org/10.3102/0002831214531321
Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others*? Teaching and Teacher Education, 27*, 259–267.
Shvarts, A. (2018). Joint attention in resolving the ambiguity of different presentations: A dual eye-tracking study of the teaching-learning process. In N. Presmeg, L. Radford, W.-M. Roth, & G. Kadunz (Eds.), *ICME-13 monographs. Signs of signification: Semiotics in mathematics education research* (pp. 73–102). Springer.

Stancliffe, R. (2019). Training the analytical eye: Video annotation for dance. *Theatre, Dance and Performance Training, 10*(2), 273–288. https://doi.org/10.1080/19443927.2019.1610039

van Dooren, W., Lehtinen, E., & Verschaffel, L. (2015). Unraveling the gap between natural and rational numbers. *Learning and Instruction, 37*, 1–4. https://doi.org/10.1016/j.learninstruc.2015.01.001

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*, 571–596.

van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244–276. https://doi.org/10.1016/j.tate.2006.11.005

van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*(4), 237–244. https://doi.org/10.1037/1076-898X.11.4.237

Völcker, M. (2019). The quality of 'good' mixed methods research: Development and discussion of an orientation framework. In J. Lüdemann & A. Otto (Eds.), *Studien zur Schul- und Bildungsforschung: Vol. 76. Triangulation und Mixed-Methods* (Vol. 76, pp. 63–99). Springer. https://doi.org/10.1007/978-3-658-24225-1_4

Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. A. (2015). Keeping an eye on learning. *Journal of Teacher Education, 66*(1), 68–85. https://doi.org/10.1177/0022487114549810

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z

Yonemoto, S. (2012). A video annotation tool using vision-based AR technology. In A. Kuijper (Ed.), *2012 international conference on Cyberworlds (CW): 25–27 September 2012, Darmstadt, Germany* (pp. 226–230). Piscataway. https://doi.org/10.1109/CW.2012.39

Bianca Steffen (MA) is a research associate and PhD Candidate at the Institute of Educational Science at the University of Paderborn, Germany. In her teaching position, she is mainly responsible for student's methods training. Her current research interests comprise workplace learning and expertise development. In her PhD project, she examines intuitive decision making.

Maikki Pouta (MA) is a university teacher of special education in University of Turku, Faculty of Education. She is currently writing her dissertation about Teachers' professional vision of student's rational number understanding. Her research interests are in expertise research, teacher education, interaction and visual processing.

Chapter 14
Data Mining and Analytics in the Context of Workplace Learning: Benefits and Affordances

Dirk Ifenthaler

Abstract Technology-based innovations in workplace learning have significantly altered both the scale and resolution of measurements for supporting learning processes in organisations. The increased availability of vast and highly varied amounts of data from settings in the context of workplace learning is overwhelming. This chapter outlines standards in data mining with a specific focus on data from workplace learning. Further, different data mining and analytics methodologies, such as Support Vector Machines or Decision Trees, are presented. An emphasis is shifted to the understanding of learning analytics which are a socio-technical data-mining and analytic practice in educational contexts. The chapter closes with an outlook on how data mining and analytics may provide benefits for future workplace learning scenarios.

Keywords Data analytics · Educational data mining · Learning analytics · Workplace learning · Technology-supported learning

14.1 Introduction

Learning at the workplace calls for a reconsideration of the design of available learning environments, with a special focus on learning supported through digital technologies (Noe et al., 2014). While workplace learning can be referred to as the construction of knowledge and competence by formal and informal means that occurs in the workplace (Billett, 2012), digitally-supported learning is defined as any set of technology-based methods that can be applied to support learning processes and learning outcomes (Ifenthaler, 2010). Emerging opportunities for

D. Ifenthaler (✉)
Learning, Design and Technology, University of Mannheim, Mannheim, Germany

UNESCO Deputy Chair on Data Science in Higher Education Learning and Teaching, Curtin University, Perth, WA, Australia
e-mail: dirk@ifenthaler.info

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_14

digitally-supported learning include game-based learning (Ifenthaler et al., 2012), simulations (Ifenthaler, 2012), Massive Open Online Courses (Ifenthaler & Schumacher, 2016b), social networks (Ifenthaler & Pirnay-Dummer, 2011), or mobile and augmented applications (Ifenthaler & Eseryel, 2013; Sampson et al., 2013). For corporate organisations, digital technologies enable the implementation of customised learning environments even on small scale. Access to digital technologies changes learning in the workplace through cost effective delivery modes, easy to access leaning resources, and flexible learning environments. Currently, digital workplace learning is implemented as formal learning environments (i.e., trainings), for example in the form of Cooperate Open Online Courses (COOCs). However, the yet underrated opportunity for digital technology in workplace learning is the support of informal learning and fostering enablers for lifelong learning (Egloffstein & Ifenthaler, 2017).

Current themes in research on digitally-supported workplace learning include big data for learning, learning and workplace analytics, people analytics, as well as bridging formal learning and informal learning through digital technologies. However, further robust empirical evidence is required in order to better support learning at the digital workplace (Ifenthaler, 2018). This chapter provides an overview on data mining and analytics in the context of workplace learning, while critically reflecting the related benefits and affordances for organisations and its stakeholders.

14.2 Data Mining and Analytics in the Context of Education

The field of social sciences and humanities are still dominated by research, development and implementation of methodologies that primarily divide the world into either *qualitative* or *quantitative* approaches. Organisations use survey instruments or interview techniques to gain insight into stakeholders learning processes or learning outcomes as well as drivers for supporting learning at the workplace (e.g., satisfaction, stress, dispositions) (Kauffeld, 2016). This relatively small toolkit for understanding complex phenomena in organisations limits the next generation of workplace learning environments when they are faced with the increased availability of big data (Gibson et al., 2019). Given the growing availability of data from vast interconnected and loosely coupled systems of administrative, academic, and personal information flowing within and across organisations and businesses, the challenge of data management, analysis, visualisation, and interpretation, which is integral to advancing knowledge and understanding of workplace learning, is constantly evolving (Gibson & Ifenthaler, 2017).

Since the early 2000s, opportunities of data mining and analytics have transformed almost every field of scientific inquiry (Collins et al., 2003; Lodge & Corrin, 2017). When applied to the educational context, these methodologies are referred to as *educational data mining* and *learning analytics* (Baker & Siemens, 2015). Educational data mining includes data mining and analytics techniques, such as visualization, clustering, classification, association rule mining, pattern recognition, or text mining, applied to the context of education (Romero et al., 2011). The analysis of educational data mining are mostly done in retrospective, i.e., big data available from educational settings are being analysed for identifying phenomena and/or predicting possible future states of the educational system. Learning analytics draw on similar data mining and analytics techniques like educational data mining (Berland et al., 2014). However, learning analytics focus (near) real-time analytics and direct support of educational phenomena (Ifenthaler, 2015), for instance adapting learning artefacts for supporting learning processes or providing personalised recommendations for achieving a specific competence level. The two methodologies educational data mining and learning analytics are described in more detail in the following subsections.

14.2.1 Educational Data Mining

Educational data mining (EDM) describes techniques and tools to analyse all kinds of data on different hierarchical levels in educational settings (Berland et al., 2014; Romero et al., 2011). In addition to the nested hierarchical character of educational data (e.g., answer level, session level, student level, educator and organisational level), the performance time, sequence of actions, and evolving elements of the learning context are also important features of relevant data in educational settings. In the context of workplace learning, EDM may be used to analyse the effects of formal training programs over the past 5 years and the possible impact on product innovation or speed of production. EDM is interdisciplinary and draws on machine learning, artificial intelligence, computer science, and classical test statistics to analyse data collected during learning and teaching. Although closely related to learning analytics, which focuses on supporting learning and performance with feedback loops to the learner and instructor in (near) real time (Ifenthaler, 2015; Long & Siemens, 2011), EDM focuses on exploring new patterns in data and on developing new models at all levels of a learning environment. For instance, EDM may identify training programs in an organization which support team-building or signify project teams which produce robust solutions in the product innovation cycle (Ifenthaler, 2014). Some common goals of current EDM practices include: (1) predicting performance and learner success for recruitment, retention and work readiness, (2) evaluating learning within learning management systems and improving instructional sequences, as well as (3) evaluating different kinds of adaptive and personalised support. Additionally, EDM is advancing research about modelling learner, domain, and software characteristics.

There are a variety of approaches and techniques to analyse educational data and different classifications. EDM involves five techniques: (1) prediction, (2) clustering, (3) relationship mining, (4) distillation of data for human judgment, and (5) discovery via models (Baker & Siemens, 2015).

- *Prediction* includes models about performance of learners, for example by analysing their behaviour in an online learning environment. For example, a predictive model can infer the successful completion of a workplace safety training which is defined by a to be predicted variable 'performance' based on a set of predictor variables, for example 'log-in frequency', 'number of forum activities', 'satisfaction with learning materials', or 'attempts in self-assessments'. Predictive models are validated using training data sets for future application in real-world settings. Successful application of predictive models in educational context have been documented in recent research (Adekitan & Noma-Osaghae, 2019; Ifenthaler & Widanapathirana, 2014; Pistilli & Arnold, 2010).
- *Clustering* methods can be used to group learners according to specific characteristics, for instance preference of learning materials (e.g., text, video, combination of both) or performance (e.g., pass vs. fail). This technique aims to discover structural components in educational data without an a priori hypothesis defined. The data points which naturally group together may help to cluster specific groups of learners (e.g., learners meeting minimum requirements) or preconditions of interest (e.g., learner with special needs). Frequent methods include factor analysis, cluster analysis, and network analysis (Gašević et al., 2019; Yousef et al., 2015).
- *Relationship mining*, which is perhaps the most often applied method in EDM, refers to identifying relationships among variables. The aim is to identify strong associations between various variables, such as learning activities, learner interaction or learner performance, and learning strategies. Frequent methods include correlation mining, pattern mining, and association rule mining (Zhu et al., 2020).
- *Distillation* of data for human judgment, aims to depict data in a way that enables stakeholders of an organisation to quickly identify structures in the data and take immediate action. Visualisations help stakeholders to quickly identify instances for increased attention, for example learners being at risk not meeting the course requirements (Bodily & Verbert, 2017; Pistilli & Arnold, 2010; Sedrakyan et al., 2018). Frequent methods include line charts, bar charts, pie charts, progress bars, textual elements, timelines, bubbles, learning paths, heat maps, interaction tables, scatter plots, and checklists (Sahin & Ifenthaler, 2021).
- *Discovery* via models, uses a pre-existing model (e.g., a predictive model) that is then applied to other data and used as a component in further analysis. For instance, a validated predictive model becomes part of a new predictive model to identify further phenomena in the nature of educational data. Discovery models are frequently based on previous predictive models, however, they may also be based on cluster analysis or human created models (based on assumptions or theoretical models) (Berland et al., 2014; Lakkaraju et al., 2015).

14.2.2 Learning Analytics

The use of static and dynamic information about learners, educators, and learning environments for real-time modelling, prediction, and support of learning processes, learning environments, as well as educational decision-making is often referred to as learning analytics (Ifenthaler, 2015). Hence, learning analytics draw on an eclectic set of data and methodologies to provide summative, real-time, and predictive insights for supporting learning, teaching, and organisational efficiency (Lockyer et al., 2013; Long & Siemens, 2011). While the field of learning analytics is receiving a lot of attention in higher education for its capacity to provide lead indicators of student failure (Sclater & Mullan, 2017) or student success (Ifenthaler et al., 2019), it has focused to date on individual courses in isolation, rather than the capabilities of educational organisations in general (Gašević et al., 2015). However, the implementation of learning analytics for workplace learning may have broad implications for the organisation (e.g., technological infrastructure, policies and regulations) and its stakeholders (e.g., students, academic staff, administrators) including changes in learning culture (Ifenthaler, 2020a). From a holistic organisational perspective, learning analytics for workplace learning may support individual stakeholders of an organisation to further develop their mental models or may assist collaborative learning processes in order to build a shared vision of the organisation (Senge, 1990).

Accordingly, the adoption of learning analytics for workplace learning is depending on several factors including technological systems, regulations and guidelines, as well as professional development of involved stakeholders (Ifenthaler, 2017a). Learning analytics ready technological systems are required to process vast amounts of available educational multimodal data, flexible data mining tools, and advanced statistical methods including machine learning algorithms (Berland et al., 2014) – this has been described in more detail in the section on educational data mining above. Learning analytics ready regulations observe data privacy legislation and ethical consideration of (semi-)automated recommendations (Prinsloo & Slade, 2014). Learning analytics ready stakeholders demonstrate educational data literacy, understood as the ethically responsible collection, management, analysis, comprehension, interpretation, and application of data from educational contexts. Accordingly, the degree of adoption of learning analytics for workplace learning within an organisation is a measure of the number of technological systems and organisational regulations implemented as well as the number of stakeholders using learning analytics or who have altered their practice because of learning analytics (Ifenthaler, 2020a).

Ifenthaler and Yau (2020) identified more than 6000 publications focussing on learning analytics and study success. Their systematic review indicates that a wider adoption of learning analytics systems is needed, as well as work towards standardised measures, visualisations, and interventions, which can be integrated into any digital learning environment to reliably predict at-risk learners and to provide adaptive support and intervention strategies. While standards for data models and data collection, such as xAPI (Experience API), exist (Kevan & Ryan, 2016), learning analytics research and development need to clearly define standards for reliable and valid measures, informative visualisations, and design guidelines for pedagogically effective learning analytics interventions (Seufert et al., 2019). In particular, personalised learning environments are increasingly in demand and are valued in workplace learning for creating tailored learning packages optimised for each individual learner based on their personal profile, containing information such as their geo- and socio-demographic backgrounds (Lacave et al., 2018), previous qualifications (Daud et al., 2017), their engagement in the recruitment journey (Berg et al., 2018), activities on websites (Seidel & Kutieleh, 2017), and tracking information on their searches (Macfadyen & Dawson, 2012). For example, learning analytics in the context of workplace learning may suggest windows for formal learning based on the individual work schedule or recommend an informal meeting with a colleague to exchange commonalities and possible synergies of projects being realised in the organisation.

Currently, major analytical strategies for learning analytics involve the above mentioned standards of educational data mining, in addition to variations of regression analysis, such as, linear regression models, logistic regression models, hierarchical linear models (da Silva et al., 2013). Other stochastic approaches include Bayesian networks and neural networks which enable adjustments to the applied algorithms based on previous results (Bartholomew, 1967). However, to identify highly non-linear and complex parameter relationships, the above-mentioned analytical strategies have obvious limitations.

Besides Random Forest (Breiman, 2001) and Decision Tree (Quinlan, 1986) approaches, Support Vector Machines (SVM) is a promising alternative data analytic approach for educational data and learning analytics (Ifenthaler & Widanapathirana, 2014). SVM is a binary classification technique based on supervised machine learning in the broad area of artificial intelligence (Drucker et al., 1997). Major applications include pattern recognition, classification, and regression modeling (Christmann & Steinwart, 2008). The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier (Cortes & Vapnik, 1995). Given a set of training examples, each marked as belonging to one of two categories; an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVM can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. The advantages of SVM can be summarized as follows (Cleophas & Zwinderman, 2013; Williams, 2011):

- SVM offer flexibility in modelling non-linear educational data.
- SVM has short training times to create new models and offer very fast testing speeds when new samples are classified.
- SVM are flexible with regard to interactions between educational parameters from different sources and hardly effected by the correlated parameters unlike most other regression techniques.

- SVM does not rely on priori-knowledge on event probabilities that are often unavailable and unreliable in education data.
- SVM can process imperfect educational data by providing a better sensitivity for modelling dependent variables.
- SVMs are inherently robust against parse data and outliers.

In summary, valid data that could be applied for learning analytics in workplace learning include a combination of learners' background information, behaviour data from digital platforms (e.g., learning management systems, games and simulations), formative and summative assessment data, and information collected through surveys. Hence, measures for learning analytics need to include reactive (i.e., direct response) and non-reactive (i.e., unobtrusive traces) data collection, i.e., multimodal data for supporting learning at the workplace (Blikstein & Worsley, 2016).

14.3 Benefits and Affordances

14.3.1 Benefits for Workplace Learning

From a holistic perspective, educational organisations and involved stakeholders can derive multiple benefits from learning analytics by using different data analytics strategies to produce summative, real-time and predictive insights and recommendations which can be associated with four levels of stakeholders: mega-level (governance), macro-level (organisation), meso-level (curriculum, learning design, educator/tutor), and micro-level (learner) (Ifenthaler, 2015). The *mega-level* facilitates cross-organisational analytics by incorporating data from all levels of learning analytics initiatives (Ifenthaler et al., 2021). Based on common data standards and open data schemas, rich datasets may enable the identification and validation of patterns within and across organisations and therefore provide valuable insights for informing educational policy making (Ifenthaler, 2021). For example, available educational data may help to implement or optimise nation-wide further development programs in order to support organisations to remain competitive. The *macro-level* enables organization-wide analytics for better understanding learner cohorts to optimise associated processes and allocate critical resources for reducing dropout or increasing retention as well as success rates (West et al., 2016). The aforementioned levels can be associated with the emerging domain of Educational Leadership Data Analytics (ELDA) which is centred at the intersection of education leadership and education data analytics (Bowers et al., 2019). For instance, training programs may be identified which are less effective for required competences in the organisation or available data may help to predict a loss of specific competences due to retirements in a specific time frame. The *meso-* and *micro-level* provide analytics insights within entities of the organisation (Ifenthaler & Widanapathirana, 2014). At the meso-level, learning designers may use analytics insights in order to improve learning artefacts or the sequencing of learning tasks as well as overall curricular design (Ifenthaler et al., 2018a). For example, videos may be further enhanced in order to overcome common errors which may have been identified in pattern recognition of previous performance data. The micro-level focusses on learning-processes of individual learners adapting to their current needs in the learning process and helping them to reflect and act on recommendations based on multiple analytics insights (Gašević et al., 2017). For instance, a learner may only spend 15 instead of 30 h in a compliance course, due to prior knowledge or competence already demonstrated.

An essential prerequisite of learning analytics benefits, however, is the perspective of data and analytics involved: (1) summative and descriptive; (2) formative and (near) real-time; (3) predictive and prescriptive. The *summative and descriptive* perspective provides detailed insights after completion of a learning phase (e.g., learning period, training programme), often compared against previously defined reference points or benchmarks. The *formative and (near) real-time* perspective uses ongoing information for improving processes through direct interventions on-the-fly. The *predictive and prescriptive perspective* is applied for forecasting the probability of outcomes in order to plan for future interventions, strategies, and actions.

14.3.2 Affordances and Challenges

While organisations choose to invest into advancing learning analytics capabilities, they will encounter several adoption challenges (Drachsler & Greller, 2016; Leitner et al., 2019; Tsai & Gašević, 2017). Accordingly, a comprehensive change management strategy for implementing actionable learning analytics seems to be pivotal. As a first step, a readiness assessment needs to be conducted while referring to the above mentioned benefits of learning analytics introduced by Ifenthaler (2015). Based on the initial step of identifying the to be achieved benefits of learning analytics, the organisation needs to carefully build a strategy based on an in-depth review of existing practices, procedures, and capabilities. Hence, the strategy includes decisions which benefits and specific features of learning analytics to be included as well as which infrastructure for a successful learning analytics implementation is required. The second step includes the readiness assessment using standardised instruments. Data for the readiness assessment is collected on organisational level, i.e., organisational readiness (e.g., existing policy, data protection regulation), with a specific focus on necessary technology, i.e., technological readiness (e.g., data warehouse, system integration), and with regard to capabilities staff members of the organisation possess, i.e., staff readiness (e.g., educational data literacy) (Ifenthaler, 2017a; Schumacher et al., 2019).

Based on the results of the in-depth readiness assessment, the resulting implementation strategy covers all areas of change (Gibson & Ifenthaler, 2020). Further, the stepwise implementation of learning analytics for workplace learning needs to be monitored and evaluated with regard to predefined and measurable KPIs (Key Performance Indicators). More importantly, the return on investment, defined as the expected gains (returns) per unit of cost (investment) of a variety of potential actions need to be monitored closely (Psacharopoulos, 2014). The return on investment may not only be conceptualised as monetary returns, but may also be returns conceptualised as retaining staff, or improving staff satisfaction (Gibson et al., 2018). Clearly, the change management process of an organisation is a time-consuming process and requires acceptance among all involved stakeholders. A change management strategy for implementing actionable learning analytics in educational organisations could be guided by the following principles (Ifenthaler, 2020a; Leitner et al., 2019; Tsai & Gašević, 2017):

- Definition of the learning analytics vision and objectives (e.g., using the above mentioned benefits matrix) and align them with the organisations mission and learning culture;
- Identification of organisational, political, or technological factors that will affect the implementation;
- Involvement and continuous information of all stakeholders including staff, management, etc.;
- Development of (and continuously update) a strategic plan focussing on short-term and long-term wins, including a needs and risk analysis as well as a clear timeline outlining responsibilities of involved stakeholders;
- Allocation of resources and identification of expertise (inside and outside of the organisation) for achieving the learning analytics objectives;
- Undertake a robust formative and summative evaluation of the learning analytics initiative to further refine the overall implementation and organisational change process.

In a recent systematic review including over 6000 initial studies on learning analytics which were published over the past 6 years, Ifenthaler and Yau (2020) indicate that a wider adoption of organisation-wide learning analytics systems is needed. While standards for data models and data collection, such as xAPI (Experience API), exist (Kevan & Ryan, 2016), learning analytics research and development needs to clearly define standards for reliable and valid indicators, informative visualisations, and design guidelines for pedagogically effective learning analytics interventions (Sahin & Ifenthaler, 2021; Seufert et al., 2019; Yau & Ifenthaler, 2020).

Further, serious concerns and challenges are associated with the application of learning analytics (Pardo & Siemens, 2014). For instance, not all educational data is relevant and equivalent for workplace learning. Therefore, the validity of data and its analyses is critical for generating useful summative, real-time, and predictive insights (Macfadyen & Dawson, 2012). Furthermore, limited access to educational data may generate disadvantages for involved stakeholders. For example, invalid forecasts may lead to inefficient decisions and unforeseen problems (Ifenthaler & Widanapathirana, 2014). Moreover, ethical and privacy issues are associated with the use of educational data for learning analytics (Ifenthaler & Tracey, 2016). That implies how personal data are collected and stored as well as how they are analysed and presented to different stakeholders (Slade & Prinsloo, 2013).

Hence, before utilising educational data and analytics for workplace learning, organisations are required to address privacy issues linked to learning analytics: They need to define who gets access to which data, where and how long will the data be stored, and which procedures and algorithms are implemented to further use the available data (Ifenthaler & Schumacher, 2016a, 2019). Slade and Prinsloo (2013) as well as Pardo and Siemens (2014) established several principles for privacy and ethics in learning analytics. They highlight the active role of learners in their learning process, the temporary character of data, the incompleteness of data on which learning analytics are executed, transparency regarding data use, as well as purpose, analyses, access, control, and ownership of data. Drachsler and Greller (2016) established the DELICATE checklist to implement 'trusted' learning analytics considering ethical and privacy aspects, suggestions of current legal frameworks and privacy fears associated with learning analytics. The checklist includes aspects such as determining the organisations' goals for implementing learning analytics, explaining intentions, involving all relevant stakeholders and the data subjects and seeking their consent but also technical aspects and how to involve external providers.

14.4 Conclusion

Technology-based innovations in workplace learning have significantly altered both the scale and resolution of measurements for complex formal and informal learning processes (Gibson & Ifenthaler, 2017; Ifenthaler, 2017b, 2020b). Recent developments in this field have heightened the need for educational data mining, machine learning, and advanced statistics to gain insights from the fine-grained process data generated in technology-rich learning environments (Ifenthaler, 2017b, 2020b; Ifenthaler et al., 2018b). Several perspectives on educational data and analytics have been identified: (1) The *data-driven perspective* utilises existing data, mostly stemming from database systems, for informing different stakeholders. While big datasets may be available in organisations, the purpose for collecting data may have been different in the first place, hence, being biased when utilised for purposes informing workplace learning. In contrast, the (2) *data-demand perspective* follows a specific analytics purpose and defines the data to be collected. This enables a well-directed analysis of educational data with direct implications for learning and teaching at the workplace. In addition, a combination of the above-mentioned approaches may be considered (Yau & Ifenthaler, 2020).

The challenges and issues discussed in this article reveal the requirements for developments in theory as well as some of the practical challenges that will need to be overcome if organisations are to achieve the vision of providing workplace learning with a 'quiet assessment' system in which the impact can be turned up at the request of learners and educators as they seek to understand the progress and outcomes of lifelong learning in the organisation. In moving forward to embrace the opportunities that could be provided by educational data mining and learning analytics for workplace learning, the challenges that remain to be addressed must not be underestimated before organisations can use (semi-)automated analytics of complex competences and understanding with confidence.

References

Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies, 24*, 1527–1543. https://doi.org/10.1007/s10639-018-9839-7

Baker, R. S., & Siemens, G. (2015). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–272). Cambridge University Press.

Bartholomew, D. J. (1967). *Stochastic models for social processes*. Wiley.

Berg, A. M., Branka, J., & Kismihók, G. (2018). Combining learning analytics with job market intelligence to support learning at the workplace. In D. Ifenthaler (Ed.), *Digital workplace learning. Bridging formal and informal learning with digital technologies* (pp. 129–148). Springer.

Berland, M., Baker, R. S., & Bilkstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1–2), 205–220. https://doi.org/10.1007/s10758-014-9223-7

Billett, S. (2012). Workplace learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 3477–3480). Springer. https://doi.org/10.1007/978-1-4419-1428-6_478

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies, 10*(4), 405–418.

Bowers, A. J., Bang, A., Pan, Y., & Graves, K. E. (2019). *Education leadership data analytics (ELDA): A white paper report on the 2018 ELDA summit*. Columbia University.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Christmann, A., & Steinwart, I. (2008). *Support vector machines*. Springer.

Cleophas, T. J., & Zwinderman, A. H. (2013). Support vector machines. In *Machine learning in medicine* (pp. 155–161). Springer. https://doi.org/10.1007/978-94-007-6886-4_15

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The human genome project: Lessons from large-scale biology. *Science, 300*(5617), 286–290. https://doi.org/10.1126/science.1084564

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/bf00994018

da Silva, J. L., Caeiro, F., Natário, I., & Braumann, C. A. (2013). *Advances in regression, survival analysis, extreme values, markov processes and other statistical applications*. Springer.

Daud, A., Aljohani, N., Abbasi, R., Lytras, M., Abbas, F., & Alowibdi, J. (2017). *Predicting student performance using advanced learning analytics.* Conference on World Wide Web Companion.

Drachsler, H., & Greller, W. (2016). Privacy and analytics – It's a DELICATE issue. *A checklist for trusted learning analytics.* Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, UK.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems 9* (pp. 155–161). MIT Press.

Egloffstein, M., & Ifenthaler, D. (2017). Employee perspectives on MOOCs for workplace learning. *TechTrends, 61*(1), 65–70. https://doi.org/10.1007/s11528-016-0127-3

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics, 4*(2), 113–128. https://doi.org/jla.2017.42.10.

Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior, 92*, 562–577. https://doi.org/10.1016/j.chb.2018.07.003

Gibson, D. C., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 29–42). Springer.

Gibson, D. C., & Ifenthaler, D. (2020). Adoption of learning analytics. In D. Ifenthaler & D. C. Gibson (Eds.), *Adoption of data analytics in higher education learning and teaching* (pp. 3–20). Springer.

Gibson, D. C., Huband, S., Ifenthaler, D., & Parkin, E. (2018). *Return on investment in higher education retention: Systematic focus on actionable information from data analytics* ascilite Conference, Geelong, VIC, Australia, 25-11-2018.

Gibson, D. C., Webb, M., & Ifenthaler, D. (2019). Measurement challenges of interactive educational assessment. In D. G. Sampson, J. M. Spector, D. Ifenthaler, P. Isaias, & S. Sergis (Eds.), *Learning technologies for transforming teaching, learning and assessment at large scale* (pp. 19–33). Springer.

Ifenthaler, D. (2010). Learning and instruction in the digital age. In J. M. Spector, D. Ifenthaler, P. Isaías, Kinshuk, & D. G. Sampson (Eds.), *Learning and instruction in the digital age: Making a difference through cognitive approaches, technology-facilitated collaboration and assessment, and personalized communications* (pp. 3–10). Springer.

Ifenthaler, D. (2012). Computer simulation model. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (Vol. 3, pp. 710–713). Springer.

Ifenthaler, D. (2014). Toward automated computer-based visualization and assessment of team-based performance. *Journal of Educational Psychology, 106*(3), 651–665. https://doi.org/10.1037/a0035505

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Sage.

Ifenthaler, D. (2017a). Are higher education institutions prepared for learning analytics? *TechTrends, 61*(4), 366–371. https://doi.org/10.1007/s11528-016-0154-0

Ifenthaler, D. (2017b). Designing effective digital learning environments: Toward learning analytics design. *Technology, Knowledge and Learning, 22*(3), 401–404. https://doi.org/10.1007/s10758-017-9333-0

Ifenthaler, D. (2018). How we learn at the digital workplace. In D. Ifenthaler (Ed.), *Digital workplace learning. Bridging formal and informal learning with digital technologies* (pp. 3–8). Springer. https://doi.org/10.1007/978-3-319-46215-8_1

Ifenthaler, D. (2020a). Change management for learning analytics. In N. Pinkwart & S. Liu (Eds.), *Artificial intelligence supported educational technologies* (pp. 261–272). Springer.

Ifenthaler, D. (2020b). Supporting higher education students through analytics systems. *Journal of Applied Research in Higher Education, 12*(1), 1–3. https://doi.org/10.1108/JARHE-07-2019-0173

Ifenthaler, D. (2021). Learning analytics for school and system management. In OECD (Ed.), *OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots* (pp. 161–172). OECD Publishing.

Ifenthaler, D., & Eseryel, D. (2013). Facilitating complex learning by mobile augmented reality learning environments. In R. Huang, Kinshuk, & J. M. Spector (Eds.), *Reshaping learning: The frontiers of learning technologies in a global context* (pp. 415–438). Springer.

Ifenthaler, D., & Pirnay-Dummer, P. (2011). States and processes of learning communities. Engaging students in meaningful reflection and elaboration. In B. White, I. King, & P. Tsang (Eds.), *Social media tools and platforms in learning environments: Present and future* (pp. 81–94). Springer. https://doi.org/10.1007/978-3-642-20392-3_5

Ifenthaler, D., & Schumacher, C. (2016a). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y

Ifenthaler, D., & Schumacher, C. (2016b). Udacity. In S. Danver (Ed.), *The SAGE encyclopedia of online education* (pp. 1149–1151). Sage.

Ifenthaler, D., & Schumacher, C. (2019). Releasing personal information within learning analytics systems. In D. G. Sampson, J. M. Spector, D. Ifenthaler, P. Isaias, & S. Sergis (Eds.), *Learning technologies for transforming teaching, learning and assessment at large scale* (pp. 3–18). Springer.

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240. https://doi.org/10.1007/s10758-014-9226-4

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development, 68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning. Foundations, innovations, and perspectives* (pp. 3–10). Springer.

Ifenthaler, D., Gibson, D. C., & Dobozy, E. (2018a). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology, 34*(2), 117–132. https://doi.org/10.14742/ajet.3767

Ifenthaler, D., Sampson, D. G., & Spector, J. M. (2018b). Linking analytics data and digital systems for supporting cognition and exploratory learning in 21st century. *Computers in Human Behavior, 78*, 348–350. https://doi.org/10.1016/j.chb.2017.10.016

Ifenthaler, D., Yau, J. Y.-K., & Mah, D.-K. (Eds.). (2019). *Utilizing learning analytics to support study success*. Springer.

Ifenthaler, D., Gibson, D. C., Prasse, D., Shimada, A., & Yamada, M. (2021). Putting learning back into learning analytics: Actions for policy makers, researchers, and practitioners. *Educational Technology Research and Development, 69*(4), 2131–2150. https://doi.org/10.1007/s11423-020-09909-8

Kauffeld, S. (2016). *Nachhaltige Personalentwicklung und Weiterbildung* (2nd ed.). Springer.

Kevan, J. M., & Ryan, P. R. (2016). Experience API: Flexible, decentralized and activity-centric data collection. *Technology, Knowledge and Learning, 21*(1), 143–149. https://doi.org/10.1007/s10758-015-9260-x

Lacave, C., Molina, A., & Cruz-Lemus, J. (2018). Learning analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology, 37*(10–11), 993–1007. https://doi.org/10.1080/0144929X.2018.1485053

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, D., Ghani, R., & Addison, K. L. (2015). *A machine learning framework to identify students at risk of adverse academic outcomes.* KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia.

Leitner, P., Ebner, M., & Ebner, M. (2019). Learning analytics challenges to overcome in higher education institutions. In D. Ifenthaler, J. Y.-K. Yau, & D.-K. Mah (Eds.), *Utilizing learning analytics to support study success* (pp. 91–104). Springer.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439–1459. https://doi.org/10.1177/0002764213479367

Lodge, J. M., & Corrin, L. (2017). What data and analytics can and do say about effective learning. *npj Science of Learning, 2*(1), 5. https://doi.org/10.1038/s41539-017-0006-5

Long, P. D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 31–40.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-Learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society, 15*(3), 149–163.

Noe, R. A., Clarke, A. D. M., & Klein, H. J. (2014). Learning in the twenty-first-century workplace. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 245–275.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Pistilli, M. D., & Arnold, K. E. (2010). Purdue signals: Mining real-time academic data to enhance student success. *About Campus: Enriching the Student Learning Experience, 15*(3), 22–24. https://doi.org/10.1002/abc.20025

Prinsloo, P., & Slade, S. (2014). Student data privacy and institutional accountability in an age of surveillance. In M. E. Menon, D. G. Terkla, & P. Gibbs (Eds.), *Using data to improve higher education. Research, policy and practice* (pp. 197–214). Sense Publishers.

Psacharopoulos, G. (2014). The returns to investment in higher education. In M. E. Menon, D. G. Terkla, & P. Gibbs (Eds.), *Using data to improve higher education. Global perspectives on higher education* (pp. 121–148). Sense Publishers.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1023/A:1022643204877

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2011). *Handbook of educational data mining*. CRC Press.

Sahin, M., & Ifenthaler, D. (Eds.). (2021). *Visualizations and dashboards for learning analytics*. Springer. https://doi.org/10.1007/978-3-030-81222-5

Sampson, D. G., Isaias, P., Ifenthaler, D., & Spector, J. M. (Eds.). (2013). *Ubiquitous and mobile learning in the digital age*. Springer. https://doi.org/10.1007/978-1-4614-3329-3

Schumacher, C., Klasen, D., & Ifenthaler, D. (2019). Implementation of a learning analytics system in a productive higher education environment. In M. S. Khine (Ed.), *Emerging trends in learning analytics* (pp. 177–199). Brill.

Sclater, N., & Mullan, J. (2017). *Learning analytics and student success – Assessing the evidence*. JISC.

Sedrakyan, G., Mannens, E., & Verbert, K. (2018). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Visual Languages and Computing, 50*, 19–38. https://doi.org/10.1016/j.jvlc.2018.11.002

Seidel, E., & Kutieleh, S. (2017). Using predictive analytics to target and improve first year student attrition. *Australasian Journal of Educational Technology, 61*(2), 200–218.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. Doubleday.

Seufert, S., Meier, C., Soellner, M., & Rietsche, R. (2019). A pedagogical perspective on big data and learning analytics: A conceptual model for digital learning support. *Technology, Knowledge and Learning, 24*(4), 599–619. https://doi.org/10.1007/s10758-019-09399-5

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Tsai, Y.-S., & Gašević, D. (2017). *Learning analytics in higher education—Challenges and policies: A review of eight learning analytics policies.* Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada.

West, D., Heath, D., & Huijser, H. (2016). Let's talk learning analytics: A framework for implementation in relation to student retention. *Online. Learning, 20*(2), 1–21. https://doi.org/10.24059/olj.v20i2.792

Williams, G. (2011). Support vector machines. In *Data mining with rattle and R* (pp. 293–304). https://doi.org/10.1007/978-1-4419-9890-3_14

Yau, J., & Ifenthaler, D. (2020). Reflections on different learning analytics indicators for supporting study success. *International Journal of Learning Analytics and Artificial Intelligence for Education, 2*(2), 4–23. https://doi.org/10.3991/ijai.v2i2.15639

Yousef, A. M. F., Chatti, M. A., Wosnitza, M., & Schroeder, U. (2015). A cluster analysis of MOOC stakeholder perspectives. *International Journal of Educational Technology in Higher Education, 12*, 74–90.

Zhu, M., Bonk, C. J., & Doo, M. Y. (2020). Self-directed learning in MOOCs: Exploring the relationships among motivation, self-monitoring, and self-management. *Educational Technology Research and Development, 68*(5), 2073–2093. https://doi.org/10.1007/s11423-020-09747-8

Dirk Ifenthaler (PhD) is a Professor and Chair of Learning, Design and Technology at University of Mannheim, Germany, UNESCO Deputy Chair of Data Science in Higher Education Learning and Teaching at Curtin University, Australia and Visiting Senior Research at University of Latvia, Latvia. Dirk's research focuses on the intersection of cognitive psychology, educational technology, data analytics, and organisational learning. He is the Editor-in-Chief of the Technology, Knowledge and Learning and Senior Editor of Journal of Applied Research in Higher Education.

Chapter 15
Addressing 'Wicked Problems' Using Visual Analysis

Eva Kyndt and Jan Aerts

Abstract Phenomena studied within the field of professional learning and development are often highly complicated as well as highly variable. In general, we accept that (professional) learning has many facets, overlapping elements and interconnected aspects. Yet, when empirically examining these phenomena, we tend to reduce the complex reality to a set of tangible variables and actors that can be explained by (at most a combination of) theoretical frameworks. As such, we treat professional learning 'problems' as if they were 'tame' problems that are well-defined, stable, and prone to control and prediction. Starting from the notion of wicked problems (Rittel HWJ, Webber MM. Policy Sci 4:155–169, 1973), the current state of the art of visual analysis as a methodology for approaching research questions in the field of professional learning is presented. Visual analysis complements and extends insights from traditional analysis as it enables researchers to discover hidden patterns and deal with big data. After giving a general introduction to what visual analysis is, this chapter presents an empirical example of how it was used to examine the implications of missingness in longitudinal multilevel data on teachers' engagement.

Keywords Visual analysis · Data Missingness · Teacher engagement · Wicked problem

---

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-08518-5_15].

E. Kyndt (✉)
Centre for Transformative Innovation, Swinburne University of Technology, Melbourne, VIC, Australia
e-mail: ekyndt@swin.edu.au

J. Aerts
Universiteit Hasselt, Hasselt, Belgium
e-mail: Jan.aerts@uhasselt.be

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_15

15.1 Introduction

Phenomena studied within the field of professional learning and development are often highly complicated as well as highly variable. In general, we accept that professional learning has many facets, overlapping elements and interconnected aspects. In addition, they are rarely static in nature; learners, (various types of) instructors and coaches, learning contexts, learning conditions as well as desired learning outcomes are dynamically evolving over time. Moreover, if we truly want to impact professional (lifelong) learning, we need to consider that different stakeholders are involved who often strive towards different – sometimes competing – goals. Yet, when empirically examining professional learning phenomena, we tend to reduce the complex reality to a set of tangible variables and actors that can be explained by (at most a combination of) theoretical frameworks. As such, we treat these phenomena as if they were "tame" problems that are well-defined, stable, and prone to control and prediction.

Despite the merits of the current traditional quantitative and qualitative research methodologies, we need to wonder whether they are still solely appropriate for understanding and tackling realistic learning questions, especially when we consider that more and more data is currently being generated and work is becoming increasingly complex. Starting from the perspective that professional learning phenomena are "wicked problems" (Rittel & Webber, 1973), we propose that visual analysis is an interesting method of analysis if we truly want to capture and understand its complexity. The notion of a "wicked problem" has been introduced by Rittel and Webber and refers to an ill-defined problem, where views vary considerably across diverse parties with a vested interest (i.e., stakeholders).

Different authors have argued that a new shift in thinking is needed to tackle ill-defined, dynamic and uncontrollable problems (Gibson & Iftenthaler, 2017; Jordan et al., 2014). As such, analysing data by means of data visualisation, in which a cognitive task is converted into a perceptive one, is suggested as one possible methodological avenue for educational research (Gibson & Iftenthaler, 2017). The visual analytical approach relies completely on **interactive** and **integrated** data visualisations for exploratory data analysis in order to identify unexpected trends, outliers or patterns. These can indicate interesting hypotheses that then can be pursued and deepened using other methods such as for example experience sampling (Seifried et al., 2022), latent profile analysis (Bauer, 2022), video-based interaction analyses (Filliettaz et al., 2022), etc. A thorough search of the literature confirmed that visual analysis is indeed still a relatively unknown method in the field of professional learning and development, and even educational sciences more broadly. In addition, the scope of learning-related research topics and questions currently being addressed with visual analysis is limited; The vast majority of the identified studies focused on e-learning, learning outcomes (incl. student trajectories) and (social) networks.

The current chapter aims to introduce visual analysis to researchers interested in professional learning and development. It starts with explaining the theory of wicked problems and illustrates why professional learning and development can be considered a wicked problem. Subsequently, we provide a high-level explanation of what visual analysis is and is not, and why it is a suitable approach to tackle wicked problems. In the final section, we illustrate an application of visual analysis. More specifically, we used visual analysis to explore the impact of multilevel missingness in our longitudinal data on teachers' individual and team work engagement.

15.2 Theoretical Background

15.2.1 Theory of Wicked Problems

The conceptualisation of wicked problems was formally introduced by design theorists Rittel and Webber in 1973. With their notion of wicked problems, Rittel and Webber were responding to the ongoing rationalisation – and subsequent inadequate problem-solving methods – of societal problems (Farrell & Hooker, 2013) and therefore stated that societal problems were inherently different (i.e., wicked) from exact science problems (i.e., tame). Farrell and Hooker (2013) argue that wicked problems have three common sources: conditions of finitude, complexity, and normativity.

Agent *finitude* refers to the fact that knowledge, cognitive capacity, and resources are limited. This finitude is, however, not solely tied to a specific individual but also to the limits of our society and even species. The fact that available resources are finite and insufficient to obtain an optimal solution is considered a necessary condition for a problem to be considered wicked (Farrell & Hooker, 2013).

With system *complexity*, Farrell and Hooker (2013) pinpoint the fact that "every aspect of our world exists of interactions between partially nested hierarchies of complex systems with multiple feedback and feedforward loops" (p. 686). As such, actions might have far-reaching consequences on different functional levels, and their effects are difficult to disentangle from (reoccurring) interactions. In addition, the outcomes of actions are difficult to predict. The complexity of wicked problems exacerbates the finitude of our resources (Farrell & Hooker, 2013).

The third feature, problem *normativity*, is inherently tied to the fact that multiple stakeholders are involved when attempting to resolve a wicked problem. It refers to the fact that human values and norms can be inextricably intertwined with problem formulation and resolution. As different stakeholders often have different values, norms and consequent priorities, a practical and coherent problem solution is often a compromise between the different stakeholders that are involved (Farrell & Hooker, 2013).

15.2.2 A Wicked Problem Perspective on Professional Learning and Development

We argue that professional learning and development can be approached from a wicked problem perspective and analyse the presence of the three conditions described above. The first condition refers to agent *finitude*. Designing perfect professional learning and development is not possible, as the resources of both the individual, organisation, and society are limited. The individual does not have infinite knowledge or time, and organisations and societies likewise are typically strained for resources. Investments (financial and otherwise) into professional learning and development always need to be balanced with (organisational) productivity, individual aspirations, and various societal needs.

Secondly, professional learning and development are inherently complex. *Complexity* arises from the fact that professional learning and development does not occur in isolation. Individuals with their own characteristics are influenced by or learn from other individuals (Kyndt et al., 2022). They are grouped within organisations providing certain affordances or hindrances, which are in turn part of sectors that may or may not have specific regulations and requirements. As such, learning and professional development is, per definition, part of a hierarchical and reciprocal complex system. In addition, learning and development is not static; it is dynamic in nature with events and experiences from the past influencing future behaviour (Kyndt & Baert, 2013).

Finally, professional learning and development is not free of *normativity*. Different stakeholders are involved, and they may have different vested interests. The priorities of the employee might differ from those that the organisation puts forward, and tensions might exist. Similarly, the needs of the society in terms of human capital might not align with the needs of specific organisations. As such, there might be considerable differences in terms of what is considered valuable or not between different stakeholders.

15.2.3 Tackling Wicked Problems with Visual Analysis

In general, two recommendations are made for addressing wicked problems (Gibson & Iftenthaler, 2017; Jordan et al., 2014; Liebowitz, 2017). First, several authors advocate combining computational power with human analytic capabilities (Jordan et al., 2014; Thomas & Kielman, 2009). This – for the field of professional learning – novel approach could enable us to deal, at least in part, with the finitude and complexity conditions of wicked problems (Farrell & Hooker, 2013). Secondly, Jordan et al. (2014) propose to address wicked problems through continued observation and curiosity, input from diverse stakeholders and collective and distributed sense-making. Conklin (2006) agrees and states that the problem-solving process is now clearly social in which a shared understanding of possible solutions is pivotal due to the normativity condition of wicked problems. According to Conklin (2006), a shared visual representation of the problem at hand facilitates a shared understanding. These two general recommendations – combining computational strengths and human analytics, and creating a shared understanding through shared representation – can be realised by addressing wicked educational problems with a visual analytical approach. Hawryszkiewycz (2014) proposed that visualisation may enhance the resolution of wicked problems as it facilitates high-level cognition among stakeholders by inducing insight, reasoning, and understanding (Edirisinghe et al., 2016, p. 297). In other words, visual analysis can empower individuals to take control of the analytic process (Keim et al., 2010).

In general, visual analytics is often described as the science of analytical reasoning facilitated by a visual interface (Thomas & Kielman, 2009), allowing for an **interaction** between a human analyst with their data. A core component of this approach is *data* visualisation, in which cognitive tasks are converted in perceptive by laying out marks (i.e., points, lines & areas) on a 2D screen using different channels (e.g., position, length, width, angle, colour, opacity); a process named visual encoding. Data visualisation can take up several roles within the analytical process, and visualisations can be positioned across a spectrum from exploratory (i.e., where the user is allowed to browse through the data without the need to have a specific question) to explanatory (i.e., where a message needs to be conveyed, e.g., a graph showing an interaction effect). Visualisations can be generic (e.g., line chart, bar chart, etc.) but will often need to be custom when complex data is considered (Sakai, 2016; Victor, 2011; see illustration in this chapter). In addition, the visual design process is typically very iterative, as the act of exploration itself will inform on what is possible or not (so-called "adjacent possible"; Sakai, 2016). For an in-depth exploration of the data visualisation process, see Munzner (2014) and Sakai (2016).

The visual analytics approach is situated at the exploration end of the data visualisation continuum; it relies on interactive and integrated visualisations for exploratory data analysis in order to identify unexpected trends, outliers or patterns. These can indicate interesting hypotheses that then can be pursued using automated algorithms and help define parameter spaces for these algorithms (Alcaide & Aerts, 2021), as different (combinations of) parameters often have unexpected and hidden effects on the results of automated algorithms and pipelines. By putting a human back into the loop to guide the analysis, interactive data visualisation has an important role in hypothesis-free data exploration (van Wijk, 2005; Keim et al., 2010). Especially in the field of social sciences, the human takes up an essential role in the analysis process. Winters et al. (2014) consider the *essential human* in their conceptual framework for characterising domain problems. They state – in line with the concept of wicked problems – that in social sciences "*problems and problem solvers are tightly coupled*" (p. 19), as such it is not possible to offer one generic approach or protocol to doing visual analysis. In essence, a visual analysis protocol can only be presented in conjunction with the topic and/or research question at hand, as we will do below.

15.2.4 What Is Not Visual Analysis?

In principle, determining if a study used visual analysis or not is straightforward; that is, it entails that the analysis of the empirical data was done using (interactive) data visualisation. However, in reality, this description is less straightforward than it might seem at first glance. To illustrate the challenges in establishing what can truly be considered visual analysis and what is not, we will elaborate on some specific types of studies that we encountered in the literature that we did not consider a visual analysis study. First, a lot of studies include a visual representation of the results, however as these visuals are not being analysed to answer the research question, they merely support the explanation of the results and are thus not considered visual analysis as defined above. Secondly, the analysis of visual data formats (e.g., video, photo's, drawings) are also not considered visual analysis because the visuals, in this case, are the data that is being analysed rather than a method for answering the research question. The same argument goes for research – using traditional methods of analysis – about the use of simulations, visualisations, images, augmented reality, and various tools as didactical materials. In addition, studies examining visual skills and spatial ability in learners are not considered, as visual abilities are the subject of study and not the method. In sum, studies that use a graphical representation for results, analyse graphical data or have visualisation and related abilities as a research topic are not considered part of the visual analytical approach if their analysis strategy does not include visualisation of data for answering research questions.

15.3 Illustration

15.3.1 Evaluating Missingness in Longitudinal Multilevel Data on Teachers' Engagement

Above we gave a general description of what visual analysis is and can offer. By presenting a concrete example of how we used visual analysis to examine the impact of missingness in longitudinal multilevel data, we hope to make this abstract concept of visual analysis tangible. It is, however, beyond the scope of this chapter to describe multiple specific designs or protocols for analyses since visual analysis entails an infinite number of possible visualisations, including existing designs as well as (to be developed) custom designs which can all be approached with an analysis protocol fitting the research question of specific studies. For an in-depth presentation of visual analysis, we refer the reader to Munzner (2014).

15.3.2 Missingness in Longitudinal Data: A Wicked Problem

Every research that has ever undertaken a longitudinal study knows that attrition over time is (almost) impossible to avoid. While different approaches exist to examine if there is (more or less) randomness in this missingness of data and different estimation methods enable statistical inference from incomplete data (Little, 1988), there is no way of truly knowing if the assumptions that are made are in fact reality (Seaman et al., 2013). We generally accept that there is a potential bias in our results but often have little insight into the direction or magnitude of this potential bias. In a similar vein, many of us studying professional learning and development know first-hand how difficult it is to achieve very high response rates within organisations, courses, programs, groups, and teams. Even in cross-sectional studies, we are usually forced to accept certain levels of potential bias and lack of clarity around magnitude and direction. As such, it is not difficult to imagine the challenge that arises when missingness at the group level is combined with attrition over time, especially if we want to draw conclusions at the different levels (i.e., individual, group/team, organisation). Especially in research on teams (e.g., team learning, group development) there are many different opinions and a lack of consensual guidelines about how many team members need to participate in order to analyse team-level constructs (Mulder, 2022).

Starting from the three conditions of Farrell and Hooker (2013), we argue that missingness in longitudinal multilevel data presents us with a wicked problem:

1. *Finitude*: Even with infinite time and (computational) resources, it is not possible to determine from the observed data whether the missing data are missing at random or not. It is only possible to test if the data is clearly not missing completely at random (Little, 1988).
2. *Complexity*: In the case of longitudinal multilevel data, missingness across time and missingness across the different levels (i.e., individual, team, organisation) are not independent of each other. If an individual or team, or organisation does not participate at each time point, this has an impact on the overall response rate, response rate at the team level and organisational level and the attrition rates across the different levels.
3. *Normativity*: As mentioned, there is a lack of universally established guidelines when it comes to the best way to deal with missing data. Different disciplines, research traditions, research fields or even different researchers have different arguments for the choices they make. In addition, we need to acknowledge that many choices are not real choices but a consequence of the available data, analytical capability, or preference of the anonymous reviewer. As such, it is clear that many different motivations and stakeholders are involved when it comes to what is considered an appropriate way to handling missingness in longitudinal multilevel data.

15.3.3 The Evolution of Teachers' (Team) Engagement

15.3.3.1 Background

Schools are often described as loosely coupled organisations due to the high level of autonomy teachers have in their jobs and the little control or coordination of school principals. Yet, there are ample studies showing that teacher personal and professional wellbeing is, to a large extent, determined by interactions with their colleagues (Kyndt et al., 2016; Meredith et al., 2020). While work engagement – a positive, fulfilling work-related psychological state of mind characterised by vigour, dedication, and absorption – has proven its relevance in many job settings, including schools, research has mainly focused on individual work engagement, often disregarding and potentially underestimating the role of the immediate social neighbourhood of the individual teacher and potential social contagion processes (Meredith et al., 2020). As such, this study set out to examine the relationship between individual and team work engagement over time. Similar to individual work engagement, Torrente et al. (2012b) define team work engagement as "a positive, fulfilling, work-related and shared psychological state characterised by team work vigour, dedication and absorption which emerges from the interaction and shared experiences of the members of a work team" (p. 107). Given that team work engagement is a *shared* state in teams, and the aim was to examine how this relation evolves over time, the first important step in the project was to gain insight into the impact of missingness at both the individual level and team level. Mulder (2022) highlighted that dealing with missing data in teams, especially in longitudinal studies, is a specific challenge for this type of research, making the analyses and interpretation of data and results daunting. As such, we believe that visual analysis, which enables us to tap into tacit knowledge held by the researcher, and which provides a wider bandwidth for exploring the data in all its complexity, will support us in gauging the impact of the multilevel missingness in our data.

15.3.3.2 Goals and Research Questions

Although there is also missingness at the school level (i.e., not all teachers of the school participated), we will focus in this illustration, in line with the measured constructs, on the missingness (across time) at the individual and team level. Our specific research questions, we want to examine using an interactive visualisation, are:

- Is drop-out biased over time?
  - *At the individual level?*
  - *At the team level?*
- What is the impact of different participation rates within teams on our findings?

15.3.3.3 Sample

The data were collected from teachers working in secondary schools in Flanders (Belgium) who are part of subject groups. Subject groups are formal structural units within schools that comprise teachers who teach the same or closely related subjects so that they can collaborate on subject-related matters (e.g., curriculum, student evaluation). In total, 2038 teachers from 400 different subject groups within 37 schools participated at least at one point in time. The mean team size of the subject groups ranged from 2 to 32 members ($M = 6.77$; $SD = 4.40$). Some teachers were a member of multiple subject groups but were asked to respond to the questions for the subject (groups) they were teaching most hours for. Figure 15.1 presents an overview of the multilevel structure of the data.

All teachers of the 400 subject groups in the 37 participating schools were invited to participate at each point in time regardless of their participation in the prior wave. Table 15.1 provides an overview of the number of responses at each point in time. Within the dataset, there are different types of missing data;

- Missingness over time: not all waves for all participants
- Missingness within teams: not all team members participated (participation rate)
- Missingness within teams over time: participation rate & drop-out over time
- Missingness of teams in schools: not all teams within schools participated
- Missingness at school level: 1 school started at T2

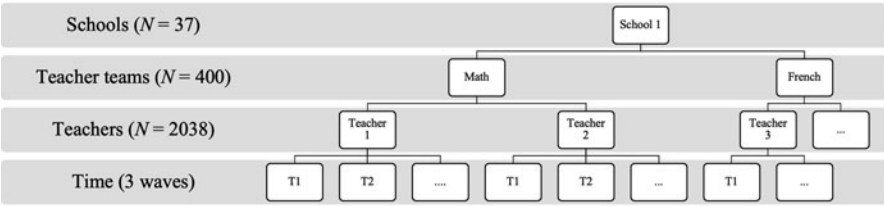

Fig. 15.1 Multilevel structure dataset

Table 15.1 Overview number of participants per wave

$N = 2038$	T1	T2	T3	All waves
T1 (*Nov–Dec 2015*)	**1604**			
T2 *(April–May 2016)*	1128	**1336**		
T3 (*Nov–Dec 2016*)	1000	948	**1278**	
All waves				**872**

15.3.3.4 Instruments

Both individual and team work engagement were measured using existing validated survey instruments. For individual work engagement, the shortened Utrecht Work Engagement (UWES) scale from Schaufeli and Bakker (2003) was used. This instrument contains three scales: vigour, dedication and absorption, each measured by three items. Team work engagement was measured using the UWES questionnaire for teams (Torrente et al., 2012a), which exist of the same three subscales. The difference between both instruments is the reference point in the items; for the individual work engagement, the participants are asked about themselves, while the team work engagement asks participants about their the team as a whole. For example, the individual item" At work, I feel full of energy" is rephrased to "During the task, my team feels full of energy". All items were answered on a 7-point Likert scale ranging from never (1) to always (7).

15.3.4 Visualising Longitudinal Multilevel Missingness

15.3.4.1 Goals of the Visualisation

The first goal of the visualisation is to provide an overview of how the (different kinds of) **missingness** affect the **patterns in the data**. This overview is meant to help decision making about the criteria that can be used for the inclusion of individuals and teams in the analysis. A second goal is to provide information about whether or not the **characteristics** of teams and individual teachers with (different levels of) missingness are different from those of the general sample. On a more abstract level, this means that the goal of the visualisation is to identify **features** in the data and to visualise **dependencies**. Features need to be identified because the goal is to find characteristics and patterns of individuals and teams that display (different types and degrees of) missingness and assess whether they are different from those of the rest of the individuals and teams. Since the patterns are dependent on the number of people and teams included that display a certain percentage of missingness, another part of the goal is to visualise dependencies.

15.3.4.2 Design Process

Exploration of Design Space As a first step in the design process, different existing visualisation methods were explored, starting from the key features of the data, including the hierarchical structure and time dimension. In addition, in order to examine patterns of missingness in the data, conditionality as well as dependencies need to be shown in the visualisation. Figure 15.2 provides examples of available options for each of the requirements. Each type of visualisation has its own benefits and disadvantages, however discussing all of these is beyond the purpose of this chapter. For further reading on this topic, we recommend Munzner (2014).

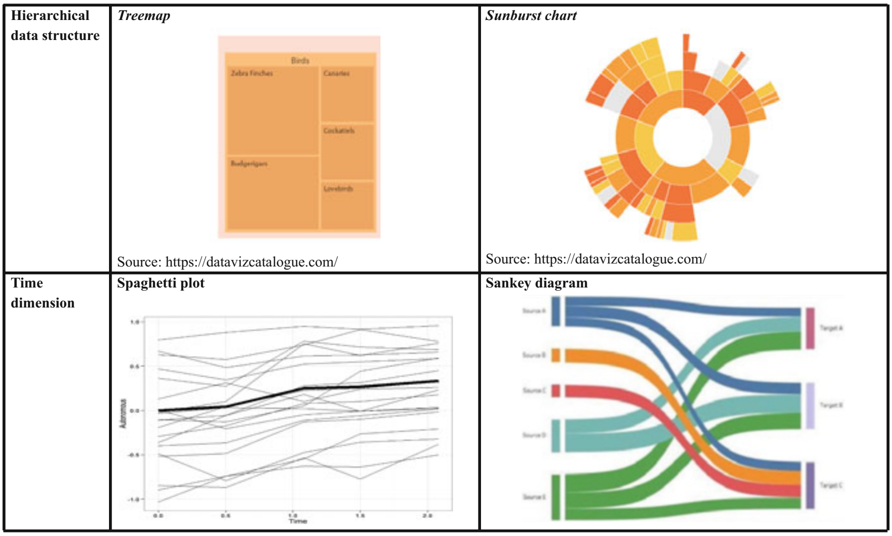

Fig. 15.2 Examples of possible visualisations

Sketching Phase In the second step of the design process, the visualisation designer went through a divergent and convergent sketching phase, in which they first sketch out a wide range of options for the visualisation. In the emergent stages of the sketching phase, these sketches were categorised according to what they display. The main categories were (1) visualisations that display the relationship between team work engagement and individual work engagement, (2) visualisations that display the flow of people from one level of engagement to another over time, (3) visualisations that show the group compositions of the teams, (4) visualisations that show how individuals score on individual work engagement and team work engagement, (5) visualisations that categorise teams in a certain way and then show individual work engagement and/or team work engagement for those categories, and (6) visualisations that display for each kind of individuals/teams what the pattern of missing data is. Some of the sketches fitted in multiple categories, which we also deemed the most promising sketches as they display more relevant information within one visualisation. In the convergent stages of the sketching phase, the designers started from the sketches that fitted as many categories as possible while also giving the clearest answers to the target questions. Subsequently, they explored how these visualisations could be enhanced by inputting ideas from the not yet included categories directly into the visualisation itself and/or by combining two visualisations.

Final Design In the final convergent stage of the sketching phase, it was decided that a **Sankey diagram** with **filters** (enabling to see a visualisation of only part of the dataset, e.g., only individuals that participated in every wave) should be the central part of the visualisation because it is able to show the patterns in the data in combination with missingness. However, because the Sankey diagram is not able to provide a simple overview of the data and does not allow for direct comparisons between individual work engagement and team work engagement, it was decided to combine it with a **line plot**. In addition, the answer options of the engagement variables were categorised to get more oversight within the Sankey plot itself. Furthermore, a filtering option was added to show only people within one of the categories in the first wave, which enables users to follow paths all the way from wave 1 to wave 3. For example, using filters, we can show a representation of the changes over time of people that scored low at T1 and compare it to the representation of the changes over time that scored high at T1.

Moreover, an **overview** window with several line plots alongside each other at different filtering settings is also provided to enable a user of the visualisation to quickly see which ranges of the filtering settings might be interesting to explore and thus interact with the visualisation as this interaction is at the core of visual analysis. Since the **background variables** of individuals are not visible within the Sankey diagram, the possibility to go on to another window was also provided. This new window compares the characteristics of the people that passed the filter or are selected, within a line or segment of one of the bars, with the characteristics of people within the overall sample.

Together, these elements of the visualisation enable the user to get a **global** idea of missingness: (a) what the missingness filters do to the data, if different patterns are visible depending on which part of the dataset is visualized, (b) a more **specific** idea of what the missingness filters do to patterns, and (c) an idea of what the missingness filters do to characteristics of people in the sample. This information gives the user a picture of how much missingness is tolerated before it impacts any trends or patterns. This information also shows the extent of the bias that is created by excluding people that are filtered out due to missingness.

Implementation The visualisation was realised with the Tableau software since this gave sufficient flexibility to implement the designed visualisation. As such, custom programming was not needed, and considerable time and effort were saved.

15.3.4.3 The Visualisation[1]

In the **overview window** (Fig. 15.3), the user can select the construct they want to display, as well as the missingness filters they want to use. The percentages represented vertically indicate the percentage of team members that completed the survey, while the NA filters in the top right enable users to include data from participants (individual and team level) that participated in specific waves. When

[1]The visualisation itself can be accessed via webadress and viewed using the free Tableau Reader software (see www.tableau.com). To see the interactivity within the visualisation, a screencast is also made. The screencast can be found at: https://www.youtube.com/watch?v=Lv5VR7HKeN4

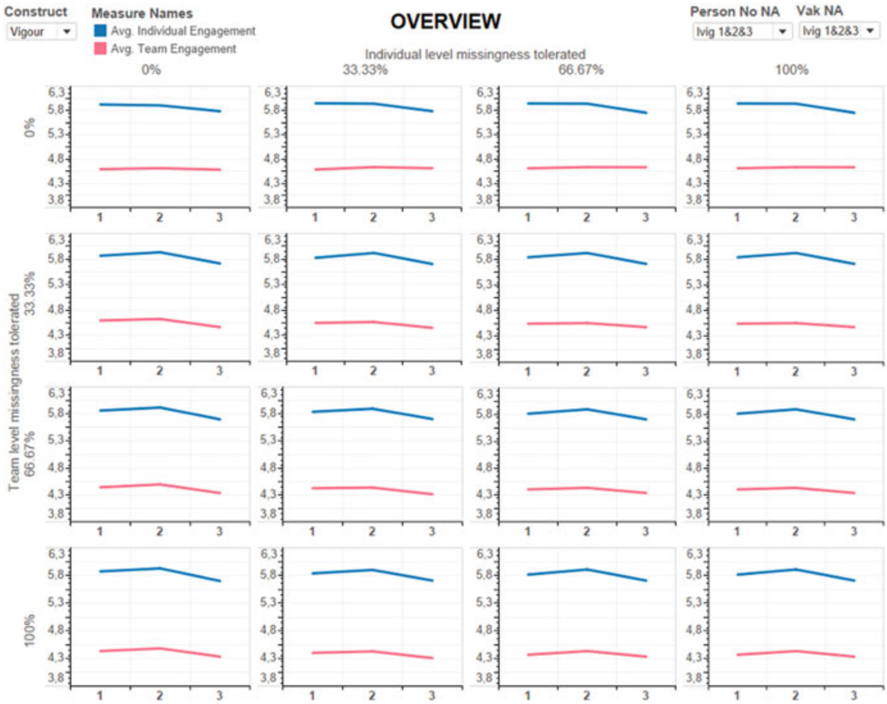

Fig. 15.3 Overview window visualisation

those are selected, an overview is given of team work engagement, and individual work engagement at different percentages of missingness tolerated.

In the **Sankey window** (Fig. 15.4), the user can explore a wider range of filtering settings. They can also explore in more depth the impact of different missingness cutoffs at the individual and team levels on the patterns in the data using the Sankey diagram. Similar to the overview window, they can select the construct they want to display and the filters they want to impose. In addition, they can also set the exact percentage of missingness they want to tolerate at the team level. As explained above, users can also selectively display the people that were in a certain category in the first wave so as to be able to follow the whole path of those people from waves 1 to 3. This can be done with the Sankey Path parameter.

In the **profiles window** (Fig. 15.5), the user can compare the characteristics of the people in the complete dataset (Right column ‘sample’) to the characteristics of the people that are included in the subset based on the filters that have been set or the people that are in a certain line or certain segment of a bar that is selected (Left column ‘Selection’). For example, Fig. 15.5 (a partial screenshot of the profiles window) shows that there are slightly less males in our subsample than there are in the complete sample. Users can go back to the Sankey window using a back button.

Fig. 15.4 Sankey diagram window visualisation

15.3.5 Findings

Using the visualisation, the domain experts developed the preliminary idea that **missingness at the individual level** does not have a large impact on the conclusions about the evolution across time. The trends seem to be the same regardless of how much individual missingness is tolerated. However, when only people are included that completed all three waves, patterns are more outspoken, and hence potential changes or dependencies might be inflated. In contrast, preliminary findings indicate that **missingness at the team level** does strongly impact the results, especially for team work engagement variables. If only teachers within teams for which all members participated are included, there is a positive trend towards the next school year for their team, while the opposite is the case when all teachers are included. An insight that would have been difficult and time-consuming to identify using traditional statistical methods.

Another insight the visualisation revealed is that there seem to be patterns in the **characteristics of people that drop out** of the study, which is similar for all engagement variables. We found that people who drop out come from all levels of engagement (high, medium, and low), which also explains why the impact of individual missingness is not so large when looking at patterns across time.

Fig. 15.5 Screenshot profiles window visualisation

Selection

Gender	
Men	35,98%
Women	64,02%

Avg. ErvS SJ1	14,85
Avg. ErvS SJ2	15,16
Avg. ErvTot SJ1	17,62
Avg. ErvTot SJ2	17,86

Interim MM1	
1	5,49%
2	94,51%

Interim MM2	
1	5,49%
2	94,51%

Interim MM3	
1	3,66%
2	96,34%

Vast MM1	
0	5,49%
1	87,80%
2	6,71%

Vast MM2	
0	5,49%
1	89,63%
2	4,88%

Vast MM3	
Missing Value	3,66%
1	90,24%
2	6,10%

Sample

Gender	
Men	37,51%
Women	62,49%

Avg. ErvS SJ1	14,15
Avg. ErvS SJ2	12,63
Avg. ErvTot SJ1	16,46
Avg. ErvTot SJ2	14,96

Interim MM1	
Missing Value	20,13%
1	6,82%
2	73,05%

Interim MM2	
Null	31,42%
1	6,53%
2	62,05%

Interim MM3	
Null	34,41%
1	5,01%
2	60,58%

Vast MM1	
Missing Value	20,13%
0	6,82%
1	64,51%
2	8,54%

Vast MM2	
Missing Value	31,08%
0	6,97%
1	54,54%
2	7,41%

Vast MM3	
Missing Value	39,52%
1	52,92%
2	7,56%

However, when the groups are explored in further detail, it can be seen that the highly engaged people who drop out are less experienced than the overall sample and also hold more temporary contracts in comparison with the overall sample. There are also slightly more males in this group. In addition, the medium engaged people who drop out are comparable to the overall sample. Finally, the low engaged people who drop out are more experienced and more often hold a permanent contract. There are also slightly more females in this group.

In addition, the visualisation also gave some **other important insights** that were not directly related to the goal of the visualisation. This illustrates how Visual Analytics can lead to serendipitous findings which were not defined beforehand. One of these insights is that the different engagement scales (vigour, dedication, and absorption) evolve relatively similarly across time. Another insight is that during the first school year, engagement seems to increase slightly while towards the second school year, engagement declines. This indicates that we should consider non-linear growth models when analysing growth models of engagement over time. A further insight is that teachers, on average, consider themselves more engaged than they consider their team as a whole. In summary, although all these effects still have to be tested statistically, the visualisation gives rise to interesting additional ideas and helps us as researchers to get the most out of our data.

15.4 Discussion

This chapter provides a high-level introduction to visual analysis as a way of approaching wicked problems and applies this perspective to the field of professional learning. Moreover, it presents a concrete example of how visual analysis was applied to explore multilevel missingness in a dataset collected to explore the relationship between individual and team work engagement over time. The current chapter focuses on the design process and functionality of the visualisation but only presents some preliminary findings that the visualisation yielded as an illustration of its value. However, it is important to note that in order for us to take an actual decision about how to deal with the impact with the multilevel missingness (e.g., excluding cases or deciding on which sort of missingness would be tolerated) as well as any conclusion around bias in our sample, our structured analysis protocol would need to be established and reported. This analysis protocol would maximise transparency and reproducibility of results, especially if several researchers execute the protocol and interrater reliability is calculated.

15.4.1 Contributions and Pitfalls of a Visual Analytical Approach

In preparation for this chapter, we searched the literature in the field of educational sciences for examples of visual analysis and identified which advantages and shortcomings they reported. In terms of advantages, various authors stated that visual analysis allowed them to identify hidden patterns in their data (Allaymoun, 2015; Bowers, 2010; Liu et al., 2016; Stikkolorum et al. 2015; Wong et al., 2016). It also facilitated complex data analysis or gave them the ability to analyse big data because the visualisation augmented their human capability (Bendinell & Marder, 2012; Bowers, 2010; Breuer et al., 2009; Li et al., 2015). They stated that visual analysis had a unique ability to generate hypotheses and potential explanations (especially in new or complex fields) but was less suited for the confirmation of hypotheses (Bendinell & Marder, 2012).

In contrast, named shortcomings entailed the steep learning curve to select and develop appropriate (custom) visualisations, potential visual overload, difficulties in generalising findings and the needed technical infrastructure (e.g., screen size and resolution, computational power) (Breuer et al., 2009; Charalampidi & Hammond, 2016; Hernández-García et al., 2014; Peretz, 2004; Stikkolorum et al., 2015; Wong et al., 2016)

15.4.2 Future Perspectives for Visual Analytics in Professional Learning and Development

Visual analysis has tremendous potential to drive new insights in the field of professional development. It enables researchers to generate hypotheses from large and potentially messy data sources that organisations have available. For example, more and more organisations are implementing extensive learning management systems and are interested in understanding the value this platform brings to their organisation (see Littlejohn et al., 2022). At this stage, however, they narrowly focus on (technical) engagement with the platform and completion of (compliance) courses. However, there is little insight into meaningful engagement with these platforms. For example, are the right individuals accessing the right resources? What does it mean if an individual is not active on the platform? And is this always something negative, or could it also be an indication that this individual does not have learning needs? Linking data from various platforms could enable us to answer this type of question but requires a different approach as our traditional statistical methods would fall short, while applying artificial intelligence without a good understanding of the data would not enable us to open this black box.

The examples above show the value visual analysis can have when working with large datasets, but also for smaller datasets, visual analysis can have its advantages, especially when the interest lies in identifying patterns and dynamics over time. At this stage, psychological and educational theories only provide limited true dynamic insights and principles. It is usually very time-intensive to decide which changes should be examined in which ways to be able to determine if changes or effects are short or long-term phenomena. In addition, identifying non-linear patterns using statistical methods requires many time points. Visual analysis cannot confirm the shape of the change over time but does enable the researcher to relatively quickly and easily compare different shapes given a set of parameters and conditions that can be imposed (such as the tolerance for missing data in our example). However, the use of visual analysis in smaller datasets is not limited to longitudinal datasets, visual analysis can offer interesting insights and starting point for any dataset that confronts the researcher with the conditions of finitude, complexity and finitude.

We agree with authors such as Hernández-García et al. (2014) and Wong et al. (2016) that researchers might need substantial training if they want to perform visual analysis by themselves, as such we do recommend engaging in interdisciplinary collaborations in which the data visualisation expert and domain expert work and learn together through an open attitude and mutual respect for each other's unique expertise, but mainly by asking lots of questions and never assuming the same term refers to the same thing.

Acknowledgements The authors wish to thank Amanda Winters and Carlos Eduardo Ortega for the technical development of the visualisation and Dr. Katrien Vangrieken for providing the data.

References

Allaymoun, M. H. (2015). Graphical visualisation of rhetorical structures in chats. *Conference proceedings of »eLearning and Software for Education« (eLSE), 01,* 407–414.

Alcaide, D., & Aerts, J. (2021). A visual analytical approach for the identification of ICU patient subpopulations using ICD diagnostic codes. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.430

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 243–268). Springer.

Bendinelli, A. J., & Marder, M. (2012). Visualisation of longitudinal student data. *Physical Review Special Topics-Physics Education Research, 8*(2). https://doi.org/10.1103/PhysRevSTPER.8.020119

Bowers, A. J. (2010). Analysing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment, Research & Evaluation, 15*(7), 1–18.

Breuer, R., Klamma, R., Cao, Y., & Vuorikari, R. (2009). Social network analysis of 45,000 schools: A case study of technology enhanced learning in Europe. In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Learning in the synergy of multiple disciplines. Lecture notes in computer science, 5794* (pp. 166–180). Springer. https://doi.org/10.1007/975-3-642-04636-0_18

Charalampidi, M., & Hammond, M. (2016). How do we know what is happening online?: A mixed methods approach to analysing online activity. *Interactive Technology and Smart Education, 13*(4), 274–288. https://doi.org/10.1108/ITSE-09-2016-0032

Conklin, J. (2006). *Dialogue mapping: Building shared understanding of wicked problems*. Wiley.

Edirisinghe, R., Stranieri, A., & Blismas, N. (2016). Information visualisation for the wicked problem of safe construction design. *Architectural Engineering and Design Management, 12*, 296–310. https://doi.org/10.1080/17452007.2016.1182890

Farrell, R., & Hooker, C. (2013). Design, science and wicked problems. *Design Studies, 34*, 681–705. https://doi.org/10.1016/j.destud.2013.05.001

Filliettaz, L., Garcia, S., & Zogmal, M. (2022). Video-based interaction analysis: A research and training method to understand workplace learning and professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 419–440). Springer.

Gibson, D. C., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education* (pp. 29–42). Springer International Publishing.

Hawryszkiewycz, I. T. (2014). Visualisations for addressing wicked problems using design thinking. In *Twenty Second European Conference on Information Systems (ECIS) 2014*. AIS Electronic Library.

Hernandez-Garcia, A., Gonzalez-Gonzalez, I., Jiménez-Zarco, A. I., & Chaparro-Pelaez, J. (2014). *Computers in Human Behavior, 47*, 68–80. https://doi.org/10.1016/j.chb.2014.10.0384

Jordan, M. E., Kleinasser, R. C., & Roe, M. F. (2014). Wicked problems: Inescapable wickedity. *Journal of Education for Teaching, 40*, 415–430. https://doi.org/10.1080/02607476.2014.929381

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the information age: Solving problems with visual analytics*. Eurographics Association.

Kyndt, E., & Baert, H. (2013). Antecedents of employees’ involvement in work-related learning: A systematic review. *Review of Educational Research, 83*(2), 273–313. https://doi.org/10.3102/0034654313478021

Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). Teachers’ everyday professional development: Mapping informal learning activities, antecedents, and learning outcomes. *Review of Educational Research, 86*(4), 1111–1150. https://doi.org/10.3102/0034654315627864

Li, X., Zhang, X., Fu, W., & Liu, X. (2015). E-learning with visual analytics. *IEEE Conference on e-Learning, e-Management and e-Services*, 125–130. https://doi.org/10.1109/IC3e.2015.7403499

Liebowitz, J. (2017). Thought on recent trends and future research perspectives in big data and analytics in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education* (pp. 7–17). Springer.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Littlejohn, A., Kennedy, E., & Laurillard, D. (2022). Professional learning analytics: Understanding complex learning processes through measurement, collection, analysis, and reporting of MOOC data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 557–578). Springer.

Liu, M., Lee, J., Kang, J., & Liu, S. (2016). What we can learn from the data: A multiple-case study examining behavior patterns by students with different characteristics in using a serious game. *Technology, Knowledge and Learning, 21*, 33–57. https://doi.org/10.1007/s10758-015-9263-7

Meredith, C., Schaufeli, W., Struyve, C., Vandecandelaere, M., Gielen, S., & Kyndt, E. (2020). ‘Contagion’ of burnout among teachers: A social network approach. *Journal of Occupational & Organizational Psychology, 93*, 328–352. https://doi.org/10.1111/joop.12296

Mulder, R. (2022). Advancing research on team learning by taking into account complexity, dynamics and context. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning – Insights from a growing field* (pp. 208–304). Springer.

Munzner, T. (2014). *Visualisation analysis and design*. Taylor & Francis Group.

Peretz, D. (2004). Using graphical profiles to study the learning and teaching of mathematics. *28th Conference of the International Group for the Psychology of Mathematics Education, 4*, 9–16.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.

Sakai, R. (2016). *Biological data visualisation: Analysis and design* (Doctoral thesis). Retrieved from KU Leuven Lirias Database (Lirias ID lirias123456789/521344).

Schaufeli, W. B., & Bakker, A. (2003). *UWES Utrecht work engagement scale. Preliminary manual* [Version 1, November 2013]. Utrecht University: Occupational Health Psychology Unit.

Seaman, S., Galati, J., Jackson, D., & Carlin, J. (2013). What is meant by "missing at random". *Statistical Science, 28*, 257–268. https://doi.org/10.1214/13-STS415

Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.

Stikkolorum, D. R., Ho-Quang, T., & Chaudron, M. R. V. (2015). Revealing students' UML class diagram modelling strategies with WebUML and LogViz. *IEEE 41st Euromicro Conference on Software Engineering and Advanced Applications*, 275–279. https://doi.org/10.1109/SEAA.2015.77

Thomas, J., & Kielman, J. (2009). Challenges for visual analytics. *Information Visualization, 8*, 309–314. https://doi.org/10.1057/ivs.2009.26

Torrente, P., Salanova, M., Llorens, S., & Schaufeli, W. B. (2012a). From "I" to "we": The factorial validity of a team work engagement scale. In J. Neves & S. P. Gonalves (Eds.), *Occupational Health Psychology: From burnout to wellbeing*. alth Psychology.

Torrente, P., Salanova, M., Llorens, S., & Schaufeli, W. B. (2012b). Teams make it work: How team work engagement mediates between social resources and performance in teams. *Pisicothema, 24*, 106–112. Retrieved from http://psicothema.com/PDF/3986.pdf

Van Wijk, J. J. (2005). The value of visualisation. *VIS 05. IEEE Visualisation, 2005*, 79–86. https://doi.org/10.1109/VISUAL.2005.1532781

Victor, B. (2011). *Up and down the ladder of abstraction* [Blog post]. Retrieved from http://worrydream.com/LadderOfAbstraction/

Winters, K. M., Lach, D., & Cushing, J. B. (2014). Considerations for characterising domain problems. *BELIV '14.* Paris.

Wong, G. K. W., Li, S. Y. K., & Wong, E. W. Y. (2016). Analysing academic discussion forum data with topic detection and data visualisation. In *IEEE international conference on teaching, assessment, and learning for engineering* (pp. 109–115). https://doi.org/10.1109/TALE.2016.7851179

Eva Kyndt (PhD) is full Professor Human Resource Development at Swinburne University of Technology (Centre for Transformative Innovation) in Melbourne, Australia. Her research focuses on the intersection between learning and working and includes topics such as the transition from education to work, workplace learning, sustainable careers, and social networks in organisations. She is the founding Editor-in-Chief of Teaching and Teacher Education: Leadership and Professional Development.

Jan Aerts (PhD) is guest professor at the universities of Leuven and Hasselt (Belgium), and is research fellow at Amador Bioscience. His main research topics revolve around visual design, interaction design, topological data analysis and (human and computational) scalability.

Part III
Research Approaches

Chapter 16
Delphi-Technique as a Method for Research on Professional Learning

Christian Harteis

Abstract The Delphi-technique, named after the antique Greek oracle, is one of the most important means to predict the future and is, thus, an important tool for research and also for educational research, particularly research on professional learning. This chapter describes the characteristics of the Delphi-technique, particularly its iterative approach that avoids group dynamic biases. The chapter explores its major (dis)-advantages against other forecast methods and describes major requirements for a Delphi-study, e.g. appropriate sample sizes. Then, an exemplary Delphi-study on the area of professional learning will be presented in very short. Finally, suggestions will be presented addressing researchers that consider conducting a Delphi-study.

Keywords Forecast · Group discussion · Experts

> Predictions are difficult – particularly with regard to the future. (Karl Valentin, a Bavarian poet and hero)

16.1 Introduction

Education aims at preparing students for the challenges of their future development. This particularly applies for children and adolescents in educational institutions that organize the transmission into the employment system and the distribution of workplaces. A major goal of primary and secondary education is the preparation of pupils for self-determined decisions on their future employment. However, education does not only aim at preparing pupils, also adults need to adapt to novel challenges in society and work. Hence, also adult and further education aims at preparing learners for future challenges. "*Non scholae sed vitae discimus*" (we do

C. Harteis (✉)
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: christian.harteis@upb.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_16

not teach for school but for life) is another saying that illustrates how much education tends to future circumstances. Thus, information about the future can be informative for educators and those who pursue educational policy.

The Delphi-technique, named after the antique Greek oracle, is one of the most important means to predict the future and is, thus, an important tool for research and also for educational research, particularly research on professional learning. This chapter describes the characteristics of the Delphi-technique, particularly its iterative approach that avoids group dynamic biases. The chapter explores its major (dis)-advantages against other forecast methods and describes major requirements for a Delphi-study, e.g. appropriate sample sizes. Then, an exemplary Delphi-study on the area of professional learning will be presented in very short. Finally, suggestions will be presented addressing researchers that consider conducting a Delphi-study.

16.2 Characteristics

In a most general description, the Delphi-technique is an iterative written procedure of group-inquiry. It can be considered as a quite strictly regulated communication process on topics about which no, just incomplete or insecure knowledge exists (i.e. the future). As a heuristic and intuitive procedure, it compensates for the lack of knowledge about the future by generating information about the future through purposeful communication with people that are supposed to have informed ideas about the future.

Within several steps of written interrogation, a group of people – mostly experts of a domain – receive questionnaires from a central research point (this may be a singular person or a research group), fill out these questionnaires and send them back to the research point. Researchers analyse the replies and merge them to a group opinion and send out this aggregated group opinion to the same group of people with new questions regarding this opinion.

16.2.1 Standard Procedure

The traditional idea of the Delphi-technique is to grasp expert opinion through repeated written interrogation of a group of experts. After each round of interrogation, the participating experts receive firstly anonymized feedback about the group opinion and secondly a task how to further deal with the group data. The aim is to strive for a consensus within the group through iterative steps of interrogation (Fig. 16.1).

At the beginning, participating experts receive an open questionnaire that explores the research topic (i.e. the future). Participants are encouraged to provide

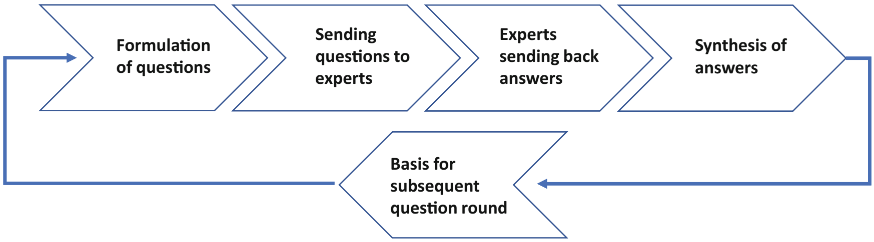

Fig. 16.1 Standard procedure of a Delphi-round

as many ideas about the topic as come into their mind and asked to send the questionnaire back to the research point. The aim of an opening round within a Delphi-process is to grasp as many ideas about the future as possible. Hence, the participants determine from the beginning the scope of content that may be considered as the forecast about the future.

It is the crucial role of the research point to organise and moderate the written group discussion. Its members develop the tasks, distribute the questionnaires and analyse and merge the findings of each round. This merge of findings shapes the basis for the next Delphi-round that again is organised by the research point. On this basis, the participants develop their opinion – despite of an anonymous circumstances of data gathering – in face of the group (i.e. their peers') opinion feedbacked after each round. Participants usually are requested to acknowledge the group opinion and to reconsider their own approach. This approach is considered to contribute to a group consensus without neglecting singular positions.

Step by step respectively round by round, the questionnaires become narrower with each round, possibly to a closed design, e.g. by giving a selection task in the second round (e.g. selection of top five ideas of round 1), and providing a ranking task in round 3. The research point holds the responsibility to focus the research goal (i.e. future forecast) across all rounds by designing the questionnaires.

16.2.2 Group of Participants

The quality of the outcome of a Delphi-process is necessarily depending on the knowledge of participants on the topic. The more knowledge the participants have, the better is the quality of the outcome. Hence, the most crucial part of a Delphi-study is the selection of participants – usually experts in the domain. The focus on experts implies that deep insiders of a domain have already thought more often and more elaborate on future development of a domain than others, and that they are the best persons to provide judgements under complete uncertainty of future for a domain. Given it would be possible to provide a perfect forecast, it would require considering an unrulable number of variables including their interrelations. Such a task is impossible to process for human spirit as well as (so far) for artificial intelligence. It is a crucial characteristic of experts to be able to cope with bounded rationality (Gruber & Harteis, 2018). Hence, experts are the most appropriate target group to be involved in Delphi-studies.

Delphi-studies do neither require representativity nor statistical significance, the number of participants usually varies between 10 and 30 experts (e.g., Brooks, 1979; Duffield, 1993), but there are also exceptional Delphi-studies with several thousand participants (Cuhls et al., 1998). In the mid of the twentieth century, several, partly experimental studies have been conducted to investigate the reliability of Delphi-forecasts. Dalkey (1969) revealed that the average group error decreases with an increasing number of participants. However, Linstone and Turoff (1975) demonstrated that the quality of Delphi-forecasts requires at least 15 participants, more participants do not notably contribute to a further increase of the forecast quality. Hence, the Delphi-technique requires just a moderate number of participants in order to generate a good forecast. However, they need to be selected by care, e.g. with regard to a variety of viewpoints. There is no standard rule but the selection depends on the research goal.

16.2.3 Iterative Procedure and Feedback

Analyses (e.g., Linstone & Turoff, 1975) revealed that experts hold a larger amount of knowledge than they activate when working on the first open questionnaire. The iterative procedure with the feedback of the group opinion stimulates the activation of experts' further knowledge, and it encourages participants to reconsider their opinion in face of the group opinion. "Basically, the controlled feedback process (. . .) allows each participant an opportunity to generate additional insights and more thoroughly clarify the information developed by previous iterations" (Hsu & Sandford, 2007, p. 2). Hence, it's plausible to assume that the group judgement becomes more valid with each additional Delphi-step. However, Linstone and Turoff (1975) revealed that the quality of group judgements only slightly increases after a fifth Delphi-step. Hence, literature on the Delphi-technique agrees that the average of rounds of usual Delphi-studies varies between three and four (Worthen & Sanders, 1987).

The Delphi-technique guarantees the consideration of each expert judgement, as extraordinary or singular it may be. A singular opinion contributed within the first Delphi-step may be taken by other participants in the second step. Similarly, a broadly shared idea of the first step can be refused in the second step. The formation of opinion during the Delphi-process, thus, is a highly dynamic process.

16.2.4 Anonymity Between the Participants

Throughout the entire Delphi-process (i.e. including the analysis of the last Delphi-round), the anonymity between the participants remains protected. It is the research group as central entity that controls and manipulates the full communication between participants by merging singular inputs and providing feedback of the group opinion. There are several reasons to be mentioned that such a Delphi-process leads to a more precise group-opinion than any other form of direct, synchronous (i.e. face-to-face as well as online) group discussion:

- The anonymous setting of data collection (i.e. participants fill out questionnaires within their regular environment) reduces the peril of group-dynamic biases. When working on the questionnaires, participants sit in their regular environment without any contact either to other participants or the research group. This setting prevents eventual interferences through dominating discussion partners and reduces peer pressure (Hsu & Sandford, 2007; Landeta, 2006).
- In contrast to alternative discussion formats, the Delphi-procedure guarantees that each singular opinion gains similar weight and importance. "Often, the group opinion is essentially determined by the opinion of the individual who talks the most" (Dalkey, 1968, p. 7).
- The iterative procedure reduces the probability of bandwagon effects, because anonymity reduces opportunities to adumbrate the prevailing position (Linstone & Turoff, 1975) as well as . . .
- . . . and influence of social desirability (Brown, 2018).
- Anonymity encourages honest utterances and relieves taking up positions that may be unpopular or off a mainstream, relieves contradicting other opinions but also relieves modifying the own opinion because there is no loss of face (Isaac & Michael, 1995).

16.3 Challenges, Advantages, Disadvantages

The description of major characteristics already revealed crucial differences to alternative formats of discussions, questionnaires and interviews. A major advantage of the Delphi-technique in contrast to face to face discussions and interviews is the avoidance of group dynamic influences, a major advantage in contrast to questionnaire studies is the flexibility of the entire Delphi-process which always can be adjusted by the research group.

16.3.1 Challenges

The Delphi-technique holds certain challenges that refer to the implementation of a Delphi-study and to the appraisement of its result.

The quality of a Delphi-forecast is necessarily related to the selection of participants that usually are considered as experts. However, depending on the domain it might be more or less clear who can be considered as an expert. In highly selective domains (e.g. professional sport, arts, medicine) quite reliable performance indicators might exist – e.g., playing strength indicator in chess (ELO number) or tennis (international ranking list) but are probably exceptions. Additionally, salaries can be considered as an indicator of expertise in selective domains – nota bene for the selection of participants for Delphi-studies. Undoubtedly, there are many persons to be found who bear a lot of knowledge on a domain and do not appear at (the top of) such rankings. However, the safest way to cover a wider and well-reflected amount of knowledge about a domain is to consider such indicators. Within scientific domains, publication indices or awards can be considered as indicators, too, without neglecting that there are much more experts to be found in these domains who do not show top values in this respect.

However, there are also professional domains that do not provide any kind of reliable performance indicators (e.g. teaching, counselling, carework). Of course, work experience can be considered as an indicator of expertise but discussions about expertise research often criticize such a pragmatic approach (researcher on expertise often apply a 10-years-rule as selection criterion for expertise). A more elaborate approach plies peer nomination but is of course related with a high effort to additionally recruit volunteers for a peer nomination. However, also selected research on expertise tries to overcome the weakness of a pragmatic way by initiating peer nomination (e.g. Berliner, 2001).

As a typical start of Delphi-studies comprises a first round of open questions. Open questions, however, require more mental elaboration than closed questions. Thus, the start of Delphi-studies requires a high level of motivation from the participants. However, this challenge might be counterbalanced through the experts' interest. Anyway, Delphi-studies usually suffer sample mortality in a similar amount as longitudinal studies do. Since participation on Delphi-studies is voluntary, a certain loss of participants is impossible to avoid. Means to avoid such a loss are to be discussed at the end of this chapter.

16.3.2 Advantages

The attractivity of the Delphi-technique results from a series of advantages. A major issue here is its adaptivity: Originally developed – and mainly used – for forecasts, it can be applied for each imaginable research question that aims at opinion formation about a topic to which no objective and clear state of the art is available. Many

Delphi-studies were on the area of technical development (e.g., Gordon & Helmer, 1964; Cuhls et al., 1998), but there are also Delphi-studies in the area of educational research, e.g. for the development of curricula for secondary education (Häußler et al., 1980) or workplace learning (Harteis, 2002). As the research group collects the input of all participants and develops the questionnaires for following rounds, the research group can control the discussion process and direct it in the way of the research question. Additionally, a Delphi-study is also flexible with regard to spatial issues. The entire communication occurs in a written way (paper and pencil as well as on the electronic way) which means that the processing of questionnaires occurs remotely.

A further advantage is that Delphi-studies spare financial budgets. There are no travel costs necessary nor do they require a big bundle of technological or material resources. Paper and pencil studies only require paper, print and postage costs, if conducted electronically via email or browsers, they do not even require those costs. Hence, even Delphi-studies with a higher number of participants are cheap and economical.

Even though that Delphi-studies suffer the for studies with repeated data gathering typical dropout of participants, the quality of the final result is quite robust against such dropout, because all efforts of the Delphi-process aim at focusing on the group consensus in sense of largest intersection of different opinions. The loss of singular contributions does not seriously harm as long a certain number of participants (15 persons are considered as minimum) manage to agree.

From participants' perspective, the Delphi-technique provides two advantages. Firstly, it allows insight into the authentic opinion of peers, because the reciprocal anonymity encourages honesty of answers. This insight on the group opinion, then, offers participants the opportunity to reflect their own position – which is often the cause to modify viewpoints. Secondly, Delphi-studies provide (relatively fast) feedback after each round.

Finally, the Delphi-technique allows to conduct and finalize an empirical study quickly. The research group holds control on the pace of the study since it conducts the analyses, develops the feedback and sets deadlines for the processing of questionnaires.

16.3.3 Disadvantages

Of course, as any other research method, the Delphi-technique also bears disadvantages. A major concern against results of Delphi-studies refers to their lack of reliability. Delphi-studies have to be considered as group discussions on subjective opinions regarding a particular topic. This implies that starting the same discussion on a different moment or with a different sample, it is highly probable that a different outcome will result. However, as described in the section of characteristics, there is no claim that Delphi-studies produce reliable and generalizable results, but the claim is that Delphi-studies provide better judgements on (future) development than other forecast techniques do. HRD practitioners may apply procedures like scenario-technique or SWOT analysis (Strength-Weaknesses-Opportunities-Threads) in order to derive requirements for professional learning. However, these procedures mainly are applied in consulting or HRD context rather than in research on professional learning.

Further, participation on a Delphi-study requires a high level of commitment and frustration tolerance, because – depending on the design of the questionnaires – open questions, particularly in round 1, may require a lot of reflection and efforts to put complex thoughts into appropriate words. A formulation of a question may challenge a participant's opinion which can lead to dropout if the commitment isn't high enough.

Complementary to the time argument as an advantage, time can also develop as disadvantage, particularly when conducted as paper and pencil study. For sure, a Delphi-study can be conducted quickly online, but if participants do not reply on time, it becomes necessary to remind and encourage them to submit their questionnaires. In a paper and pencil version, the postal way requires a certain time, and participants have to take (and find) a timeslot to fill out the questionnaires. Typical seasons of the year (e.g., summer holidays, Christmas break) may cause additional delays, too.

Discussions on disadvantages of Delphi-studies often refer to the strong dependency of the quality of its outcomes and the quality of the selection into the sample (e.g. Jorm, 2015; Linstone & Turoff, 1975). The Delphi-technique particularly focuses explicit and implicit knowledge on a domain, and it is this knowledge that is considered to contribute to the quality of the outcome. However, this applies not only for the Delphi-technique but for all empirical ways of data gathering that the selection of test persons, interview partners or subjects of observation that the quality of the sample selection has crucial impact on the quality of the outcome.

16.4 An Example of a Delphi-Study on University Didactics Development in Germany

Delphi-studies in educational areas often address future developments for curricular or professional requirements, and so does the example of a four step Delphi-study provided here. The example comprises descriptions of preparation and organization processes the organization of the four rounds and a reflection of the outcomes (Paetz et al., 2011).

16.4.1 Preparation

The area of research was university didacts in Germany. Since the so-called Bologna reforms started to transform the European landscapes of universities, the challenges for German universities were particularly high, because there was no tradition of (consecutive) three-circles study programs (i.e. BA, MA, PhD) but rather a tradition of four to 5 years study programs with degrees that were incompatible with or even unknown in other European countries. The main aims of the Bologna reforms aimed at establishing a European landscape of university programs that allow students flexible changes in order to increase European mobility and competitiveness (Eurodyce, 2009). Major goal of these university reforms has been to transform the national systems into a commonly shared framework of tertiary education until 2010.

16.4.1.1 Formation of Research Group and Development of Research Goals

The project group, which comprised five researchers, was formed during 2008 exactly in a time when these changes occurred and raised uncertainty across all academic staff. Hence, the application for the research grants used for this study already anticipated the decision for a Delphi-study. However, the discussions started with formulating research goals which were the development of a competence framework for academic teaching that fulfill the requirements of an European landscape of study programs. During the 2000s, there was a controversy if the new framework of study programs – that then also included module handbooks, descriptions of study outcomes and students' performance requirements in terms of time invested – allow to stick to traditional ways of teaching of if (completely) new requirements will arise. Within this context, the goal was formulated to cover the full breadth of academic teaching which not only consists of teaching hours but also of assessment practices and administrative duties that contribute to quality management for study programs. It became clear that the study will address quite a broad focus. In order to be able to explain the research ideas more concretely, a number of (still relatively abstract) research question were defined:

- What are the future relevant skills and competences that enable university teachers to steadily perform on a high level?
- What are the consequences particularly of the Bologna reforms for the requirements that university teachers have to fulfil in future?
- What results of these consequences for the academic staff that works in departments of university didactic?
- How can university didactic address current and future challenges of transformation?

16.4.1.2 Recruiting of the Sample

For the recruitment of experts, two strategies applied. A first check of departments of university didactics revealed that the vast majority of permanent staff was well established researchers on higher education. Hence, a first search addressed specialists in the field by scanning databases for authors of German and international publications. On this basis, a list of the authors with most publications on higher education and university didactics was created. As second search strategy, peer nomination was applied. For this reason, all directors of departments of university didactics in Germany, Austria, and Switzerland have been contacted and asked to nominate three experts from German speaking countries. These directors nominated academic staff as well as freelance trainers in higher education.

Initially, the plan was to recruit 30 experts for the Delphi-study. After receiving both lists, the decision was made to invite the 25 highest ranked experts from both lists. Due to a high degree of overlapping, 34 experts were invited, thus. As Table 16.1 reveals, not all initially invited, and participation fluctuated and decreased finally.

After only 30 of 34 experts reacted on the first invitation, only those were invited for the following rounds. However, one expert who was not able to reply to first round but indicated further interest was additionally invited for the following rounds.

The participants were asked for some personal data. Of the 31 experts participating round 1, 16 were female and 15 male. Their average age was 50.3 years (SD = 10.0). Work experience in the field of university didactics was 16.9 years (SD = 11.4). 28 experts described that they regularly provide teaching for university didactics, 27 conduct research on higher education and university didactics, and 22 also acted as freelancers in the area providing courses and counselling. 25 indicated their publication output: The mean of publications was 48.0 (SD = 59.2), the median – which is more stable against outliers – was 30 (quartile 1 = 15, quartile 3 = 50). The publications on research on university didactics showed a mean of 33.6 (SD = 47.9; MD = 20.5; Q1 = 9.5; Q3 = 40).

The vast majority passed study programs in the humanities or social sciences, but there were also six experts with degrees in engineering, sciences or business management.

Looking at these data, the sample was a promising selection of groups for a Delphi-study.

Table 16.1 Response rate across the four rounds

Round	Number of invited experts	Participation
1	34	30 (88%)
2	31	26 (84%)
3	31	28 (90%)
4	31	20 (65%)

16.4.1.3 Designing the Study

Due to the nature and variety of research questions, a four step Delphi-study was planned, and this plan was clearly communicated to the participants (Fig. 16.2).

To anticipate a result that was unexpected: Many experts declined in the opening round that the Bologna-reforms really require novel skills and competences, but they require a different priority of those. The further development of the study should confirm this first estimation.

For the first round, open questions required open answers about the experts' understanding about requirements in academic teaching, assessing and in academic administration. The feedback was analysed through qualitative content analyses. The second round aimed at an aggregation of data in order to allow a selection of most important contributions from round 1. In round 3, this selection was to prioritized. Additionally, scenarios for future development were developed from the open

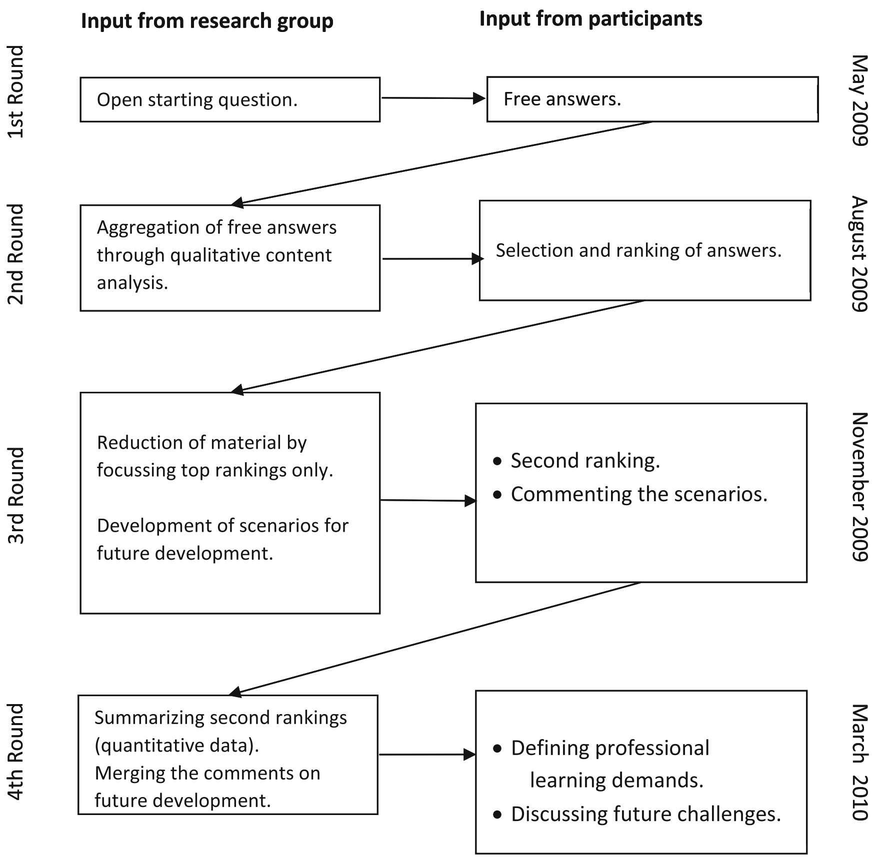

Fig. 16.2 Workflow plan for the study

questions in round 1 through the research group. These scenarios needed to be judged. Finally, round 4 aimed the development of a group consensus.

16.4.2 The Four Delphi Rounds in More Detail

The detailed description of the four rounds explains the concrete questions and tasks for the participants, the analytic work of the research group and a very brief presentation of major outcomes.

16.4.2.1 Round 1: Understanding Competence

The goal of the opening round was to grasp the experts' understanding of skills and competences, and to learn about their general assumptions about skill requirements of academic teachers, particularly when considering the transformations through the Bologna reforms. However, they were asked not only to consider their closer working area but the entire variety of an university.

Therefore, the participants received a list of optional skills and competences including an exemplary operationalisation. They were asked to select a random number of skills and competences, that they consider (a) important for university didactics and (b) for academic teachers after the completion of the Bologna reforms.

The full data were analysed by qualitative content analysis (Mayring, 2004). In total, the participants provided 1382 singular suggestions of different operationalizations. In a first step of analysis, redundant suggestions have been deleted, and individual descriptions have been paraphrased and deductively clustered correspondingly. After this step, 975 suggestions of important competences remained, that have now been inductively clustered to a final frame of 57 suggestions of skills and competences that were related to the three areas of academic teaching introduced by the research group: Instruction, assessment, and academic administration.

Not surprisingly, the understandings of the term *competence* referred to well-established positions presented in (German) research literature: *Competence* is capability and the will to act in professional context in order to solve problems, based on skills knowledge, proficiency, experience, and orientations (Butler, 1978).

Table 16.2 shows – just for illustrative purposes – the three most often and the three fewest selected and operationalized competences in the list of 57.

From an observation perspective the choice of fewest selected competences may surprise, since they doubtlessly belong to important characteristics of academic teachers. However, at least the fact that these characteristics have been chosen by anybody can be read of indicator of any importance. The full list has been taken into the tasks for round 2.

Table 16.2 Top and bottom of the number of selection

Suggested competence	Number of selections
1. Assessment competence	48
2. Use of teaching methods	45
3. Coping with determining factors	36
. . .	
55. Supporting development of personalities	3
56. Project management	3
57. Responsibility	3

Table 16.3 Results for instruction

Suggested competence	Number of selections
1. Use of teaching methods	21
2. Supporting independency	16
3. Know-how and expertise	16
. . .	
21. Enthusiasm	9
22. Innovation	8
23. Individualization	8

Table 16.4 Results for assessment

Suggested competence	Number of selections
1. Assessment competence	25
2. Feedback quality	23
3. Empathy	21
. . .	
17. Enthusiasm	10
18. Innovation	9
19. Individualization	8

16.4.2.2 Round 2: Selection

The goal of the second round was to reveal, which of these 57 competences were considered most important for each of the three areas of academic teaching. Hence, the aim of this round was to reduce the list through selection by importance. The concrete task for the participants in this round was:

> Please choose for each of the three areas of academic teaching – i.e. *instruction*, *assessment*, *academic administration* – 15 of these 57 competences that you consider most important for future academic teaching.

Therefore, the participants received for all of these three areas the list of the 57 competences, combined with operationalizations and anchors that fit best to these three areas. This way, it was possible simply to count the number of selections in this round in order to learn which competences have been conserved most often.

Tables 16.3, 16.4, and 16.5 show again the most often and fewest chosen competences in round 2.

Table 16.5 Results for academic administration

Suggested competence	Number of selections
1. Abilities of cooperation	23
2. Communication skills	22
3. Leadership	19
. . .	
21. Ethical attitudes	10
22. Evaluation competence	9
23. Planning skills	8

By this, the analyses for round 2 have been completed that provide participants the information basis for round 3.

16.4.2.3 Round 3: Weighting

The third round followed two goals. Firstly, one aim of this round was to weight the results of the former round and to come to a group opinion which ten competencies for each area of academic teaching were considered most important. Therefore, the first task was:

> Please create for each of the three areas of academic teaching – i.e. *instruction*, *assessment*, *academic administration* – a ranking list of the ten most important competences. Please number your ranking list from 1 to 10 and use 1 as most important item.

Secondly, a part of the research interest was to judge the consequences of the Bologna-reforms. Resulting from the answers in round 1 in which participants described their view on the consequences of Bologna, the research group developed a 3 × 3 matrix of scenarios that sketch (possible consequences) of Bologna: For each of the areas of academic teaching a scenario with positive, negative, and no serious consequences. The task here was:

> Please provide your personal viewpoint on each of these nine scenarios as concrete as possible by giving comments to each of them. Their perhaps provoking formulation is intended in order to make argumentation and a delineation easier. As a final personal conclusion, please judge the consequences of Bologna reforms general on a rating scale ranging from −2 (very negative) to +22 (very positive).

The analyses, hence, comprised a quantitative transformation of ranking places into points by turning the scale (rank 1 = 10 points, rank 2 = 9 points and so on).

Tables 16.6, 16.7, and 16.8 show the five competences that were considered most important for each area of academic teaching.

To remark here: When comparing to Table 16.4, one can see a significant change f group opinion. Whereas *enthusiasm* scored third lowest in round 2, it ended no at rank 5 in round 3.

Regarding to the experts' judgement of the consequences of Bologna-reform, the general judgement on the ranking list ended with a mean of +0.5 (SD = 1.7), hence, in tendency positive. The inductive categorizations of the reaction towards the

Table 16.6 Most important competences for instruction

Suggested competence	Points
1. Use of teaching methods	161
2. Know-how and expertise	160
3. Supporting independency	143
4. Self reflection	98
5. Enthusiasm	97

Table 16.7 Most important competences for assessment

Suggested competence	Points
1. Assessment competence	222
2. Feedback quality	150
3. Know-how and expertise	113
4. Communication skills	111
5. Counselling	100

Table 16.8 Most important competences for academic administration

Suggested competence	Points
1. Abilities of cooperation	144
2. Coping with determining factors	126
3. Innovation	117
4. Staying power	110
5. Communication skills	108

scenarios reveal that criticism refers rather to ways that the reforms have been implemented than to the goals of the reforms. The open answers can be summarized here just in following hypotheses:

- Academic teaching with all its facets becomes more important for the success of universities.
- All academic programs will apply output and competence orientation.
- The perception of individual degrees of freedom for teaching under the regime of module descriptions differs widely and depends on the individual situation.
- The general idea of modularization is appreciated but its implementation is often a problem.
- The general aims of the Bologna-reforms with regard to instruction are appreciated but the implementation often reproduces traditional practices.
- Course-related assessment is appreciated.
- However, the increasing number of assessments is a problem.
- An important function of academic administration is quality management and the provision of services, academic staff gains additional managerial tasks.
- Bologna supports the development of profiles.

16.4.2.4 Round 4: Demand for Further Education

After having developed a future-oriented competence profile, the goal of round 4 was to identify the demand for training and further education. The task for the participants to choose ten competences across all three areas of academic teaching for which they see academic teachers least prepared and, thus, highest demand for training and education. Therefore, the participants received the full list of competence suggestions from round 3 and could indicate their choice by marking with a cross. The analysis simply comprised the counting of crosses. Tables 16.9, 16.10, and 16.11 show the five most often chosen competences for the three areas of academic teaching.

In total, the highest number of selections covered instruction (71), followed by academic administration (59) and assessment (58).

16.4.3 Reflection of Outcomes

As a first, general interpretation of the outcomes, it is to acknowledge that the participants in general judge the Bologna reforms positive in their tendency. That was quite a surprising outcome, because these reforms have been discussed very critically beyond academic staff whereas educational policy and executive boards of

Table 16.9 Highest training demand for instruction

Suggested competence	Number of selections
1. Competence-orientation	13
2. Knowledge on teaching methods	11
3. Supporting independency	11
4. Individualization	9
5. Use of teaching methods	7

Table 16.10 Highest training demand for assessment

Suggested competence	Points
1. Feedback quality	11
2. Competence-orientation	11
3. Assessment competence	10
4. Counselling	9
5. Coaching for learning	6

Table 16.11 Highest training demand for academic administration

Suggested competence	Points
1. Moderation skills	13
2. Leadership	12
3. Problem-solving capabilities	12
4. Abilities of cooperation	12
5. Communication skills	5

universities were open-minded in 2010. At least, it was their task to assert these reforms against resistance from widely traditionally socialized academic staff.

Along the results of round 3, it was possible to define a model of competences that are important for the future. Looking at details, there is no sensation to be found because the singular competence suggestions all are widely discussed for a long time. However, since at the beginning a list of 59 competence suggestions were presented for selection form which each of them appears plausible and reasonable, the end result comprised (only) 30 – from which the tables above only displayed the half. Here, the basic assumption of Delphi-studies applies: When focusing important competences for academic teaching and university didactics, the outcomes of the Delphi-study is better information than focusing, e.g., those competence suggestions that were sorted out during the Delphi-process.

Additionally, the fourth round provided a well-grounded judgement on demands for training and further information. That probably is the most interesting result for the target group of this study, namely staff at departments for university didactics. And, from a later perspective it is to acknowledge that the offer of training for university teachers steadily increased during the last years (e.g., Jaramillo-Baquerizo et al., 2019; Ödalen et al., 2019).

16.5 Suggestions for Conducting Delphi-Studies

The Delphi-technique is a very flexible instrument which allows adaptations at each round which can arise from researchers' insight interests but also from participants' input. It may require high workload at particular stages of the Delphi process, especially when tasks comprise open answers. As described above, in order to fully realize the advantages of Delphi-studies, some prerequisites have to be given. However, a real "gold standard" of Delphi-studies does not exist (Powell, 2003). This final paragraph provides suggestions that support the realization of advantages and that may assist a smooth implementation of a Delphi-study.

16.5.1 Research Group

The composition of the research group is crucial for the success of every research endeavour. Even though it is of course technically possible to run a Delphi-study with a singular person, it is not wise to do so – at least if the classical approach is chosen with an open and wide beginning which implies qualitative analyses. The experts' contributions can be very heterogeneous and comprise a variety of ideas and interpretations. A multiple view on these contributions is very helpful. Hence, it is wise to establish a research group with a manageable number of people, typically between 3 and 5 persons.

Delphi-studies do not require complex statistical analyses but – at certain moments – delicate but crucial qualitative analyses. Hence, members of the research group ideally hold experience in qualitative data analysis. Expertise in the domain which is investigated, though, is less important. A high level of expertise may even cause biased interpretations of some extraordinary inputs through the participants. However, a certain familiarity with the domain is necessary in order to understand input adequately.

16.5.2 Research Process

The formation of a group consensus requires several steps of interaction. Hence, Delphi-studies should at least consist of three rounds in order to enable participants to overthink and modify their original opinion in a reasonable amount. However, Linstone and Turoff (1975) evaluated the quality of several Delphi-forecasts and state that the quality does not gain quality notably after a fifth round. Literature on Delphi-studies suggest three or four rounds as sufficient for the consolidation of a consensus (Worthen & Sanders, 1987).

The analyses of input and the appropriate preparation of documents for the participants may require a certain amount of time. In times when only paper and pencil communication via post, a regular time interval between two rounds were 3 or 4 months (Gordon & Helmer, 1964). Nowadays, of course, interaction via emails and web interfaces accelerate sending process, sometimes also analyses if quantitative online questionnaires are implemented. However, qualitative analyses require time as the preparation of the communication with the participants does. Additionally, participants may require a certain while for processing the tasks, and follow up inquiries may become necessary if participants miss deadlines. Hence, it is important to develop an adequate schedule which does not need to be revised during the running study. However, electronic communication is more effective than paper-based interaction (Brown, 2018).

16.5.3 Composition of the Sample

The composition of the sample is of similar importance than the composition of the research group. It is of highest importance that the participants hold expertise in the domain that is to be investigated. As discussed above, domains differ in their clearness that indicates expertise. Where clear indicators exist, the selection of persons that should be invited to participate is not a challenging task. However, if such clear indicators do not exist, it is important to put high efforts into the nomination process. A multiple strategy appears appropriate (Landeta, 2006), the study presented in the third paragraph provides an example.

There are two reasons to finally recruit a certain amount of experts for a Delphi-study (Iqbal & Pipon-Young, 2009): Firstly, they suffer a similar sample mortality as all studies with multiple measurements do; secondly, in order to grasp a wide range of experts' (implicit) knowledge, a number of experts need to participants across all rounds. Hsu and Sandfort (2007), thus, recommend to recruit number of about 30 experts in order to be able to expect that the materials of all rounds are processed by at least 15 experts – the number that Linstone and Turoff (1975) claimed as prerequisite for a good forecast.

16.5.4 Interaction with Participants

For the recruitment as well as for the communication in each round, an appropriate communication with the participants is crucial (Flostrand et al., 2020). It is important to describe the aims and objects of research in a convincing and an appealing way, to be clear with regard to the expectations on participants' commitment and to provide a reliable time schedule. Violations of declaration or (frequent) modifications of the schedule cause disappointment which may end in breaking off participation (Landeta, 2006).

The quality of feedback is a further factor of similar importance (Kennedy, 2004). One of experts' usual motivation to participate Delphi-studies is that they have interest in learning about peers' opinion in their domain. Hence, the feedback of the results of the former round is not only the fundament for the new questionnaire but also informative for the participants. Additionally, it is important to include singular or outsider opinions in the feedback in order to make sure that all participants can recognize their own contribution to the group opinion.

16.6 Conclusion

This article aimed at describing the Delphi-technique as an appropriate instrument for research on professional learning and development. Due to globalized markets, economic and technological development, and digitalization, the future of work and employment holds many future challenges for which currently no reliable information exists. There might arise increased interest in research that, firstly, provides an outlook on this development and its consequence. Secondly, it might also become more important to get insight on future demands for professional learning. A Delphi-study is a promising instrument to explore these challenges in a way that is considered to be superior compared to alternative ways of forecast generation.

References

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*, 463–482.

Brooks, K. W. (1979). Delphi technique: Expanding applications. *North Central Association Quarterly, 53*(3), 377–385.

Brown, J. (2018). Interviews, focus groups and Delphi techniques. In *Advanced research methods for applied psychology: Design, analysis and reporting* (pp. 95–106).

Butler, F. C. (1978). The concept of competence: An operational definition. *Educational Technology, 18*(1), 7–18.

Cuhls, K., Blind, K., & Grupp, H. (1998). *Delphi `98 Studie zur globalen Entwicklung von Wissenschaft und Technik*. Fraunhofer Institut für Systemtechnik und Innovationsforschung.

Dalkey, N. C. (1968). *Predicting the future* (RAND Corporation Paper P-3948). RAND.

Dalkey, N. C. (1969). *The Delphi method* (RAND Corporation Paper RM-5888-PR). RAND.

Duffiled, C. (1993). The Delphi technique: A comparison of results obtaining two expert panels. *International Journal of Nursing Studies, 30*(3), 227–237.

Eurodyce. (2009). *Bologna with student eyes 2009*. ESU.

Flostrand, A., Pitt, L., & Bridson, S. (2020). The Delphi technique in forecasting – A 42-year bibliographic analysis (1975–2017). *Technological Forecasting and Social Change, 150*, 119–173.

Gordon, T. J., & Helmer, O. (1964). *Report on a long-range forecasting study* (RAND Corporation Paper P-2982). RAND.

Gruber, H., & Harteis, C. (2018). *Individual and social influences on professional learning. Supporting the acquisition and maintenance of expertise*. Springer.

Harteis, C. (2002). *Kompetenzfördernde Arbeitsbedingungen*. Gabler.

Häußler, P., Frey, K., Hoffmann, L., Rost, J., & Spada, H. (1980). *Physikalische Bildung. Eine curriculare Delphi-Studie*. IPN.

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1–8.

Iqbal, S., & Pipon-Young, L. (2009). Methods – The Delphi method – A guide. *Psychologist, 22*, 598–600.

Isaac, S., & Michael, W. (1995). *Handbook in research and evaluation*. Educational and Industrial Testing Services.

Jaramillo-Baquerizo, C., Valcke, M., & Vanderlinde, R. (2019). Professional development initiatives for university teachers: Variables that influence the transfer of learning to the workplace. *Innovations in Education and Teaching International, 56*(3), 352–362.

Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian and New Zealand Journal of Psychiatry, 49*, 887–897.

Kennedy, H. P. (2004). Enhancing Delphi research: Methods and results. *The Journal of Nursing, 45*, 504–511.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change, 73*, 467–482.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Addison-Wesley.

Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. V. Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 159–176). SAGE.

Ödalen, J., Brommesson, D., Erlingsson, G. Ó., Schaffer, J. K., & Fogelgren, M. (2019). Teaching university teachers to become better teachers: The effects of pedagogical training courses at six Swedish universities. *Higher Education Research and Development, 38*(2), 339–353.

Paetz, N.-V., Ceylan, F., Fiehn, J., Schworm, S., & Harteis, C. (2011). *Kompetenz in der Hochschuldidaktik. Ergebnisse einer Delphi-Studie vor dem Hintergrund der Bologna-Reformen* [Competence in university didactics. Results of a Delphi-study with background bologna-reforms]. VS Verlag für Sozialwissenschaften.

Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing, 41*, 376–382.

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. Longman.

Christian Harteis (PhD) is Full Professor of Educational Science and Educational Management at Paderborn University, Germany. His research focuses on learning through and for work, professional development and research on expertise. He conducted various empirical studies in industrial and service enterprises, and is currently a member of an interdisciplinary Graduate School investigating human-centered approaches of digitalization of working life. Before taking the position at Paderborn university in 2010, he worked at the Universities of Regensburg, Ulm, and Bamberg. He acts as an Editor-in-Chief for the journal *Vocations and Learning*. Studies in Vocational and Professional Education, and is a co-editor of this book series.

Chapter 17
The Narrative Approach to Research Professional Identity: Relational, Temporal, and Dialogical Perspectives

Katja Vähäsantanen and Maarit Arvaja

Abstract In this chapter, we present the narrative approach as applied in the field of professional learning. The specific aim is to present the methodological opportunities and concerns it raises in research on professional identity within the sociocultural frames of work environments. We utilise examples from empirical studies that have employed a range of narrative methods to collect and analyse datasets. The datasets include individuals' written and spoken narratives, encompassing the told experiences of their identities, and the meanings underlying these. We illustrate how a particular strength of narrative research lies in its ability to portray temporal pathways through the phenomena under investigation, and the relational and dialogical aspects pertaining to them. We also discuss the main challenges in the approach, and how it may be possible to conduct credible and significant research despite the challenges. We conclude with a discussion on future avenues of narrative research in the field of professional learning.

Keywords Narrative approach · Professional identity · Sociocultural framework · Thematic narrative analysis · Narrative positioning analysis

K. Vähäsantanen (✉)
Finnish Institute for Educational Research, Department of Education, University of Jyväskylä, Jyväskylä, Finland
e-mail: katja.vahasantanen@jyu.fi

M. Arvaja
Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland
e-mail: maarit.arvaja@jyu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_17

17.1 Introduction

Research on professional learning has evolved in different contexts[1] and in many directions (e.g. Billett et al., 2014). One emergent trend is the investigation of professional learning through the lens of professional identity, as noted by Tynjälä (2013). Professional identity can be understood as individuals' experience-based understanding of their professional interests, goals, ambitions, positions, values, identifications, and future prospects (Arvaja, 2016; Beijaard et al., 2004; Vähäsantanen et al., 2020). Professional identity is largely negotiated in relation to the sociocultural frameworks of work environments; however, it is also based on the agentic activities of individuals, which produce both transformations and continuities in professional identities over time (Beijaard et al., 2004; Kira & Balkin, 2014). In other words, as understood within a sociocultural framework (Vähäsantanen, 2022), professional identity is always negotiated and constructed at the temporal and emotional interface of the personal and the social dimension. In efforts to elaborate the relational nature of professional identity, one should consider especially how professional identity can be balanced or tensioned in relation to work contexts, and the kinds of emotions that emerge from professional identity – work relationships (Kira & Balkin, 2014).

Professional identity is undoubtedly at the core of changing working life, but we would argue that insufficient use has been made of *narrative approaches* in explorations of professional learning, including professional identity (for exceptions, see e.g. Goodson et al., 2010; Vähäsantanen, 2022). In this chapter we elaborate the narrative approach as applied to professional identity. We emphasise that the narrative approach refers to a frame of reference, and that within it, different methods can be used to collect and analyse data (Biesta et al., 2008; Caine et al., 2013). The general idea is to investigate individuals' experiences with a view to understanding their (professional) lives, including their identities (Spector-Mersel, 2010). Overall, the narrative approach emphasises that the narratives[2] of individuals give insights into how individuals make sense of themselves, and of the experiences, feelings, and events in their lives (Goodson et al., 2010; Riessman, 2008).

The specific aim here is to address the methodological opportunities and concerns associated with the narrative approach in research on professional identity. In the present chapter, a sociocultural framework is applied as a theoretical umbrella for

[1] Although professional learning can take place in educational trajectories before entering work contexts, this chapter focuses on persons who are actually employed as professionals in the work context.

[2] Here, we consider narrative as a sequence of events and experiences which is significant for the narrator, and which reveals identity (Bamberg, 2012; Biesta et al., 2008; Moen, 2006). The word 'narrative' is often used in different ways in the literature and also in our chapter: (i) what the subject tells (i.e. the *told* narrative), and (ii) what researchers reconstruct based on told narratives (i.e. the *constructed* narrative). Here, we understand that an interview and diary can include one or several told identity narratives. During the analysis, these told identity narratives can be constructed as a single thematically coherent and/or temporal narrative.

understanding professional identity in narrative research, which is viewed from temporal, relational, and dialogical perspectives. This chapter will first introduce the main theoretical and methodological backbones of narrative identity research. It will then illustrate how one can collect empirical data on professional identity in a variety of ways (including both written and spoken narratives), and how different analytical methods can emphasise different principles in investigating professional identity. Selecting from different methods applicable to the narrative framework, we (i) focus on interviews and diaries as narrative data-collecting methods, and (ii) consider two kinds of narrative thematic analyses and a form of narrative positioning analysis. For this purpose, we shall utilise methodological and empirical examples drawn from published professional identity studies (Arvaja, 2016, 2018; Ursin et al., 2020), and also apply a novel analytical process. Finally, this chapter will consider the strengths and future avenues of narrative research as applied to professional learning.

17.2 Theoretical and Methodological Considerations in Narrative Identity Research

In narrative identity research, the focus is on the actual construction of narratives, and the role they play in the social construction of identity (e.g. Wortham, 2001). Hence, narratives are not merely a way of telling someone (or oneself) about one's life; rather – in the words of Rosenwald and Ochberg (1992, p. 1) – they are 'the means by which identities may be fashioned'. In this sense, identity is understood as itself taking the form of a narrative. Viewed in this light, a narrative approach to data collection and analysis is a natural and viable means of understanding professionals' identity.

In the kind of narrative data collection that aims to study professional identity – such as via interviews or a diary – professionals are usually prompted to narrate experiences, and to reflect on their doing, thinking, and feeling as professionals. This telling of the narrative often involves a *temporal* perspective (e.g. Bamberg, 2012; Riessman, 2008), including space-time transitions that connect the here-and-now with previous and anticipated events and experiences (Biesta et al., 2008). Past experiences and understandings are used as a reflective mirror for evaluating the present, or new ones are incorporated for shaping the future self (Lee & Schallert, 2016). Thus, through told narratives, professionals do not just represent a current professional identity; they also construct and negotiate the identity through narrating the self within various time scales (Bamberg, 2012; McAlpine, 2016; Sfard & Prusak, 2005). In this sense, the narrative approach makes it possible to see the construction of professional identity, including *transformation* in the professional identity, via the subject's narrated past, present, and future (Wortham, 2001).

Nevertheless, the mere representational, constructive, and transformative value of a told narrative is one-sided, insofar as narrative includes also a *relational* aspect

(Akkerman & Meijer, 2011; Wortham, 2001). In representing and constructing identity, the narrator simultaneously positions the self in relation to the social and cultural context, which comprises notably other people and institutions, and what these stand for (Sfard & Prusak, 2005; Wortham, 2001). All in all, narrative identity construction can be seen as a negotiation between a person and a sociocultural context; in this sense, it is not purely something that takes place in the mind of an individual (Goodson et al., 2010; Vähäsantanen, 2022).

In this chapter, in line with narrative research, we understand professional identities as constructed by means of narratives resulting from intra/interpersonal dialogues about meaningful experiences (Assen et al., 2018). Such told narratives can be (as in our material) created via interviews and diaries, which may provide a window to understand professional identities (Biesta et al., 2008; Goodson et al., 2010). In this sense, told narratives can be seen as a surface of contact with a lived and experienced life, and with the individual's identity in relation to the sociocultural context.

By means of narrative analysis, researchers interpret and draw conclusions from narratives that are told, written, or visualised (Riessman, 2008). Through analysis, researchers construct an understanding of human actions and experiences, and form meanings regarding the phenomenon under investigation (e.g. Polkinghorne, 1995). In narrative identity research, the analysis can reveal how people at a given moment tell about, represent, and construct their identities (Assen et al., 2018). Narrative analysis can further organise experiences, events, and happenings – and also identities – into a temporal and/or context-bound whole by means of a 'plot' (Biesta et al., 2008; Moen, 2006; Polkinghorne, 1995).

In practice, there are different versions of narrative methods (e.g. Lieblich et al., 1998). In this chapter, we focus on sociocultural and dialogical approaches to narrative identity construction and its analysis. Thus, the analysis of told narratives goes beyond the mere informational value of 'telling about oneself'; hence, it notably stresses the *constructive, relational, and transformative* nature of narrative (Caine et al., 2013; Hermans, 2001; McAlpine, 2016).

For present purposes, in the case of interview research, Sect. 17.3 shows how professional identity can be analysed *thematically* by employing categorical and holistic lenses. *Categorical* analysis focuses on identity themes across told narratives, as recounted by several persons, seeking thus to provide a comprehensive picture of all identifiable kinds of identities across participants at given time. *Holistic* analysis, for its part, focuses on the identity themes of an individual participant to illustrate how, in that person, different identities are connected and overlap over time. Section 17.4 introduces a *dialogical* approach to study professional identity via learning diaries. Wortham's (2001) method on narrative positioning is used to illustrate how the professional self is narratively constructed by positioning oneself in relation to others (including both people and institutions) and their voices.

17.3 A Thematic Perspective on Professional Identity

17.3.1 Telling Professional Identity Narratives in Dialogical Interviews

In the following sections, we present analytical and empirical examples from studies conducted with professionals working in a university. The participants in total covered 20 academics (males and females) working, for example, as university lecturers and researchers at different stages in their academic careers. The data from these people were all obtained via the same thematic *narrative interview* protocol (for other narrative data collection methods, see e.g. McAlpine, 2016). The narrative interview method was used in order to elicit individuals' told narratives about their professional identity, viewable from different aspects and temporal perspectives (e.g. Riessman, 2008). Different aspects of professional identity (understood mainly as professionals' interests, ambitions, and values) emerged as themes in the interviews. In particular, the interviews included the following themes, which could be positioned on a three-dimensional temporal continuum: (i) work history, (ii) current work, identity, and agency, and (iii) future professional prospects.

Following previous narrative procedures (e.g. Riessman, 2008), no rigorous guidelines or strictly formulated questions were applied in the interviews, the aim being to hear as authentic a narration as possible. In relation to each theme, we asked one general (open-ended) question; subsequently we asked more specific questions on themes relating to the interviewees' narration. For example, in the case of the second theme, the open-ended question was formulated as 'Would you please describe your experiences of your current work?' while the more specific questions included: 'Could you tell us about your main interests and values at work?' and 'What possibilities are there to enact your main interests in your current work?' This illustrates how the interviews posed questions on the relational dimension of professional identity, covering how professional identity can be constructed, expressed, and realised in relation to the work environment (Kira & Balkin, 2014). In other words, the more specific questions above sought to gain a narration covering the balanced and tensioned relationship between professional identity and the context.

During the interviews, the interviewees were encouraged to freely describe their perceptions, experiences, and thoughts concerning the themes in question. The overall rationale for this was to create a supportive narrative practice for telling about one's experiences and identities. At the same time, the interviews constituted dialogical situations, such that both the interviewer and interviewee were able to exchange ideas and experiences. All in all, the told narratives should be understood as *relationally* produced through an interpersonal *dialogue* between the researcher and the participant, situated within a specific *context* (McAlpine, 2016; Riessman, 2008; Vähäsantanen & Saarinen, 2013). Narrative interview methods have certain advantages compared to, for example, diaries (Case 3, Sect. 17.4.1), or the use of an empathy-based story method (e.g. Wallin et al., 2020) that both produce written texts. Narrative interviews support free telling, and allow the interview to ask clarifying questions, with possibilities for gaining profound data and/or opportunities to hear individuals' own experiences. On the other hand, interview studies are time-consuming, and might tend to produce socially accepted answers in a dialogue.

17.3.2 Thematic Methods to Narrative Analysis

The following two Sects. 17.3.3 and 17.3.4 demonstrate how the interview data described above can be analysed via two *thematic narrative analysis methods*. The methods differ in important respects, but both are thematic. Thus, in the both cases, the main theme (i.e. professional identity) encompassed how people perceive themselves as professionals, and also the relational dimensions of professional identity. The main purpose of analysis processes was to identify what individuals explicitly tell about their (professional) identity (see also Lieblich et al., 1998). To some extent, *thematic narrative analysis* resembles *traditional thematic analysis*, insofar as both methods are interpretative methods and aim to identify themes from datasets (McAllum et al., 2019). However, the methods in question also have significant differences. In segmenting the data into themes across interviews, traditional thematic analysis does not indicate how events are linked together, and does not seek to give a sense of an overall 'whole' (McAllum et al., 2019). By contrast, narrative researchers do not divide up the data; rather, they work with contingent sequences within the data, that is, with the temporal and/or sequential linking of events/themes, actions, and actors (McAllum et al., 2019; Riessman, 2008).

The thematic narrative analytical methods described in this chapter were used for different purposes. An important difference lies in the fact that the aim of the *categorical-thematic* analysis (used in Case 1, Sect. 17.3.3) was to identify all the various kinds of thematic descriptions of identity relevant to constructing cultural identity narratives at a given time, using data from several participants. In particular, the notion of a *thematic plot* in the constructed narratives refers to how identity descriptions are sequenced with related themes (i.e. agency, emotions, and context). By contrast, the *holistic-thematic* analysis (used in Case 2, Sect. 17.3.4) aimed to identify and construct holistically an *individual* narrative on professional identity; it was based on a single interview, and created, notably, a *temporal plot* encompassing the evolution of particular professional identity themes (without neglecting emotions and context). In this way, the analysis of Case 2 sought to illustrate the *temporal path* of identity, including transitions in the occurrence of identity themes. It analysed professional identity themes within a case (i.e. one participant), in contrast to Case 1, which analysed professional identity themes across the cases (i.e. participants). Both cases included the aim of presenting the findings as narratives in manner that is fairly typical in (thematic) narrative research (McAlpine, 2016), but not in traditional thematic analysis. A narrative's meaning is derived from its plot, which sequentially and/or temporally orders themes and events, thus constructing a whole in which the parts are contingent upon one another (McAllum et al., 2019).

17.3.3 A Categorical-Thematic Analysis Aimed at Cultural Identity Narratives (Case 1)

In this section we demonstrate how a categorical-thematic analysis can be applied to identify different cultural identity narratives, resulting in general statements across a number of academics (for somewhat similar examples see Ylijoki & Ursin, 2013). Our examples and reflections here are based on published research originally conducted to investigate professional identity, agency, and emotions among academics working in a university (Ursin et al., 2020).

In the categorical-thematic analysis, the main focus of the analysis was not on an individual academic; rather it was on a fairly coherent communicative act that appeared to be representative across academics (Lieblich et al., 1998). The total analytical process included the following stages. Firstly, we identified different told narratives of identity within and across eight individual interviews, and identified and categorised three different descriptions of identity, based on the content of the empirical accounts. The labels applied were *teacher, researcher*, and *combined researcher-teacher*. Secondly, we identified and categorised (from the empirical accounts) the main *individual* and *sociocultural factors* that constrained or resourced each identity; these included, for example, the work community and the academic position. That is, this analytical stage also aimed to recognise the relational dimension of identity; that is, how professional identity is negotiated in relation to social context, but also to individual backgrounds. Thirdly, we identified how the emotions and/or agency were presented in the case of each different identity description. For example, we identified what kinds emotion emerged when participants told about their identity. Fourthly, we categorised *agency* on the basis of its strength (i.e. as weak, moderate, or strong), and the *emotions* according to their nature (i.e. as negative or positive).

Fifthly, we employed a culturally-based narrative analysis. Such an analysis is based on the notion that there are culturally available narratives which individuals draw on (and modify) to make sense of their experiences, and that an individual may draw on more than one narrative in explaining the experiences (see McAlpine, 2016). The goal of this analysis was to derive coherent narratives based on the told narratives of the interviewees (which had been categorised during previous phases) in order to see the kinds of identities that were constructed in the interviews, and also, the kinds of emotions and agency related to them. Based on the differences in these themes, five identity narratives were constructed and presented in a thematic narrative form (see Table 17.1). The narratives thus did not include a temporal 'identity plot' over time, encompassing identity transformations. They could nevertheless be said to embody plots, to the extent that the researchers organised the narratives in a sequentially thematic manner (Polkinghorne, 1995). In this sense, a plot tied together the main theme (i.e. professional identity) and various related themes (e.g. the relationship between identity and sociocultural context, and the emotions emerging from the relationships) in order to demonstrate the sequence of different themes.

Finally, we categorised notably the identity narratives in terms of *balance* (which means in particular an alignment between the identity and the context, in terms of

Table 17.1 Cultural narratives of professional identity

	Identity narratives	Relational dimension of identity	Agency	Emotions
Balanced	*Confident teacher*	Firm and socially constructed together with colleagues and students.	Strong agency promoted by colleagues, superiors, and one's own interests.	Confidence, meaningfulness, satisfaction, enjoyment, happiness.
	Passionate researcher	Channelled by intrinsic motivation, with position permitting full commitment to research.	Strong agency supported by one's own interests and colleagues.	Enthusiasm, enjoyment, satisfaction, meaningfulness, excitement.
Tensioned	*Insecure teacher*	Lack of competencies to be a knowledgeable teacher; a restrictive working environment.	Weak agency, framed by unsupportive colleagues and lack of one's own resources.	Insecurity, disappointment, exhaustion, non-appreciation.
	Inadequate researcher	Lack of competencies to be a proper researcher with experiences of institutional barriers.	Weak agency, framed by lack of one's own resources.	Inadequacy, exhaustion, insecurity, guilt, anxiety, sense of unfairness, fear.
Combined	*Ambivalent academic*	Balanced and at times tensed relationship between teacher and researcher identities.	Moderate agency, framed by superiors and individual resources.	Satisfaction, appreciation, meaningfulness, enjoyment, confusion.

Modified from Ursin et al. (2020)

producing positive emotions), and *tension*, which refers in particular to an emotionally unpleasant mismatch between the identity and the context (cf. Gergen & Gergen, 1986). Lying somewhere between these extremes there was also a narrative of *ambivalence*, which combined elements from both views concerning teacher and researcher identities, and which attached to them variable emotions. In other words, the categories of balance, tension, and ambivalence encompass different thematic plots pertaining to identity narratives. Table 17.1 provides an overview of the five narratives of professional identity among academics. The narratives in the written form are presented in our previous study (Ursin et al., 2020).

17.3.4 A Holistic-Thematic Analysis: To Encompass Temporality in Identity (Case 2)

Here we move towards a holistic, case-centred strategy that encompassed a temporal dimension in professional identity (e.g. Spector-Mersel, 2010). In theoretical terms, temporality in professional identity means taking into account answers to the following questions: 'Who was I as a professional in the past?', 'Who am I as a professional at the moment?', 'Who do I want to become as a professional in the future?' (Beijaard et al., 2004; Vähäsantanen, 2022). Working within a socioculturally oriented narrative framework, Sfard and Prusak (2005) emphasise also the temporal dimensions of identity by distinguishing two subsets of significant narratives about the person, namely (i) *actual identity* – which involves factual assertions consisting of narratives about the actual state of affairs, and (ii) *designated identity* – which consists of narratives presenting a state of affairs which is expected, wished for, or obligated in terms of becoming part of the person's identity. We shall take up each of these notions below.

The analysis (as conducted for the purposes of this chapter) aimed to reveal the temporal complexity and uniqueness of a professional identity trajectory via an interview with one academic (pseudonym Leana). To this end, a *holistic-thematic* approach was utilised (Lieblich et al., 1998; for a somewhat similar example see Vähäsantanen & Eteläpelto, 2011). Thus, the main focus of the analysis was to reveal how the main theme (i.e. professional identity) was told and constructed over time, that is, how it evolved from beginning to end during the interview. The analysis also aimed to encompass the relational and emotional aspects of professional identity.

In practice, the analytical process included the following phases: (i) reading the interview, (ii) searching for told narratives of professional identity and naming such identity expressions according to their nature, paying particular attention to indicators of teacher identity, researcher identity, and combined identity, (iii) focusing on the relational and emotional dimensions of professional identity in the selected told narratives – which in practice meant that we identified how the sociocultural context supported or constrained a specific identity, and the kinds of emotions that emerged from such balanced or tensioned relationships, and (iv) categorising the told narratives according their temporal frame (which could be the past, the present, or the future).

Finally, to illustrate Leana's professional identity trajectory, we developed a narrative with a temporal plot, here synthesising the contents into a coherent developmental account (Polkinghorne, 1995), with temporal descriptions of different kinds of identities, proceeding from the past to the future. The temporal phases of Leana's identity trajectory were named as (i) remembering her previous professional identity, (ii) constructing her current professional identity, and (iii) designating her future professional identity.

In the text below (see also Table 17.2), the constructed narrative on Leana's professional identity trajectory is presented in such a way as to include these temporal phases. Each phase demonstrates a specifically temporal dimension in the conveying of Leana's professional identity – including its relational and emotional dimensions – within the interview.

Remembering a Previous Professional Identity Related to her past professional identity (Beijaard et al., 2004), Leana remembered clearly 'who she was in the past' (Example 1 below), but she also described her identity in terms of 'who she did not

Table 17.2 A temporally organised individual narrative of professional identity

	Description of identity		Relational dimension of identity	Emotional dimension of identity
Remembering a previous identity	*Teacher identity*	The professional passion had been teaching.	The job description allowed teacher identity to flourish, together with intrinsic motivation.	Enacting the teacher identity created pleasant emotions, such as enjoyment.
	Researcher identity	The researcher identity was not the desired identity.	The job description emphasised teaching.	Reluctance towards the researcher identity.
Constructing a current identity	*Combined identity*	The teacher and researcher roles were united in practice, but the combined identity was not as desired.	The job description pushed for combining teaching and researching.	The combined identity felt 'insane' and stressful.
	Researcher identity	Professional passions were associated with being a researcher.	The job description did not support fully this identity.	Along with awareness of the research interest, the hope of situational change was present.
Designating a future identity	*Researcher identity*	The future ideal was to focus on research.	The future job description enabled, and intrinsic motivation boosted this identity.	The researcher identity was mainly associated with excitement.
	Back-up identity	This future identity offered an escape from possible unwanted scenarios that did not support the research identity.	Previous education made this option possible.	The back-up identity emerged from fears concerning the future, and a possible escape from an unwanted scenario.

want to be in the past' (Example 2 below). With reference to her career, Leana indicated that teaching had been her main task at the university during the years she had worked there, and that she had taught countless different courses. She emphasised her past identity from the perspective of being a teacher, and she characterised herself as being a teacher who really enjoyed teaching, and the contact with the students:

> Example 1: Well, I thought very strongly that I was a teacher, and I enjoyed teaching and being in contact with the students.

Leana's job description was also in line with her intrinsic motivation towards teaching, and it allowed her teacher identity to flourish. In addition, Leana mentioned her past researcher identity in terms of something that she could never have imagined earlier in her career:

> Example 2: I even had a feeling that I would never be a researcher because I could not sit on my own in front of a computer.

Constructing a Current Professional Identity When Leana talked about her current professional identity, her narration revealed a negotiation between the kind of actual identity she had in practice and the kind of identity she wanted to engage in (Sfard & Prusak, 2005). At the present moment, Leana had, for the first time, begun her own research as part of her doctoral studies. In her narrative, Leana emphasised that being a teacher and a researcher were strongly combined in her current professional identity, and also in her work in practice. Nevertheless, this kind of combined identity was tensioned, and the combined identity aroused feelings of confusion and stress due to challenges in combining these two sides of her identity.

> Example 3: Now it's a kind of crazy combination of research and teaching, so that I do have a bit of like a schizo feeling about this – when am I a researcher, and when am I a teacher? So perhaps I'm now between these two identities as well, so am I a researcher or am I a teacher, or am I now both? So, I am like, having problems in terms of time and at this identity level.[3]

Leana also gave indications of another kind of current identity, that of being purely a researcher. This was now the identity that she wanted to have, and one that she felt excited about; however, her work duties, including a heavy load of teaching, were not really in line with it. Leana felt as if research would be a way to be excited, and to learn and develop (Example 4); however, she lacked the agency to change a situation in which she felt that teaching sometimes got in the way of this desired identity:

> Example 4: Now I do find it terribly wonderful that I've got a chance and time to do research, and I'm really excited... it is now proper research [my main interest], as I would want it to be, and I actually find some teaching periods disturbing. Now I just feel these kinds of things, that what I want now, like really want, is to do that research, and to take it further, and to learn with it.

Designating a Future Professional Identity In talking about her future identity, Leana described both a designated identity and a back-up identity – i.e. one that would come into play if her designated identity could not flourish. Regarding the future, she now had months in which she could mostly do research in her new academic position. Thereafter, the future was still open and uncertain. Thinking about the future evoked dreams and hopes as well as doubts and fears. These feelings, together with relational and individual factors, affected her possible identities. Leana's designated identity seemed to focus notably on researching. Such an outlook seemed to be supported from a relational perspective, and it awakened many

[3] Examples 3 and 4 in this section were originally published in Ursin et al. (2020).

positive feelings, including excitement, satisfaction, and the desire to grow both professionally and personally:

> Example 5: I am so excited that now, in the autumn, I'll get to – that I'll have time for research and I am looking forward to it so much, it is such an exciting phase at this point of life.

On the other hand, the uncertainty of the researcher – work continuum – together with possible problems related to not having enough time and support on the way – had also led to the construction of a *back-up* identity. Leana's education offered possibilities to work outside the university, so she also pondered the opportunity to move away from the academic world if necessary. Nevertheless, with her current interest in research, this possible identity seemed a relatively unlikely option. It was the emotion of fear for the future (with a possible lack of options to realise her research ambitions) that seemed to guide this back-up identity, rather than any desire to change jobs.

To conclude, Leana's professional identity encompassed different forms and transformations over time. Within one form of her identity there were also various viewpoints; these included the relational dimension (that could support or constrain), and emotions (that were not always consistent), together with a range of motives and expectations.

17.4 A Dialogical Approach to Narrative Self-Construction

In this section we demonstrate how the *dialogical* approach to narrative self-construction (Wortham, 2001) – interpreted via the framework of *Dialogical Self Theory* (Hermans, 2001) – provides a methodological basis to study the relational and temporal construction of professional identity. In Dialogical Self theory, identity can be seen as consisting of multiple *I-positions* with diverse perspectives (Hermans, 2001). The I-position of a person is 'a particular voice that has been internalised in one's Self-presentation' (Akkerman et al., 2012, p. 230). Building on Bakhtin's ideas (1984), voice, in turn, can be conceptualised as 'a speaking personality bringing forward a particular perspective of the world' (Akkerman et al., 2012, p. 229).

According to Wortham (2001), the self is narratively constructed through positioning different voices from the social world in relation to each other, and by positioning oneself with respect to these voices. We demonstrate how this narrative (or interactional) positioning, and in particular two of its aspects – namely *voicing* and *evaluating* (here building on accounts of voicing and ventriloquation by Bakhtin, 1984) – provide a useful methodological tool for analysing the relational construction and transformation of professional identity.

17.4.1 Narrative Positioning (Case 3)

We use the learning diaries of one academic, Anna (pseudonym), to exemplify the analysis and process of narrative positioning. Most of the learning diary excerpts we present here appeared in a previous study, which dealt with the process of positioning in teacher identity work (see Arvaja, 2016, 2018). Anna wrote in total 18 diaries while participating in a year-long course on pedagogical studies for adult educators (PSAE). She simultaneously worked as a researcher and teacher at the university. For each diary, the teacher of the programme assigned a topic to be written about, such as: 'My own learning history: which experiences would I like to zoom in on?; What do I think about learning?; How do I react to differences in pedagogical situations?'. The students were instructed to reflect freely on the topics assigned (see also Arvaja, 2016, 2018).

In narrative positioning, the narrator (Anna) establishes both her own and other characters' voices (see also Wortham, 2001). The narrated self in a learning diary is *voiced* (i.e. characterised) through the telling of events, characterisations, and experiences related to Anna herself. These can be identified mostly by the use of first-person singular pronouns (*I, me, my, mine*). The characters articulated (e.g. named) in the narrative can be seen as external *others* whose voices belong to other individuals, groups, or institutions. The voice(s) of a character represent a recognisable social type or a recognisable type of person with their related characteristics, values, or ideologies (Wortham, 2001; Wortham & Gadsden, 2006). Through *evaluation*, the narrator establishes varying degrees of distance from the voices in the narrative via differentiation from or identification with these voices – for example by taking a critical or supportive stance with regard to them (Wortham, 2001; Wortham & Gadsden, 2006). The narrator uses different indexical cues, such as evaluative verbs, adjectives, and nouns to voice and evaluate the self and others (see also Wortham, 2001, pp. 70–75).

The prime focus in the narrative positioning analysis is on the interconnection of the voices and evaluations that seem to be relevant to Anna's professional I-positioning. In tracing interconnections, a particular aim is to identify recurrences and repetitions in characterising (voicing) and in evaluating, since recurring voices and positionings indicate that the narrator wants to *reinforce* a particular sense of self (Wortham, 2001). In the present case, a closer focus on Anna's characterisations and evaluations reveals how she positions herself in relation to (and through) her narrated self and different relevant others. It is through the processes of voicing and evaluating different characters and her narrated self that Anna (the narrator) constructs her professional identity.

In demonstrating Anna's narrative positioning, we focus especially on narrative data in which *boundary experiences* (Assen et al., 2018; Henry & Mollsteadt, 2022; Hermans, 2013) are reflected. In a boundary experience, the uncertainties, challenges, or disappointments faced in one's life (e.g. in one's work) can led to disruption and reconfiguration within the dialogical self (Henry & Mollsteadt, 2022). In these situations, the person often experiences tensions between different

I-positions (or between internal and external voices) causing feelings of discomfort and helplessness, or a feeling of being unable to cope (Geijsel & Meijers, 2005; Ligorio & Tateo, 2007). Nevertheless, the effort to overcome a boundary experience can open up a moment for identity negotiation, often leading to a process of re/de-positioning (Assen et al., 2018) aimed at restoring continuity and consistency in the dialogical self (Hermans & Hermans-Konopka, 2010).

17.4.2 An Example of Narrative Positioning

In the following illustration of narrative positioning, we show how pedagogical studies can be regarded as an external other – an entity that provided new perspectives, allowing Anna to confront and identify the tensions underlying her work and professional identity. The pedagogical studies triggered boundary experiences, i.e. situations in which Anna felt tensions between her different I-positions. These led to a re-negotiation of her professional identity and I-positioning.

At the beginning of the studies, Anna characterised herself (in other words, 'voiced' herself; see Wortham, 2001) as a teacher who preferred to enact a particular teacher 'role' in the teacher-student relationship. However, as the studies proceeded Anna re-negotiated this characterisation in her internal dialogue:

> Example 1: I myself have found it good, at least earlier, that I have formed a kind of teacher role for myself, since it is handily there in-between the self and the student in case of difficult situations. (Diary 1).
>
> Somehow, I earlier thought that teachers are of a certain type, I mean they have a specific kind of role. [. . .] But it would not need to be like this, I mean it's absolutely wonderful that bringing out one's own persona seems to be supported in PSAE! In my earlier days when I was shy and timid, I was probably suffering the most from the fact that I was unable to bring out my own self from behind the protective shell in any sphere of life. Now that I have not had such a strong protective shell for years, I'm somehow annoyed with the idea that I should behave against my personality to be a real teacher. So it's good that things don't have to be like this. Adopting a role has always been hard for me anyway, I mean taking on a role that feels artificial. (Diary 10)[4]

Example 1 above illustrates Anna's renegotiation of her I-position as a teacher. This renegotiation appears in Anna's differentiation (i.e. her evaluation, see Wortham, 2001) from the notion of teachers being a certain type of person (i.e. cultural typification, Akkerman & Meijer, 2011). The pedagogical studies, with their idea of 'bringing out one's own persona', represent an external other, with whose voice Anna now identifies herself. This idea triggers a boundary experience, enabling Anna to observe and explore her I-positions from a meta-position (Hermans, 2013), and become aware of the conflicting voices between them. Anna now seems to feel that her previously adopted teacher I-position does not support her personal I-position, insofar as she cannot now fully identify herself with that teacher

[4] All the examples in this section have been published originally in Arvaja 2016 and/or 2018.

positioning. The self-dialogue here manifests itself in an internal dialogue between Anna's collective 'teacher' voice and a personal voice representing her true subjective feelings, i.e. the authentic voice (Marková, 2006).

Anna's storytelling self (Wortham, 2001) also brings her past internal voice into the space of the negotiation. Hence, there are two parallel voices that contradict Anna's present I as a person. These are (i) the external voice of 'the teacher' with an (artificial) role, and (ii) the internal voice of a past (yet still acknowledged) shy self bearing a protective shield. Anna positions her present self as distant from both of these voices, since these would hinder the actualisation of her true self. In her narrative she creates new linkages between her (current) personal I-position and the (now properly understood) teacher I-position, and redefines these so that they are more coherent and dialogically integrated (Hermans & Hermans-Konopka, 2010). This identity re-negotiation is apparent in the expressions of relief and joy in Anna's narrative. At the end of the studies, Anna characterises these harmonised I-positions:

> Example 2: As a teacher, I am easy to approach, and I'm genuinely interested in students and their learning and guiding them, and I can be empathetic to students. [. . .] I think that being one's own personal self makes it easier to interact with students, since then the interaction situations are more genuine. Moreover, acting as oneself is less exhausting in the psychological sense than enacting a given role. (Diary 17)

In the above example Anna explicitly characterises and represents herself as a (now more authentic) teacher. In this self-positioning Anna speaks with a student-centred voice, as a person who is genuinely interested in students and their learning, and who values a genuine relationship with the students. Across several instances, Anna repeats – and thus reinforces (Wortham, 2001) – an identity of being oneself as a teacher, as opposed to taking an artificial role (see also Example 1). Through Anna's narrative self-construction (Examples 1 and 2) we can see how, during the year-long pedagogical studies, Anna's dialogical self (as regards her teacher and personal I-position) has moved from a state of decentring (disharmony) to a state of centring (harmony), resulting in a more coherent sense of self (Hermans, 2013). Anna has come to an understanding of the kind of teacher she wishes to be.

The pedagogical studies seemed to give rise to another boundary experience leading to a reconfiguration within Anna's dialogical self:

> Example 3: It's frightening, too, how my enthusiasm for teaching/guidance is increasing along with PSAE, since the shift of my interest increasingly outside my current job description is weakening my motivation to work in my current job [research]. (Diary 8)
>
> For me, the danger of PSAE actually lies in that I think it's one reason why I am now experiencing such a strong work identity crisis. [. . .] I mean, PSAE is increasingly opening my eyes to the idea that teaching/education is work that I want to do. (Diary 10)

Example 3 above demonstrates how Anna feels distress in her dual-status academic position. Anna's inner tension and confusion are reflected in the emotional tone of her narrative, insofar as she uses strong, emotionally loaded words: 'frightening', 'danger', and 'identity crisis' (e.g. Ligorio & Tateo, 2007). The external other (i.e. pedagogical studies) promotes new awareness ('opening my eyes'), and offers Anna a new frame of reference to assess and critically reflect on her current professional positions from a meta-position. As a response to (the perspective voiced by) this external other, Anna re-defines her professional I-positioning. Anna's I-position as a researcher is weakening, while her I-position as a teacher is strengthening in the course of the pedagogical studies.

Anna takes a critical position towards the university from her teacher I-position – a position which is based on a student-centred perspective (see Example 2). Anna feels that the reformed university curriculum does not enable her to actualise this perspective:

> Example 4: The latest curriculum reform and time limits set for students to finish their studies have, in my opinion, driven university education in a direction that resembles mass production. Earlier, students could write a master's thesis for two different subjects, but now this is prevented. Also, the studies must proceed more straightforwardly, which may prevent choosing more unusual subjects as one's minor. We can say, therefore, that academic freedom has also been reduced for students. (Diary 15)
>
> Discovering one's potential, talents, etc. requires time and time for oneself. As efficiency thinking has spread into university as well, it may affect students' possibilities to recognise and apply their strengths. (Diary 16)

In Example 4, Anna's narrative is double-voiced, in the sense that different voices from the social world are positioned in a dialogue at the level of ideologies associated with these voices (Wortham, 2001). Anna voices the university as a managerial institution. This becomes apparent through indexical cues, including the mentions of 'mass production', 'freedom. . .reduced', 'efficiency thinking', and 'time limits'. As a counterpart to this, Anna sees that traditional academic values and humanistic ideals, including students' opportunities for time, unique choices, and discovering one's own potential, are deteriorating in the university (hence positioning students as weak in terms of agency). It is important to recognise that in this situation Anna does not *explicitly* represent her voice and I-positioning; it is rather that by speaking through (voicing and evaluating) various 'others' she interactionally positions herself with regard to the characteristics, values, and ideologies these others stand for. Anna identifies with the students' position, while simultaneously reinforcing her own position as critical yet powerless against the university and its managerial power.

Anna's redefined professional I-position (identity) leans strongly towards teaching (see Example 3). However, she feels that 'the university world' does not support this I-positioning:

> Example 5: Even though I would truly like to help students in their choices and study-related problems, I feel that the university world does not give much of a chance for this. Because lecturers' posts are filled by research merits (and I don't believe this will change), if you would consider a career, then everything 'extra' like giving more time to students should be cut off. This is one reason why I want to leave university, as so many things there seem inhuman. [. . .] In the university world, self-interest can surpass humanity, and if you want to help students more than you 'must', then at worst it can be 'suicide career-wise' at the university. [. . .] Because I have actually already decided to try and seek a job elsewhere, I can now also invest more in both teaching and students. (Diary 16)

This narrative further confirms – and makes explicit – an emotionally and morally charged conflict within Anna's professional self (see also Example 4). While Anna, as a teacher, would like to help students and give them time, as an academic she ought to publish as much as possible (thus gaining the associated research merits), which limits the time given to students. These aims and motives are in clear contradiction, causing distress in Anna's professional positioning. Anna voices herself as a caring and empathetic teacher who works for the students, whereas the university is voiced as an indifferent and inhuman community that does not give due credit to a devoted teacher. In her narrative, Anna repeatedly constructs a huge distance between the institutional voice and her personal and professional voice. To maintain her professional identity and to truly enact her re-defined teacher I-position, Anna feels she is internally constrained to leave the university (or to commit 'suicide career-wise'). This shows a progressive movement (Hermans & Hermans-Konopka, 2010) in Anna's narrative self-construction, insofar as she constructs a more agentic position through a critical stance, and a decision to act upon her own values and sense of self identifying possibilities for the future. Table 17.3 presents a summary of Anna's narrative positioning within her dialogical self.

To conclude, the construct of narrative positioning provides conceptual and analytical tools for exploring the irreversible relationship of 'I and the other' in the context of professional identity. Professional identity can be seen as *positioned into being* through positioning one's self with respect to the relevant characters and their respective voices, as presented and evaluated in the narrative.

Table 17.3 Anna's narrative positioning in the dialogical self

	Negotiation of teacher and personal I-positions	Voicing and evaluating	Negotiation of researcher and teacher I-positions	Voicing and evaluating
Start of the pedagogical studies	Separation between teacher and personal I-positions	Identifying with the collective 'teacher' voice (different from the authentic, personal voice)	Dual status: academic position, separate teacher and researcher I-positions	Identifying with the assumed collective voice of 'the teacher' (researcher not voiced at this point)
	Disharmony between personal I-positions and presumed teacher I-position	Differentiating the past ('shy') personal voice and the collective 'teacher' voice from (present) authentic, personal voice	Strengthening the teacher I-position, weakening the researcher I-position	Identifying with the voice of the pedagogical studies (being oneself as a teacher), distancing oneself from the managerial and research-centred voice of the university
End of the pedagogical studies	Harmony and integration between the personal and the teacher I-position	Identifying the teacher voice with the personal voice (being oneself as a teacher)	Teacher I-position as the core of professional identity	Identifying with the student-centred teacher voice, based on humanistic ideals and traditional academic values

17.5 Methodological Discussion on Narrative Professional Identity Research

Our chapter indicates that both of the research tools applied (interviews and diaries) provide an excellent starting point for exploring the complexity, uniqueness, and different sides of professional identity (itself a socially constructed and experiential phenomenon) through written and spoken narratives. In the diaries and interviews, professional identity is reflected and constructed in relation to the sociocultural context of work. In addition to this relational perspective, the datasets offer a timeline to observe possible transformations in identities: (i) through time layers reflected by the narrator (reflections on the past, present, and future) within one interview/diary, or (ii) across several datasets through long-term data collection. Both methods also serve as a reflective practice by which past experiences are used to reflect and evaluate new ones (e.g. Lee & Schallert, 2016) – a practice which can encompass telling about and reflecting both the current identity and the future-oriented identity (Sfard & Prusak, 2005). Thus, the strengths of data collection as presented here include an opportunity to capture individuals' experiences, and to gain information on professional identity, within its context, from different temporal perspectives. For example, the method of empathy-based stories captures comprehensively individuals' future scenarios and social representations regarding a particular phenomenon, but it is less able to capture individuals' lived experiences (Wallin et al., 2020).

Interviews and diaries have specific strengths and limitations as data collection methods. As compared to interviews (here used in Cases 1 and 2), the learning diary (as used in Case 3) makes it possible to monitor the ongoing process of identity construction in a less retrospective and exhausting manner. However, as compared to interview research, in a diary study the researcher has no opportunity to ask for clarifications via a dialogue that might provide a deeper perspective on professional identity.

The narrative approaches used in the analysis included: *categorical-thematic analysis* (Case 1), *holistic-thematic analysis* (Case 2), and *narrative positioning analysis* (Case 3). All the cases demonstrate how the expressions and construction of professional identity are relational in nature. In/through the narratives, professional identity was negotiated in relation to 'others' (people and institutions with related values, practices, and ideologies), thus reflecting the sociocultural context of the work. In addition, in Case 3 the narrative positioning analysis provided tools to analyse the professional's internal dialogue and to depict diverse voices behind the self- and other positionings, with possibilities to uncover (mis)alignments between the work/educational context and different sub-identities (I-positionings). In Cases 2 and 3, the tensions and misalignments detected through the analysis made it possible to trace the renegotiation or transformation in the identity of the professionals in question. However, whereas in Cases 1 and 2 expressions of identity within the narratives were mostly traced through *explicit representations* of a person's own thinking, doing, and valuing as a professional, in Case 3 the identity expressions could also be traced more implicitly, via the narrator's voicing and evaluating (speaking through) the other characters in the narrative.

Cases 2 and 3 demonstrated in particular how it is possible to trace *temporality* in narratives, through the connections formed by the narrators, linking their past, present, and future. These cases also made it possible to recognise identity *trajectories* – that is, renegotiation and transformation in professional identity – via one interview (Case 2) and via long-term diary research with one participant (Case 3). This illustrates how a *holistic* approach is applicable if we wish to understand the phenomenon under examination over time (see also Biesta et al., 2008; Riessman, 2008), including the changes and trajectories in professional identity observable in our cases. However, a weakness in Cases 2 and 3 was that they utilised data from individual participants. In order to increase the trustworthiness and transferability, the analytical procedures could be continued via an analysis similar to the one presented here, but conducted over several participants. The outcome could be, for example, different individually-based professional identity trajectories (each based on one participant), or different general professional identity trajectories (each based on several participants sharing more or less similar trajectories).

In contrast to the holistic approach, Case 1 used a categorical-thematic analysis to analyse professional identity across several subjects. It produced different identity narratives with a thematic plot rather than a temporal plot. This kind of analysis provides diverse perspectives on professional identity, increasing the credibility and possible transferability of the findings. On the other hand, although useful for making more general statements based on several subjects, one could say that category-centred approaches may tend to overlook the sequential features that are the hallmarks of narrative (Riessman, 2008; Spector-Mersel, 2010).

The cases indicate ways in which narrative inquiry can reveal unique perspectives and deeper understandings of professional identity, incorporating temporal and relational viewpoints. Different methods have their specific strengths. For example, traditional thematic analysis segments the data (McAllum et al., 2019), while narrative analysis methods reveal how identities are transformed and how different identities are related to each other over time. Narrative methods can also show how identities are sequenced with other themes, such as the sociocultural context. A further advantage of narrative approach is that researchers' constructed narratives are easily accessible to readers, so that the readers can use results practically and pedagogically (McAlpine, 2016) as a tool to reflect their identities. However, as with all methodologies, there are also limitations in the narrative approach. These limitations involve notably the use of subjective and interactional data, and small datasets. We should also be mindful that the told narratives of professionals capture only a limited number of experiences, and give only partial information on identities. Researchers then form interpretations based on the information provided, seeking to construct a coherent narrative (see also McAlpine, 2016).

Insofar as knowledge is taken to be subjective, socially constructed, and partially incomplete (Riessman, 2008), narrative research does not aim to convince the readers that the narratives found by the researcher correspond to reality or to some 'true' state of affairs. Rather, narrative research seeks to convince the reader that certain narratives are believable, and to present narratives that the reader can identify with (Moen, 2006). Since narrative researchers now prefer to use the term 'transferability' (involving findings that can be applicable to other similar settings) instead of 'generalisation' (e.g. Riessman, 2008), our aim here has been to address different ways in which transferability can be recognised and achieved in narrative research. When the narrative findings are compared and confirmed via the findings of other studies, the credibility and transferability of the narrative findings is also enhanced.

17.6 Conclusions and Future Avenues: Narrative Identity and Learning Research

We conclude that scholars in the field of professional learning could benefit from the use of narrative research in efforts to explore and contribute to theory on professional identity, and in particular, to professional learning. The narrative approach can reveal the complex and temporally unique pathways that are present, while shedding light also on the relevant relational and sociocultural perspectives on the phenomena under investigation (see also Biesta et al., 2008; Goodson et al., 2010).

In the field of professional learning and development, one of the main aims is to develop workplace pedagogy to enhance professional learning. Scholars now see identity work as way of supporting professionals' learning, and have introduced some methods to support professional identity work. These include in particular narrative arts-based frameworks, including creative writing and visual identity methods (Schellings et al., 2021; Vähäsantanen et al., 2020). Such methods are particularly capable of revealing the emotional, embodied, and tacit (non-linguistic and non-cognitive) aspects of professional identity work.

In future, we will need a more elaborated, research-based understanding of professional identity trajectories, together with narrative pedagogies to support identity work in working life. We believe that these endeavours could have a meaningful impact on individuals and work organisations, and we invite readers to start their own journey in the field of narrative identity and learning research.

Acknowledgements This chapter is derived in part from articles published in Journal of Further and Higher Education (6.12.2018), available online https://www.tandfonline.com/doi/full/10.1080/0309877X.2018.1541971; Teaching and Teacher Education (3.8.2016), available online https://doi.org/10.1016/j.tate.2016.07.024; and Teaching in Higher Education (21.9.2017), available online https://www.tandfonline.com/doi/full/10.1080/13562517.2017.1379483.

References

Akkerman, S. F., & Meijer, P. C. (2011). A dialogical approach to conceptualizing teacher identity. *Teaching and Teacher Education, 27*(2), 308–319.

Akkerman, S. F., Admiraal, W., & Simons, R. J. (2012). Unity and diversity in a collaborative research project. *Culture & Psychology, 18*(2), 227–252.

Arvaja, M. (2016). Building teacher identity through the process of positioning. *Teaching and Teacher Education, 59*, 392–402.

Arvaja, M. (2018). Tensions and striving for coherence in an academic's professional identity work. *Teaching in Higher Education, 23*(3), 291–306.

Assen, J. H. E., Koops, H., Meijers, F., Otting, H., & Poell, R. F. (2018). How can a dialogue support teachers' professional identity development? Harmonising multiple teacher I-positions. *Teaching and Teacher Education, 73*, 130–140.

Bakhtin, M. (1984). *Problems of Dostoevsky's poetics* (C. Emerson, Trans.). University of Minnesota Press.

Bamberg, M. (2012). Why narrative? *Narrative Inquiry, 22*(1), 202–210.

Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education, 20*(2), 107–128.

Biesta, G., Goodson, I., Tedder, M., & Adair, N. (2008). *Learning from life: The role of narrative* (Learning Lives Summative Working Paper No 2). University of Stirling: The Learning Lives Project.

Billett, S., Harteis, C., & Gruber, H. (Eds.). (2014). *International handbook on research in professional and practice-based learning*. Springer.

Caine, V., Estefan, A., & Clandinin, D. J. (2013). A return to methodological commitment: Reflections on narrative inquiry. *Scandinavian Journal of Educational Research, 57*(6), 574–586.

Geijsel, F., & Meijers, F. (2005). Identity learning: The core process of educational change. *Educational Studies, 31*(4), 419–430.

Gergen, K., & Gergen, M. (1986). Narrative form and the construction of psychology science. In T. Sarbin (Ed.), *Narrative psychology: The storied nature of human conduct* (pp. 22–44). Praeger Press.

Goodson, I. F., Biesta, G., Tedder, M., & Adair, N. (2010). *Narrative learning*. Routledge.

Henry, A., & Mollstedt, M. (2022). Centrifugal–centripetal dynamics in the dialogical self: A case study of a boundary experience in teacher education. *Journal of Constructivist Psychology, 35*(2), 795–814.

Hermans, H. J. M. (2013). The dialogical self in education: Introduction. *Journal of Constructivist Psychology, 26*(2), 81–89.

Hermans, H. J. M. (2001). The dialogical self: Toward a theory of personal and cultural positioning. *Culture & Psychology, 7*(3), 243–281.

Hermans, H. J. M., & Hermans-Konopka, A. (Eds.). (2010). *Dialogical self theory: Positioning and counter-positioning in a globalizing society*. Cambridge University Press.

Kira, M., & Balkin, D. B. (2014). Interactions between work and identities: Thriving, withering, or redefining the self? *Human Resource Management Review, 24*(2), 131–143.

Lee, S., & Schallert, D. (2016). Becoming a teacher: Coordinating past, present and future selves with perspectival understandings about teaching. *Teaching and Teacher Education, 56*, 72–83.

Lieblich, A., Tuval-Mashiach, R., & Zilber, T. (1998). *Narrative research: Reading, analysis and interpretation*. Sage.

Ligorio, M. B., & Tateo, L. (2007). "Just for passion": Dialogical and narrative construction of teachers' professional identity and educational practices. *European Journal of Social Psychology, 5*(2), 115–142.

Marková, I. (2006). On the "inner alter" in dialogue. *International Journal of Dialogical Science, 1*(1), 125–147.

McAllum, K., Fox, S., Simpson, M., & Unson, C. (2019). A comparative tale of two methods: How thematic and narrative analyses author the data story differently. *Communication Research and Practice, 5*(4), 358–375.

McAlpine, L. (2016). Why might you use narrative methodology? A story about narrative. *Eesti Haridusteaduste Ajakiri, 4*(1), 32–57.

Moen, T. (2006). Reflections on the narrative research approach. *International Journal of Qualitative Methods, 5*(4), 56–69.

Polkinghorne, D. (1995). Narrative configuration in qualitative analysis. In J. A. Hatch & R. Wisniewski (Eds.), *Life history and narrative* (5th ed., pp. 5–23). Routledge.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Sage.

Rosenwald, G., & Ochberg, R. (Eds.). (1992). *Storied lives: The cultural politics of self-understanding*. Yale University Press.

Schellings, G., Koopman, M., Beijaard, D., & Mommers, J. (2021). Constructing configurations to capture the complexity and uniqueness of beginning teachers' professional identity. *European Journal of Teacher Education.* https://doi.org/10.1080/02619768.2021.1905793

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34*(4), 14–22.

Spector-Mersel, G. (2010). Narrative research: Time for a paradigm. *Narrative Inquiry, 20*(1), 204–224.

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*(1), 11–36.

Ursin, J., Vähäsantanen, K., McAlpine, L., & Hökkä, P. (2020). Emotionally loaded identity and agency in Finnish academic work. *Journal of Further and Higher Education, 44*(3), 311–325.

Vähäsantanen, K. (2022). Professional identity in changing workplaces: Why it matters, when it becomes emotionally imbued, and how to support its agentic negotiations. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning* (pp. 29–46). Springer.

Vähäsantanen, K., & Eteläpelto, A. (2011). Vocational teachers' pathways in the course of a curriculum reform. *Journal of Curriculum Studies, 43*(3), 291–312.

Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2020). University teachers' professional identity work and emotions in the context of an arts-based identity coaching program. In L. McKay, G. Barton, S. Garvis, & V. Sappa (Eds.), *Arts-based research, resilience and well-being across the lifespan* (pp. 233–256). Palgrave Macmillan.

Vähäsantanen, K., & Saarinen, J. (2013). The power dance in the research interview: Manifesting power and powerlessness. *Qualitative Research, 13*(5), 493–510.

Wallin, A., Pylväs, L., & Nokelainen, P. (2020). Government workers' stories about professional development in a digitalized working life. *Vocations and Learning, 13*, 439–458.

Wortham, S. (2001). *Narratives in action. A strategy for research and analysis*. Teachers College Press.

Wortham, S., & Gadsden, V. (2006). Urban fathers positioning themselves through narrative: An approach to narrative self-construction. In A. De Fina, D. Schiffrin, & M. Bamberg (Eds.), *Discourse and identity* (pp. 314–341). Cambridge University Press.

Ylijoki, O.-H., & Ursin, J. (2013). The construction of academic identity in the changes of Finnish higher education. *Studies in Higher Education, 38*(8), 1135–1149.

Katja Vähäsantanen (PhD) is a senior researcher at the Finnish Institute for Educational Research, University of Jyväskylä, Finland. She is an Adjunct Professor in the field of working life research and development, University of Tampere, Finland. Her current research interests include lifelong guidance as well as professional identity, agency, learning, and emotions in work organisations. She is also interested in the potential of different methodological approaches, notably the narrative approach and its applicability to research on learning in working life. Her recent publications include articles in Vocations and Learning, Journal of Workplace Learning, Frontline Learning Research, Studies in Continuing Education, and Professional Development in Education.

Maarit Arvaja (PhD) works as a senior researcher at the Finnish Institute for Educational Research, University of Jyväskylä. She is an Adjunct Professor in the field of collaborative learning and interaction in learning environments. Her current research focuses on teacher identity and professional development, higher education pedagogy, and dialogical and narrative approaches to learning. Her recent publications include articles in European Journal of Teacher Education, Teaching in Higher Education, Teaching and Teacher Education, Journal of the Learning Sciences, and The Internet and Higher Education.

Chapter 18
Capturing Actions of Communities: Towards Virtual Ethnography and Digital Tools in Researching Organizations and Workplace Learning

Soila Lemmetty, Kaija Collin, Vlad Glăveanu, and Susanna Paloniemi

Abstract In this chapter, we introduce the ethnographic research methodology as a starting point in organisational and workplace learning research. In approaching learning at work as a practice-based and communal phenomenon, this strategy has been found suitable for studying its nature and practices. We focus on new and innovative ways to conduct ethnography—especially virtual ethnography and digital tools. In this chapter, we first briefly describe the background of ethnographic methodology. We then move on to consider why workplace learning should also be studied in virtual environments and how digital tools, such as subjective cameras, can be utilized in conducting ethnographic research. Then, we present two empirical case examples. The first case illustrates the study of informal learning using ethnography in a virtual environment. The second case illustrates the use of subjective cameras in a subjective, evidence-based ethnographic process. Using these case examples, we show how the basic principles of ethnography can be strengthened and applied in virtual environments and with the help of digital tools in workplace learning research. We also consider some potential ideas for further ethnographic research on workplace learning alongside ethical matters related to ethnographic and virtual ethnographic research.

Keywords Ethnography · Organization ethnography · Virtual ethnography · Workplace learning · Subjective evidence-based ethnography

S. Lemmetty (✉) · K. Collin · S. Paloniemi
Department of Education, University of Jyväskylä, Jyväskylä, Finland
e-mail: soila.j.lemmetty@jyu.fi; kaija.m.collin@jyu.fi; susanna.paloniemi@jyu.fi

V. Glăveanu
Department of Psychology and Counselling, Webster University Geneva, Bellevue, Switzerland
e-mail: glaveanu@webster.ch

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_18

18.1 Introduction

Ethnography is a research strategy that focuses on cultures, cultural practices, and people who act as parts of these cultures. It is linked to the social sciences and is concerned with "human studying human" (Reed, 2017), such as by focusing on the practices and interactions of people in different contexts (Fetterman, 2010; Hammersley & Atkinson, 2007). The emphasis on cultural understanding in ethnographic research is rooted in anthropology, and it goes back 2 centuries, to when anthropologists traveled to new continents to understand unfamiliar cultures, people, and their everyday lives. The subjects of anthropological and ethnographic research have changed over time, but ethnographic research has a multidisciplinary nature, and this expansion into new areas and fields over time is expected. As an empirical and qualitative approach, an ethnographic research strategy can be applied, especially when the aim of the study is to describe, structure, and theorize about complex and often hidden cultural phenomena (Paloniemi & Collin, 2010). Ethnography is also a relevant strategy when the researcher wants to understand social and psychological phenomena, processes, and their effects on people and communities (Heyl, 2001) or to describe and understand different situations and events from the participants' perspectives (Cohen et al., 2007; Pole & Morrison, 2003). The benefit of ethnography is that it enables a comprehensive and detailed description of the target phenomenon (Heyl, 2001). Thus, the ultimate aim of ethnographic research is to describe what happens in various contexts and environments, as well as how people act, see, and interpret their own actions and those of others.

Research on adult learning, particularly in the workplace context and in different communities, has increased over the last 20 years, focusing, for example, on the concepts of workplace learning, organizational learning, and human resource development. Hence, most previous studies have been conducted in organizational and management research and in educational and adult learning research. Although the methodologies used to study adult learning have been diversified, most research has focused on building theories or utilizing quantitative research methods (Li & Bartunek, 2009). According to Marjolein et al. (2006), previous studies that focused on learning processes at work were based on questionnaires and interviews. They suggested that diverse methods and research tools should be applied to explore professional learning in the work context. The methods used in the study of workplace learning and its processes have been conventional, and a balance between quantitative and qualitative research approaches has also been recommended (Li & Bartunek, 2009).

To contribute to the contextual understanding of workplace learning, previous studies have called for the use of an ethnographic research strategy, specifically observational research (Lemmetty & Collin, 2020; Collin, 2006; Baskett, 1993). As workplace actors, activities, problems, and tools become increasingly complex and diverse, the benefits of ethnography as a methodological strategy are increasingly recognized. Ethnography is especially useful (1) when we know little about the phenomenon at hand, (2) when we aim to collect detailed information and reach a comprehensive, situational-based understanding, and (3) when we want to connect the retrieved information to the context in which it emerged. Therefore, the focus of interest in the work community and organizational learning may be on interactions, relations, networks, and work practices. Consequently, the use of ethnography in the study of workplace learning has increased in recent years (e.g., Collin, 2006; Snoeren et al., 2016; Riera et al., 2020; Lemmetty & Collin, 2020; Lemmetty, 2020a, b). For example, the ethnographic strategy has been employed to examine learning at work and relevant phenomena in hospitals (e.g., Valleala et al., 2015; Hägg-Martinell et al., 2017), technology organizations (e.g., Collin et al., 2018; Lemmetty & Collin, 2020; Lemmetty, 2020a, b), and educational contexts (e.g., Pastuhov & Sivenius, 2020). Growing interest in ethnographic methodology has also occurred in the broader context of organizational and workplace research, which Rouleau et al. (2014) describe as an "ethnographic turn" (p. 2). Thus, while *organizational ethnography* is not a new idea (e.g., Selznick, 1949), it has gained acceptance as a research method and even a paradigm in recent decades (van Maanen, 1986; Rouleau et al., 2014).

The reasons for the growing use of the ethnographic research approach in organizational and management research derive from the changes taking place in working life. Work and practices in organizations have become more complex in terms of intensified competition, the time and space of the work, and the wide variety of technological tools and platforms used in the work (van Maanen, 1986; Yanow, 2009). At the same time, social media and virtual channels have exploded in popularity—especially in the "Google generation" born after 1993—and have forced researchers to exploit and examine them (Murthy, 2013). Regarding digitalization, the importance of remote work is increasing, especially in the fields of expert and knowledge work. Most recently, the global COVID-19 pandemic has increased the role of remote work. In these contexts, organizational and workplace learning ethnography can no longer be unambiguously linked to a certain physical place and time.

Remote work creates new opportunities and challenges for the learning and development of employees and organizations. As formal learning situations move online and informal processes become revised to accommodate increasingly diverse combinations of remote work and "normal" work, how they are viewed in ethnography, must also change. As a result, there has been growing interest in exploiting methods that help to grasp subtle shifts in organizational life (Rouleau et al., 2014). There have been advances in the use of social media, and the huge variety of technological tools has renewed the workplace and increased the variety of settings that can be used in ethnographic research (e.g., the internet, mobile phones, videos, and GPS). In this context, the focus should be on how to promote organizational ethnography as a paradigm (e.g., by writing about organizations in a different way) (van Maanen, 1986). It is not only a matter of researching online environments or applying new digital tools; it is also a complete re-evaluation of research practices and processes, starting by defining the concept of "the field," the role of the researcher in the new context, and ethical issues related to virtual ethnography (Rouleau et al., 2014).

In this chapter, we first briefly describe the background of ethnographic methodology. We then move on to consider why workplace learning should also be studied in virtual environments and how digital tools, such as subjective cameras, can be utilized in conducting ethnographic research. Then, we present two empirical case examples. The first case illustrates the study of informal learning using ethnography in a virtual environment. The second case illustrates the use of subjective cameras in a subjective, evidence-based ethnographic process. Using these case examples, we show how the basic principles of ethnography can be strengthened and applied in virtual environments and with the help of digital tools in workplace learning research.

18.2 Virtuality and Digital Tools in Ethnographic Research in Organizations and Communities

18.2.1 Background of Ethnographic Research on Organizations and Workplaces

The aim of ethnographic research is to describe and explain the activities of people in their environments (Croucher & Cronn-Mills, 2015). The trajectory of ethnographic research does not always appear to advance, and its findings are not always generalizable or reproducible (Davies, 1999). Ethnography is well suited for exploring community activities, social interactions, and cultures, especially in small groups (Pole & Morrison, 2003) and in their own contexts (e.g., Coffey, 1999; Hammersley & Atkinson, 2007). Thus, the ethnographic research approach has been utilized to study learning at work by focusing on the diverse sociocultural and contextual circumstances interwoven in the learning processes of individuals, collectives, or organizations (see, for example, Lemmetty, 2020b; Riera et al., 2020).

Previous studies on workplace learning have noted that ethnographic research progresses from the examination of an extensive research area (e.g., learning situations) or entity to a detailed focus on smaller, more structured areas (e.g., the specific learning acts of individuals or groups) (Collin, 2006; Lemmetty, 2020a). In this funnel-like process, data collection, analysis, and interpretation are intertwined (Davies, 1999). All stages of ethnography, from data collection and analysis to writing, overlap and cannot be separated in practice (Kunda, 2013). Additionally, in ethnography, it is not necessary to commit to a single theoretical frame of reference or to adhere to a theory-based research approach; rather, the richness of perspectives can be increased through theory triangulation (Hammersley & Atkinson, 2007). Although there are no clear guidelines for conducting ethnographic research—for its process or stages—some basic starting points, concepts, and areas related to ethnographic research should be examined.

A key concept in ethnography is *field.* In organizational ethnography, it is important to understand how a field is defined and the factors that should be considered in its examination (Tuncalp & Le, 2014). Burrell (2009) defined the field as "the stage on which the social processes under study take place" (p. 182). Thus, by focusing on interactions between people, ethnographic observation in the field offers the possibility of understanding the sociocultural nature of the phenomenon being studied (e.g., workplace learning). The ethnographic field is thus the entity in which processes, social interactions, goals, and boundaries emerge. At the same time, the field is defined as the process of empirical data formation, and its boundaries are ultimately determined by the research questions, interpretations, and writing (Fingerroos, 2003).

On the other hand, *fieldwork* refers to the length of time a researcher spends in the field and the activities that they perform while in the field. During fieldwork, the aim of the researcher is to gain an understanding of the research phenomena from the perspective of the subjects. It is also important to link and relate the researcher's interpretations and understanding to the wider culture and context. Ethnography is well suited to viewing organizations in their broad (e.g., societal) contexts (Watson, 2012).

Although the concept of *context* is widely used in organizational research, it is vaguely and weakly defined (see more about the concept of context in organizational research in the article by Ostroff, 2019). In ethnography, context often refers to the broader understanding and knowledge that underpins the research phenomenon and the relationships between micro-observations and the macro level. In conducting their research field ethnographers refer to the context of a work community or organization or a broader entity, such as a particular industry, event, state, or phenomenon. Thus, context is the broad setting in which the phenomenon under study is viewed, while field is the object of research about which observations are made and through which the context is understood.

In ethnographic fieldwork, parallel data collection methods, such as observations, field diaries and notes, interviews, discussions, and documents, are commonly utilized (O'Reilly, 2011). Typically, observation data, such as field notes and field tapes, are supplemented by interviews or discussions with participants. Combining data from observations and interviews can also be described as *cross-validation* (Cohen et al., 2007). Observations by the researcher can produce rich descriptions of actions performed in the setting, while interviews enable the members of a work organization to voice their experiences. In any case, the importance of fieldwork in initiating or advancing the researcher's ongoing process of interpretation is crucial; observing and interviewing the participants are the main means by which a comprehensive description of the phenomenon or culture is constructed (Heyl, 2001). Observations can be conducted by monitoring people in selected contexts or shadowing key persons in their everyday work (Hammersley & Atkinson, 2007). The focus of observations can be on different agents and participants in an organization, such as representatives of professional groups and teams (Emerson et al., 2001). The duration of shadowing or observing work can vary from a few hours to several days. It might sometimes be appropriate to shadow an entire work shift (e.g., in a hospital) or to shadow only for a short period (e.g., in an office). Observations can be analyzed as primary empirical data, or if they are not the focus of the analysis, they can offer useful contextual information in locating key persons and key incidents within their setting.

During fieldwork, the researcher makes choices about what is relevant to the research and the situations in which the observations should be made. These events (i.e., incidents) can be described as arenas in which the activity that forms the field takes place (Zilber, 2014). *Key events* are interesting situations that occur in the field, which the researcher recognizes during fieldwork and/or afterwards, when analyzing their field notes (Lemmetty, 2020a). Emerson (2004) described three factors as characteristics of key events: (1) they are not necessarily dramatic or profoundly relevant from the perspective of the examinees; (2) they are part of everyday life; and (3) their significance is not necessarily clear at the start of the analysis, but they arouse the researcher's interest in examining them more closely. Because the processes of data collection and analysis partly overlap in ethnography, a preliminary understanding of key events arises during observations but is deepened and refined in the analysis stage. Zilber (2014) presented several questions that can be asked to support the analysis of key events:

> Who organized the event? Are they central or peripheral actors in the field? What are the (manifested and hidden) interests of the event? Is this event part of a series of field-level events? Or perhaps the first or the last instance of a series? Does the studied event attract a variety of field-level actors or a specific subsection? What other compatible events take place in the field? What is the centrality of the studied event, given those other events and compared with them? (p. 105)

Many methods of analysis have been applied to the ethnography of workplace learning (e.g., Collin, 2006; Lemmetty & Collin, 2020; Lemmetty, 2020a, b). Ethnographic data should be analyzed in its context, considering cultural and other contextual aspects (e.g., work aims and tasks and the nature of work). Thus, in ethnographic research conducted in the contexts of workplaces and organizations, it is typical to analyze discourses, practices, cultures, structures, and interactions. Within these frames, many basic qualitative analysis methods might be appropriate for addressing the research questions. For example, when focusing on content, thematic content analysis may be a suitable analysis method, whereas discourse analysis might be appropriate for analyzing interactions and discussions. Ethnographic analysis usually aims to create a multifaceted picture of the phenomenon at hand, which is enabled by combining different analytical tools with ethnographic analysis.

18.2.2 Studying Workplace Learning in Virtual Environments

Virtual ethnography is implemented in ways that differ from traditional face-to-face ethnography, for instance, by observing interactions in digital environments or on the internet (Purli, 2007; Kozinets, 2015). It has been described as a technique that relies on virtual participant observations of interactions in online chat rooms, group blogs, social networks, and discussion boards. However, workplace learning has rarely been studied by virtual ethnography. The area can therefore be considered partially undeveloped as a methodology. Because (expert) work is increasingly being done remotely, people more often interact in virtual environments in traditional workplaces (e.g., in Slack, Teams, and Hangouts). Thus, it is suggested that a remote environment also affects learning.

Workplace learning is often categorized as *informal* or *formal* (Lave & Wenger, 1991; Marsick & Watkins, 1990). Formal learning refers to learning in structured situations such as training or courses (Tynjälä, 2008). Since the beginning of the 2000s, in addition to in-house training, online courses and training have been utilized in the development of employees' competence and skills. They have been seen as useful because they often offer opportunities for flexible learning that is detached from place and time. They also leave the individual with greater responsibility for his or her own learning. For example, employees can decide when and how to learn (Conn, 2000). Informal learning refers to all learning that takes place during work and through work practices (Janssen et al., 2017; Kyndt & Baert, 2013). Acquiring new knowledge through developmental work, sharing experiences with colleagues, and collective problem solving have been considered forms of informal learning at work. As remote work and virtual work environments become more common, informal learning at work will increasingly take place within the framework of virtual environments.

However, the distinction between informal and formal learning has also been criticized, as it implies that work and jobs are informal in terms of learning, but pedagogical frameworks that influence the progression and planning of learning processes, artifacts, and practices are also emerging in workplaces (Billett, 2014). Thus, the following question has been raised: Why is learning in the workplace less formal than the learning and training that occurs in schools and courses? As access to virtual work environments and digital learning tools has increased, it seems that the boundaries between informal and formal learning have become increasingly blurred. As learning requirements grow, it should be possible to continuously develop new knowledge and skills during work (Lemmetty, 2020b). Virtual environments can aid learning processes, but they also place constraints and delays on learning. However, very little research is available on the pedagogical characteristics, frameworks, artifacts, and individual or collective practices that emerge in virtual work environments and how they affect learning processes and culture. Because virtual ethnography can provide useful approaches and tools for examining these phenomena, its importance in future research on adult education should not be underestimated.

One aspect of the study of learning at work is the communal and the practical actions attached to it. The term *community of practice* refers to a group of people whose members share a particular skill or competence (Lave & Wenger, 1991; Fuller & Unwin, 2003). A community can be built around a shared goal, or it can be formed to create or compile the necessary information. The sharing of different individuals' knowledge is essential in a community of practice.

In workplace learning, a community of practice is often a group of workplace members who share information and learn from each other. The sharing of information and participation in a community of practice requires interactions and meetings between community members. As workplaces become increasingly multifaceted, communities of practice not only combine or engender face-to-face encounters but also function through a variety of communication tools. Ethnographic research, which can be used to examine the activities of communities of practice at the individual, community, and organizational levels, offers opportunities for investigating the formation of virtually-functioning communities of practice. Online ethnography is less obtrusive, more accessible, and more convenient than conventional ethnography. Furthermore, it makes possible higher degrees of participation in observed communities.

18.2.3 Subjective Evidence-Based Ethnography and the Study of Human Actions

One of the most ambitious aims of ethnography is to capture actions and thoughts in their context—which is the workplace environment in this case—and as they unfold. This requires being able to relate what people do and say to where they are, what they use, and with whom they interact. In ethnographic investigations, these psychological, social, and material dimensions are combined in ways that are rare in other methodologies and acutely needed in scientific research (Glăveanu, 2020a). However, ethnographers face a great practical and methodological challenge: they are tasked with the difficult job of recording activity in an unintrusive and descriptive manner and focusing on behavior (e.g., the thoughts, emotions, and intentions of the observed individuals) to make inferences from it. Of course, the key issue is that the participants cannot both perform their activity and express reflections about what they think and how they feel while performing it. This is the longstanding challenge of introspection—the impossibility of being both observer and observed at the same time (Danziger, 1980). In ethnography, it is not feasible to ask participants to play both roles, so their thoughts and intentions are usually inferred later, based on an analysis of field notes. Although there are sustained guidelines on how to write field notes efficiently and in a way that does justice to the lived experiences of both researchers and participants (Clifford, 1990), ethnographic descriptions remain intrinsically selective. However, this is not necessarily a limitation because seasoned ethnographers are especially skilled in focusing on and conveying what is essential in their writing and reflections, but novice researchers could experience it as a formidable challenge.

One way to address the various difficulties associated with ethnographic work is to conduct *subjective evidence-based ethnography* (SEBE) (Lahlou et al., 2015; Glăveanu, 2019). This method is based on the use of “subjective cameras,” which are audio and video devices placed at eye level to record human activity from the perspective of the participant. SEBE is considered *subjective* because it helps researchers capture the first-person perspectives of people wearing a subjective camera. It is *evidence-based* because this perspective is documented through the use of a recording that becomes material evidence that is then analyzed by the researcher together with the participant. Finally, it is *ethnographic* because it reveals the path of human action in its context. Indeed, with the help of subjective cameras, while we cannot not see participants *per se*, we can notice a great deal about their environments and their interactions within these environments.

Thus, the SEBE methodology is particularly well suited to the study of human activity, which can be underpinned by several theories, including *activity theory* (Engeström & Kerosuo, 2007) and *pragmatist accounts of action* (Joas, 1993). In particular, the latter is based on the premise that we at all times occupy various positions in the world, which are simultaneously physical, social, and symbolic. For example, we are positioned through our bodies, by the social roles we adopt, and in discourse and culture. Specific to human action, however, is the fact that we are never "trapped" in a single position (Gillespie & Martin, 2014). On the contrary, we often exchange positions and include their respective perspectives in dialogues that expand our spheres of possible action (Glăveanu, 2020b). These acts of repositioning and exploration of the possible, in addition to their associated constraints, are enabled by social and material interactions. In this regard, SEBE has proven to be particularly useful; it helps us document the positions and perspectives of the participants as they unfold in relation to the positions and perspectives of others within a rich material context.

A unique feature of SEBE is that it cuts across classic dichotomies, such as those between the subjective and the objective. Practically, the video recording offers an "objective" account of what was seen, heard, and done by the participant. In the interview, the participant complements this account with a subjective retelling of their actions, intentions, beliefs, and emotions. However, most importantly, SEBE operates at an intersubjective level, as the research findings are not derived from the video or the interview alone but from the account co-constructed by the researcher and the participant. Their dialogue illuminates the recorded actions and helps organize them according to specific frameworks, while simultaneously refining and advancing these frameworks.

Another notable research dichotomy transgressed by the use of this innovative methodology is that between thought and behavior. Many behaviorist methods and techniques have been used in organizational research (e.g., Ebert & Freibichler, 2017), but there is also emergent phenomenological literature that emphasizes lived experience (e.g., Anosike et al., 2012). Through the use of SEBE, we are able to bring together the study of behavior and the study of affect and cognition in different stages, from the recording to the interview. This avoids a common problem in ethnographic studies caused by the use of think-aloud protocols (Jääskeläinen, 1998) that ask participants to continuously narrate their flow of consciousness. In such cases, the participant's course of action is disturbed. In SEBE, think-aloud protocols are not necessary because first-person recording offers the unique possibility of resituating streams of consciousness within the pathway of action.

The final dichotomy transgressed by SEBE is of an ethical nature and concerns the power balance in the relationship between the researcher and the participants. Traditionally, when the participant agrees to take part in the study, the researcher oversees all aspects of the research. Participants can withdraw or have their data withdrawn, but in most cases, they act as passive actors in the research process. However, in SEBE, the participants are granted the power to start and stop the recording as they see fit, to watch it first if they are interested in doing so, and to return all or select parts of the recording to the researcher. While this autonomy might seem to create a power imbalance that favors the participants, the emphasis is on trust and the effort to build trust, which is the hallmark of any ethical relationship.

In view of these benefits, where is SEBE applied? It is used in a variety of contexts, from the transfer of professional expertise in industry and power plants to studying the ways in which Polish mothers take care of their children (Lahlou et al., 2015). Notably, this methodology has also been used to study creativity, specifically the practices of Easter egg decoration by craftswomen in northern Romania (Glăveanu & Lahlou, 2012). In these distinctive contexts, SEBE has been used to provide evidence of the complex dynamics through which people engage with their social and material environments and learn from them.

Because creative learning has a perspectival nature (Glăveanu et al., 2019), this methodology has many promising applications, including in workplace learning. In other words, participants learn from the intersection of their perspectives with those of others, and both can be well evidenced with the help of subjective cameras because people's actions reveal their position, perspective, and repositioning. In both learning and creativity, we reposition ourselves vis à vis others or the problem at hand. The ability to capture the moments in which such shifts in action and knowledge occur and to study them would be extremely valuable for workplace learning scholars. Moreover, having multiple participants in a workplace wear such cameras could help document how various perspectives harmonize or conflict. Hence, in SEBE, interviews are not only an occasion to study perspectival forms of workplace learning but also to actively stimulate the perspectives and processes of mutual learning and knowledge construction (Ness, 2017). Thus, SEBE could be a research and intervention tool that is useful for researchers and practitioners alike.

18.3 Empirical Case Examples

Next, we present two empirical examples. The first example illustrates the study of informal learning using ethnography in a virtual environment, and the example second illustrates the use of subjective cameras in a SEBE process. Using these examples, we show how the basic principles of ethnography can be strengthened and applied in virtual environments and with the help of digital tools. Also, we aim to provide an understanding of the benefits and limitations of virtual ethnography and digital tools in ethnographic research.

18.3.1 *Virtual Ethnography in Slack*

In a recent research project called HeRMo: Ethical Human Resource Management Supporting Creativity in Finnish Growth Companies (2018–2022), we examined employee creativity and learning in technology work (Collin et al., 2021; Lemmetty & Collin, 2020). One of the aims of our research project was to understand personnel's learning at work in different work environments and situations. In technology-based work, there is an increasing demand for competence development and workplace learning, perhaps more so than in any other industry. For example, the demands of learning are well reflected in the fact that working in the software industry requires a deep understanding of key processes. In technological work, continuous learning is not only a requirement of companies and the industry; it is also motivated by the employees themselves (Fuller & Unwin, 2003; Riddell et al., 2009) and considered an essential part of organizational operations (Scheeres et al., 2010). The development of technologies is reflected in everyday life in concrete terms; it does not make sense to formally teach employees about the latest technologies, as those technologies may be old in a week, a month, or a year (Lemmetty & Collin, 2020). Work in the field of technology is strongly problem-based, and employees are involved in short-term loops of problem-driven learning (e.g., Havnes & Smeby, 2014; Nerland, 2008; Collin, 2006). Thus, it is suggested that learning in technology work takes place informally in everyday work and related interaction situations.

One essential environment for interaction in the field of technology is the virtual discussion environment, in which employees have the opportunity to seek advice and receive guidance from colleagues regardless of their physical location. For this reason, we decided to explore the manifestation of learning in virtual discussion channels, which we examined through virtual ethnography. In our case study, we asked: What kinds of learning situations manifest in the virtual discussion platform, and what are the strengths or challenges of researching learning in the virtual discussion platform? We selected one of the target organizations of our project as the case organization for our virtual ethnographic study: a medium-sized technology company in which staff have access to the Slack discussion channel. Participants in our case study—designers, engineers, developers, and other expert employees—developed new technological applications, environments, and solutions to meet the needs of different industries and customers but also used digital environments in their own everyday work. The data from the case study comprised 260 pages of saved authentic discussions from a virtual channel. Based on our case study, we highlight how we applied the basic principles of ethnography—its concepts and methods—in our virtual ethnographic study.

18.3.1.1 Entering the Virtual Field

Entering the field is the starting point for ethnographic research. When the object of research is a work organization and the field is understood to be the space in which processes, social interaction, goals, and boundaries emerge (Burrell, 2009), it can be said that a field is not one and the same permanent state or moment but rather an entity that includes a wide variety of environments and situations within an organization and its operations. At the beginning of the HeRMo project, the basic operations and practices of the organization were discussed with two key people. At this stage, these key people described that in addition to physical workspaces and situations, the virtual Slack environment is an important environment for staff interaction. In Slack, things that come up are discussed, support and feedback are received from others, and problems are solved together. For this reason, it was agreed that, in addition to traditional observation and interview data, the discussions in the Slack channel would also be used as research material. This was also justified from the perspective of the *constructivist* conception of knowledge that guided the research. From the constructivist viewpoint, learning is built on social and linguistic interactions (Gergen, 1999), and its construction is observed more so at the level of the community (Bergen & Luckmann, 1996) than at the level of individual psychological structures and processes (Gergen, 1999). Thus, by focusing on interactions between people, ethnographic observation in the field offers researchers the possibility of understanding the sociocultural nature of workplace learning.

Our research team gained access to the Slack channel. The platform had several discussion channels thematically grouped around different topics. There were two channels for everyday and open discussion: (1) a channel focusing on work-related issues and (2) a channel for general and informal discussions. The research process began by exploring the channels and making general observations. First, it was found that several short discussions took place on both channels on a daily basis. Reading previous conversations and following up on ongoing discussions made it possible to make our first observations. In the informal discussion channel, the interactions were largely focused on workplace humor, discussion about leisure activities, and finding a lunch venue. The conversation was sporadic, made up of changing and occasional topics. In contrast, in the work-related channel, the discussions focused on the challenges and problems encountered at work, requests for help, and the sharing of information, as can be seen from the material example below.

Nickname A [12:12 PM]

> In this context [refers to a previous discussion], what is the best way to share individual documents to clients (e.g. a meeting checklist)? Email directly to customers, right?

Nickname B [12:15 PM]

> If you have to share [indicates the storage location of the folder], then the entire folder could be shared with the sharing settings “[indicates that setting].”

Nickname B [12:15]

> But are there any other ways to share meeting checklists? [Discussion continues…]

From the point of view of conducting research on workplace learning, work-related discussions seemed to provide the most relevant material for the purposes of our study. For this reason, it was decided that we would limit our data collection to the work channel. After a preliminary understanding of the discussion channels, our actual virtual fieldwork began.

Several factors are involved in planning and conducting fieldwork, such as the times at which and situations in which the researcher may be present. The researcher must accept that it is not possible to do fieldwork around the clock, and temporal boundaries must be set (Tuncalp & Le, 2014). To ensure that research data are adequate, it is necessary to assess how extended periods of fieldwork provide sufficient and reliable data (Hammersley & Atkinson, 2007). Thus, once our understanding of the scope, pace, and amount of discussion in the channel under study was tentatively established, we made a detailed plan for data collection. The researcher saved Slack channel conversations for 1 week. The researcher's role in the channel was to be a silent follower and observer—not a participant. Thus, the researcher did not influence the discussion in any way. So that an understanding of the field and phenomenon under study could be constructed, the researcher read through the discussions of each day at the time the recordings were saved. It is important to note that the researcher who followed the Slack discussion also made physical observations and conducted interviews in the organization during the same period. In this way, the researcher's interpretation of the Slack data was combined with other observations made by the researcher. During the saving and reading of the Slack material, remarks and notes were written about the discussion. After 1 week of data collection, it was noticed that certain features were repeated throughout much of the discussions. Also, it was initially possible to ensure that the material provided an understanding related to the themes and questions of the study.

18.3.1.2 Glimpses of Learning Processes Through Events

To develop an understanding of the target organization and its learning situations, it is necessary to study the actors, their processes and practices, the frames that limit or enable them, and the goals that the organization seeks to achieve (Zilber, 2014; Lemmetty, 2020b; Collin, 2006). From this perspective, we wanted to look at the kinds of learning situations manifested on the Slack channel. We found that many questions were asked on the Slack channel, which were related to, for example, experiences of good practice, challenges with customers, lack of information, or individual solutions. Also, employees asked for tips on where or from whom to get information. The example below shows a typical discussion in which an employee put a description of a problem they were facing on the channel and received a response from a coworker in less than 1 min, after which other people participated in the conversation. The event can be seen as a key event from the perspective of the learning situation, as it reflects a typical problem-solving event that was observed in the data.

Nickname A [11:45 AM]

> does anyone else have a problem with the Hangouts Chrome extension blinking orange on the taskbar even though no new messages have arrived?

Nickname B [11:46 AM]

yes! :angry:

Nickname C [11:51 AM]

Yes, same problem here…

Nickname D [11:58 AM]

Yes

Nickname E [12:02 PM]

It's because if you're signed into Hangouts in Gmail and then, because it's a separate software, it will blink before you go to Gmail to read it. So you have to sign out of Hangouts in Gmail. [Colleague] talked about this.

Nickname F [12:38 PM]

It blinks here too! How do I sign out of Hangouts in Gmail?

Nickname G [12:40 PM]

Click on your own image to enter the menu

Nickname F [12:40 PM]

Oh!

Nickname E [1:14 PM]

For clarification: click on the image on the left. If you click the image on the right, everything disappears.

Nickname F [1:15 PM]

Yep, I found it! :slightly_smiling_face:

In examining a research target, problems are often identified regarding the extent of the field and its delineation. This reveals a particularly interesting question in the study of learning: When can we consider a learning situation to have begun or ended? According to the constructivist view of learning, learning is constructed based on previous understanding and knowledge; it is an ongoing process (Wenger, 2009). Organizational learning projects are increasingly complex entities, and they operate across different environments. For example, training may be completed at an educational institution, but the application of what is learned takes place in the workplace. This perspective emerged strongly in the Slack material, especially when considering the benefits or limitations in the study of learning. If discussion on a virtual platform is the only material used in a research project, it excludes many other observations regarding learning processes. The material example below illustrates this challenge in concrete terms:

Nickname A [9:37 AM]

Hello everyone, Customer asked how easy it is to implement dynamic pricing in a ticket store? Any vision for this? In addition, tickets should be able to be priced differently depending on the time the ticket was purchased

Nickname B [9:38 AM]

No. you can come to talk with the [specific team name] team. We can show you the current implementation.

Nickname A [9:38 AM]

Alright, I will come there soon!

This example shows a glimpse of the learning process and that it was triggered by a customer question. However, an adequate response would seem to require a joint discussion with a particular team, as well as a demonstration of the "current implementation" mentioned in the example. For this reason, the conversation was not continued on the Slack channel, and the questioner moved from his workstation to engage in face-to-face conversations with the other team. Thus, it remains unclear how the incident progressed and how it ended. Several similar examples were found in the data. From this, it can also be seen that Slack serves as a good platform in the first stages of the learning process; however, in complex learning situations that require a great deal of discussion and illustration, Slack does not seem to be a useful environment.

In workplace learning research, identifying the learning context has been described as an important starting point in considering what people need to learn in a specific organization and why (e.g., Collin, 2006). In organizational research, the ethnographer links their observations within a broader context and at a specific time, in a particular place and sociocultural space, which contributes to the formation of theoretical generalizations. *Contextualization* refers to connecting data to the cultural environment and the historical macro level of time and place. For example, based on our research in technological organizations, learning seems to be highly intertwined with the problem-solving situations that emerge in everyday work and interaction, such as finding answers to client questions (see also Lemmetty & Collin, 2020). Moreover, in this field, learning is linked to the technological skills and competencies that people need to perform their work (Lemmetty, 2020b).

18.3.2 A Subjective Evidence-Based Ethnography (SEBE) Study of Craft Practices

A SEBE project usually follows a series of steps (Glăveanu & Lahlou, 2012). First, as in any other research process, a question needs to be formulated, which in this case was a question that is amenable to the purposes of ethnography (e.g., how participants act and interact in specific situations, how they learn, or how they create). Second, one or more participants should be identified whose experience is relevant to the research question and who is available for an in-depth study of their actions, including filming. Third, participants are asked to wear a camera while performing the activities of interest, often continuously, for a long time. This requirement is uncomfortable because the small camera is either part of a readymade headset or placed on things that are used regularly by the person (e.g., a pair of glasses) or are easy to use (e.g., a visor). It is also likely that participants are inconvenienced in psychological terms by wearing a camera, although research has shown that they soon forget that their activity is being recorded. Moreover, they have full control over which activities are filmed and whether the recording is returned to the researcher. Fourth, the researcher reviews the recording and considers it through the lenses of the initial questions and the theoretical framework. At this stage, it is important to remain open to being surprised by the data and to use the findings to question assumptions. Finally, one of the most important parts of SEBE is an interview between the researcher and the participant based on the findings from the recorded data. If time allows, the interview can include viewing the entire recording. The interview can be conducted in multiple sessions or, more commonly, in critical moments that capture something of interest from the researcher's perspective. One of the most useful aspects of SEBE is that, at this stage, because participants view a first-person recording of their actions, they can be more easily re-situated within it, and even after 1- or 2-week intervals, they can remember surprisingly well what they were doing, what they were about to do next, and, most importantly, what they thought or felt at specific points in time (Tulving, 1972).

The interview ultimately allows the researcher and participant to construct a joint account of what happened—one that includes more than a retelling of behaviors but is a true account of human actions comprised of intentions, meanings, reactions, and the effects of materials and other people on a participant's activity.

In what follows, we will briefly illustrate the use of SEBE by referring to a project focused on learning and creativity in community contexts in the case of craft practices (for details, see Glăveanu & Lahlou, 2012; Glăveanu, 2013). The main aim of this research was to understand how creative actions unfold in the case of experienced and novice Easter egg decorators in north Romanian communities. While this does not directly concern workplace contexts, key methodological lessons can be derived from this case study.

18.3.2.1 Setting Up SEBE

One of the most important moments in conducting an SEBE has to do with entering the field. Because subjective cameras allow participants to record their own activities, it is crucial to first establish a close and ethical relationship with them—one based on transparency and accountability. In the craft case study, this was accomplished by starting with interviews of the participants that established the main points of reference concerning their activities and situated these activities within their life course. Through this, it was possible to choose which work sequences should be observed. In the research referred to here, the choice was made to include experienced egg decorators and novices (children) and observe them within a familiar context: painting eggs at home or at the local museum. In this way, adding the task of recording the session caused minimal disturbance to the participants' routines and allowed them to focus on the work itself rather than the presence of the researcher.

Careful consideration must also be given to when the observed creative learning process begins and ends. This is important given that decorating eggs is not always a linear activity, and the unit of analysis can be established at the level of one egg or a batch of eggs worked on at the same time. Finally, the position of the subjective camera needs to be checked, and researchers ought to help participants set up their cameras (e.g., attach them to a pair of glasses if they have glasses) and explain how the recording starts and stops. Ethically speaking, the participants can use their right to withdraw from the research even after the recording has been made, and they can decide whether to give the tape to the researcher or only provide parts of it.

18.3.2.2 Analyzing SEBE Data

One distinguishing characteristic of the data obtained through SEBE is that it can easily overburden the researcher. Recording videos, either with a subjective or a more "traditional" kind of camera, leaves younger researchers with considerable material to analyze. The important thing about SEBE is that video analysis is only part of the analytical process. With the egg decorators, the main interactions being analyzed were those between researcher and participant as they watched the subjective camera recording. In this case, due to the limited number of hours being recorded, it was possible to look over a whole video quickly. The focus was on the "pathways" of different actions (e.g., what kinds of phases or stages are specific to decoration and what their succession is) and how these action patterns differed between experts and novices. Unsurprisingly, the former took less time than the latter to decorate, and they also presented more habitual forms of painting eggs. In contrast, novices often went back and forth between stages and were sometimes a bit less certain of the outcome. SEBE data can be analyzed using different analytical procedures, from content and thematic analysis to conversational and discursive forms. This is because the main piece of data is not the video itself but the actions and interactions taking place in the researcher–researched–recording triad. Ultimately, this is where creativity unfolds and becomes manifest.

18.4 Concluding Thoughts

In this chapter, we have described the idea and aim of ethnography, explained different ethnographic tools, and reflected on emerging future directions for utilizing virtual ethnography in researching workplace and organizational learning. Based on the core ethnographic concepts of field, context, and key events, we have described the roots and nature of ethnography as a methodology used in researching the social worlds of work as interpreted by acting and experiencing individuals and communities. Thus, at its best, ethnography connects and cuts across the objective and subjective worlds of work and learning, thus enabling researchers to grasp the phenomena at hand in a multifaceted manner. The findings of such research may lead to a deeper understanding of phenomena, which is sorely needed in scientific research because of the fast pace of changes in working life. This connection among psychological, social, and material dimensions is rare in other methodologies. Without close and detailed studies of the worlds of work, our conceptions of them will remain inadequate. Therefore, ethnography can be utilized to shed light on human actions that are rarely, if ever, predictable or in line with current theory or expert opinion. In organizational research, ethnographic work often takes issue with managerial claims or with highly generalized concepts that say little but assume much about what people do and think in their everyday working lives, including real actions that occur (or do not occur) in work settings.

However, changing requirements have challenged the work that people do and the learning embedded in work practices. As we have moved toward a deeper understanding of learning in workplaces and organizations, we have witnessed the emerging importance of remote work, virtual work, and the variety of digital tools utilized at work. By approaching learning in organizations in a mainly informal way, employees can no longer avoid learning in and through technological tools, apps, blogs, Google, YouTube, and virtual discussion platforms (Lemmetty & Collin, 2020). Therefore, we need new tools to adopt our research methods in the study of these changing learning environments. When the nature of learning changes, we need new applicable methodologies, such as virtual ethnography, to achieve a deeper understanding of learning in workplaces and organizations. As a method and methodological approach, ethnography needs to evolve.

The potential of virtual ethnography and SEBE—the empirical case examples of new ethnographic tools provided in this chapter—for the study of workplace learning and multi-site working life is tremendous. However, in embarking on research, it is important to understand several ethical issues and the choices a researcher should make during the research process. The researcher's critical reflection on their role is especially important in virtual ethnography. Furthermore, because it is easy to enter the field, the informant's protection and consent might be neglected. It is also important for the researcher to consider ethical issues in relation to the goal and purpose of the research and to ensure that their choices are ethical and that the phases of the research process are visible. Clear guidelines for ethical reflections and choices in virtual ethnography do not always exist. However, Boellstorff et al.'s (2012) *Second Life* and Kozniets' (2015) *Netnography: Redefined* provided useful starting points.

Despite its laborious nature and the variety of ethical issues that need careful consideration throughout the research process, ethnography offers a methodology with a rich range of tools for the future study of learning in workplaces and organizations. Because ethnography is increasingly utilized, it may provide a continuously developing methodology that can meet the requirements for a deeper understanding of learning, especially in increasingly multifaceted areas of working life. Connecting new methods with the spirit of ethnography and considering ethical issues will lead researchers, especially ethnographers, along many interesting developmental paths. Moreover, it is possible that future research could return to the roots of ethnographic discovery because "the most important element of fieldwork is being there" (Fetterman, 2010), regardless of what "there" might mean in the future.

References

Anosike, P., Ehrich, L. C., & Ahmed, P. (2012). Phenomenology as a method for exploring management practice. *International Journal of Management Practice, 5*(3), 205–224.

Baskett, H. K. M. (1993). *Workplace factors which enhance self-directed learning*. Seventh International Symposium on Self-Directed Learning, West Palm Beach, FL, USA, pp. 21–23. https://files.eric.ed.gov/fulltext/ED359354.pdf

Berger, P. L., & Luckmann, T. (1996). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor Books.

Billett, S. (2014). Mimesis: Learning through everyday activities and interactions at work. *Human Resource Development Review, 13*(4), 462–482. https://doi.org/10.1177/1534484314548275

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. (2012). *Ethnography and virtual worlds: A handbook of method*. Princeton University Press.

Burrell, J. (2009). The field site as a network: A strategy for locating ethnographic research. *Field Methods, 21*(2), 181–199. https://doi.org/10.1177/1525822X08329699

Clifford, J. (1990). Notes on fieldnotes. In R. Sanjek (Ed.), *Fieldnotes: The makings of anthropology* (pp. 47–70). Cornell University Press.

Coffey, A. (1999). *The ethnographic self: Fieldwork and the representation of identity*. Sage. https://doi.org/10.4135/9780857020048

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge.

Collin, K., Herranen, S., Paloniemi, S., Auvinen, T., Riivari, E., Sintonen, T., & Lemmetty, S. (2018). Leadership as an enabler of professional agency and creativity in information technology organisations. *International Journal of Training and Development, 22*(3), 222–232. https://doi.org/10.1111/ijtd.12130

Collin, K. (2006). Connecting work and learning: Design engineers' learning at work. *Journal of Workplace Learning, 18*(7/8), 403–413.

Conn, A. N. (2000). *Self-directed learning in the workplace*. Doctoral dissertation, University of Maryland.

Collin, K., Lemmetty, S., Forsman, P., Glăveanu, V. P., Auvinen, T., Riivari, E., Keronen, S., & Jaakkola, M. (2021). Supporting creativity and learning at work: Practices and structures from growth companies. In S. Lemmetty, K. Collin, V. P. Glăveanu, & P. Forsman (Eds.), *Creativity and learning: Contexts, processes and support* (pp. 245–272). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-77066-2_11

Croucher, S. M., & Cronn-Mills, D. (2015). *Understanding communication research methods. A theoretical and practical approach*. Routledge.

Danziger, K. (1980). The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences, 16*(3), 241–262.

Davies, C. A. (1999). *Reflexive ethnography. A guide to researching selves and others*. Routledge.

Ebert, P., & Freibichler, W. (2017). Nudge management: Applying behavioural science to increase knowledge worker productivity. *Journal of Organization Design, 6*(1), 1–6.

Emerson, M. R. (2004). Working with "key incidents". In C. Seal, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 457–472). Sage.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2001). Participant observation and fieldnotes. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 352–368). Sage.

Engeström, Y., & Kerosuo, H. (2007). From workplace learning to inter-organizational learning and back: The contribution of activity theory. *Journal of Workplace Learning, 19*(6), 336–342.

Fetterman, D. M. (2010). *Ethnography step by step*. Sage.

Fingerroos, O. (2003). Refeksiivinen paikantaminen kulttuurien tutkimuksessa [Reflexive localization in the study of cultures]. *Elore, 10*(2) available at: https://doi.org/10.30666/elore.78407

Fuller, A., & Unwin, L. (2003). Learning as apprentices in the contemporary UK workplace: Creating and managing expansive and restrictive participation. *Journal of Education and Work, 16*(4), 408–426. https://doi.org/10.1080/1363908032000093012

Gergen, K. J. (1999). *An invitation to social construction*. Sage.
Gillespie, A., & Martin, J. (2014). Position exchange theory: A socio-material basis for discursive and psychological positioning. *New Ideas in Psychology, 32*, 73–79.
Glăveanu, V. P. (2013). Creativity and folk art: A study of creative action in traditional craft. *Psychology of Aesthetics, Creativity, and the Arts, 7*(2), 140–154.
Glăveanu, V. P. (2019). Studying creativity as a social process: The use of subjective cameras. In I. Lebuda & V. P. Glăveanu (Eds.), *The Palgrave handbook of social creativity research* (pp. 163–173). Palgrave.
Glăveanu, V. P. (2020a). A sociocultural theory of creativity: Bridging the social, the material, and the psychological. *Review of General Psychology, 24*(4), 335–354.
Glăveanu, V. P. (2020b). *The possible: A sociocultural theory*. Oxford University Press.
Glăveanu, V. P., & Lahlou, S. (2012). Through the creator's eyes: Using the subjective camera to study craft creativity. *Creativity Research Journal, 24*(2–3), 152–162.
Glăveanu, V. P., Ness, J. I., Wasson, B., & Lubart, T. (2019). Sociocultural perspectives on creativity, learning, and technology. In C. A. Mullen (Ed.), *Creativity under duress in education? Resistive theories, practices, and actions* (pp. 63–82). Springer.
Havnes, A., & Smeby, J.-C. (2014). Professional development and profession. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 915–986). Springer. https://doi.org/10.1007/978-94-017-8902-8
Hägg-Martinell, A., Hult, H., Henriksson, P., & Kiessling, A. (2017). Medical students' opportunities to participate and learn from activities at an internal medicine ward: An ethnographic study. *BMJ Open, 7*(2), e013046. https://doi.org/10.1136/bmjopen-2016-013046. PMID: 28196948; PMCID: PMC5318643.
Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice*. Routledge.
Heyl, B. S. (2001). Ethnographic interviewing. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 369–383). Sage.
Jääskeläinen, R. (1998). Think-aloud protocols. In M. Baker (Ed.), *Encyclopaedia of translation studies* (pp. 265–269). Routledge.
Janssen, L., Smet, S., Onghena, P., & Kyndt, E. (2017). The relationship between learning conditions in the workplace and informal learning outcomes: A study among police inspectors. *International Journal of Training and Developing, 28*(2), 92–112. https://doi.org/10.1111/ijtd.12095
Joas, H. (1993). *Pragmatism and social theory*. University of Chicago Press.
Kozinets, R. (2015). *Netnography: Redefined*. Sage.
Kunda, G. (2013). Reflections on becoming an ethnographer. *Journal of Organizational Ethnography, 2*(1), 4–22. https://doi.org/10.1108/JOE-12-2012-0061
Kyndt, E., & Baert, H. (2013). Antecedents of employees' involvement in work-related learning: A systematic review. *New Media & Society, 83*(2), 1506–1522. https://doi.org/10.1177/1461444817698776
Lahlou, S., Le Bellu, S., & Boesen-Mariani, S. (2015). Subjective evidence-based ethnography: Method and applications. *Integrative Psychological and Behavioral Science, 49*(2), 216–238.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University.
Lemmetty, S. (2020a). Employee opportunities for self-directed learning at technology organisations: Features and frames of self-directed learning projects. *Studies in Continuing Education, 43*(2), 139–155.
Lemmetty, S. (2020b). *"Self-learning is present every day—In fact, it's my job." Self-directed workplace learning in technology-based work*. Doctoral dissertation. University of Jyväskylä. https://jyx.jyu.fi/handle/123456789/71221
Lemmetty, S., & Collin, K. (2020). Self-directed learning in creative activity: An ethnographic study in technology-based work. *Journal of Creative Behaviour*. [early online]. https://doi.org/10.1002/jocb.438

Li, E.-S., & Bartunek, J. (2009). Research methods for organizational learning: The transatlantic gap. *Management Learning, 40*(4), 439–447.

Marjolein, G. M. C., Berings, A., Doornbos, J., & Simons, P. R. (2006). Methodological practices in on-the-job learning research. *Human Resource Development International, 9*(3), 333–363. https://doi.org/10.1080/1367886060089355

Marsick, V. J., & Watkins, K. E. (1990). *Informal and incidental learning in the workplace*. Routledge.

Murthy, D. (2013). Ethnographic research 2.0: The potentialities of emergent digital technologies for qualitative organizational research. *Journal of Organizational Ethnography, 2*(1), 23–36. https://doi.org/10.1108/JOE-01-2012-0008

Nerland, M. (2008). Knowledge cultures and the shaping of work-based learning: The case of computer engineering. *Vocations and Learning, 1*(1), 49–69. https://doi.org/10.1007/s12186-007-9002-x

Ness, I. J. (2017). Polyphonic orchestration-facilitating creative knowledge processes for innovation. *European Journal of Innovation Management, 20*(4), 557–577.

O'Reilly, K. (2011). *Ethnographic methods*. Routledge.

Ostroff, C. (2019). Contextualizing context in organizational research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 39–65). American Psychological Association. https://doi.org/10.1037/0000115-003

Paloniemi, S., & Collin, K. (2010). Discursive power and creativity in inter-professional work. *Vocations and Learning, 5*(1), 23–40. https://doi.org/10.1007/s12186-011-9064-7

Pastuhov, A., & Sivenius, A. (2020). Doing ethnographic research in adult education. In B. Grummel & F. Finnegan (Eds.), *Doing critical and creative research in adult education* (Case studies in methodology and theory). Brill.

Pole, C., & Morrison, M. (2003). *Ethnography for education: Doing qualitative research in educational settings*. Open University Press.

Purli, A. (2007). The web of insights: The art and practice of webnography. *International Journal of Market Research, 49*(3), 387–408.

Reed, I. A. (2017). Ethnography, theory and sociology as a human science: An interlocution. *Ethnography, 18*(1), 107–129.

Riddell, S., Ahlgren, L., & Weedon, E. (2009). Equity and lifelong learning: Lessons from workplace learning in Scottish SMEs. *International Journal of Lifelong Education, 28*(6), 777–795. https://doi.org/10.1080/02601370903293252

Riera Claret, C., Sahagún, M. Á., & Selva, C. (2020). Peer and informal learning among hospital doctors: An ethnographic study focused on routines, practices and relationships. *Journal of Workplace Learning, 32*(4), 285–301. https://doi.org/10.1108/JWL-11-2018-0141

Rouleau, L., de Rond, M., & Musca, G. (2014). From the ethnographic turn to new forms of organizational ethnography. *Journal of Organizational Ethnography, 3*(1), 2–9. https://doi.org/10.1108/JOE-02-2014-0006

Scheeres, H., Solomon, N., Boud, D. & Rooney, D. (2010). When is it OK to learn at work? The learning work of organisational practices. *Journal of Workplace Learning, 22*(1/2), 13–26. https://doi.org/10.1108/13665621011012825

Selznick, P. (1949). *T. V. A. and the grass roots*. University of California Press.

Snoeren, M., Raaijmakers, R., Niessen, T., & Abma, T. (2016). Mentoring with(in) care: A co-constructed autoethnography of mutual learning. *Journal of Organizational Behavior, 37*, 3–22.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic.

Tunçalp, D., & Lê, L. P. (2014). (Re)Locating boundaries: A systematic review of online ethnography. *Journal of Organizational Ethnography, 3*(1), 59–79. https://doi.org/10.1108/JOE-11-2012-0048

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154. https://doi.org/10.1016/j.edurev.2007.12.001

Valleala, U. M., Herranen, S., Collin, K., & Paloniemi, S. (2015). Fostering learning opportunities through employee participation amid organizational change. *Vocations and Learning, 8*(1), 1–34. https://doi.org/10.1007/s12186-014-9121-0
Van Maanen, J. (1986). Power in the bottle: Drinking patterns and social relations in a British police agency. In S. Srivasta (Ed.), *Executive power* (pp. 204–239). Jossey-Bass.
Watson, T. J. (2012). Making organisational ethnography. *Journal of Organizational Ethnography, 1*(1), 15–22. https://doi.org/10.1108/20466741211220615
Wenger, E. (2009). A social theory of learning. In K. Illeris (Ed.), *Contemporary theories of learning* (pp. 209–219). Routledge.
Yanow, D. (2009). Organizational ethnography and methodological angst: Myths and challenges in the field. *Qualitative Research in Organizations and Management, 4*(2), 186–199. https://doi.org/10.1108/17465640910978427
Zilber, T. (2014). Beyond a single organization: Challenges and opportunities in doing field level ethnography. *Journal of Organizational Ethnography, 3*(1), 96–113. https://doi.org/10.1108/JOE-11-2012-0043

Soila Lemmetty (PhD) works as a postdoctoral researcher at the Department of Education, University of Jyväskylä. Her research focuses on the phenomena of adult learning, wellbeing and creativity, which she examines in the context of working life. She has utilized an ethnographic strategy and a variety of qualitative methods of analysis extensively in her research. Lemmetty defended her PhD in 2020 and has worked at the University of Jyväskylä since 2015 in various research, teaching and other positions.

Kaija Collin (PhD) works as an Associate Professor at the University of Jyväskylä, Department of Education, Finland. Her research interests focus on workplace and organizational sustainable learning, creativity and leadership and professional development and HRD in various organizational contexts mainly with the help of qualitative methodology. Collin has widely published internationally and nationally. She also has long experience as an editor and reviewer in the fields of adult education, HRD, leadership and workplace learning. She also has a long experience in supervising master and doctoral students.

Vlad Glăveanu (PhD) is an Associate Professor and Head of the Department of Psychology and Counselling at Webster University Geneva, Associate Professor II at the Centre for the Science of Learning and Technology (SLATE), University of Bergen, Norway, founder and Director of the Webster Center for Creativity and Innovation (WCCI) and founder and President of the Possibility Studies Network (PSN). His work focuses on creativity, imagination, culture, collaboration, and societal challenges.

Susanna Paloniemi (PhD) is a senior lecturer in the Department of Education, University of Jyväskylä, Finland. Her current research and teaching interests include professional agency, workplace learning, and emotions in work organisations. She has utilised various research methods in her research, and thus holds an interest in methodological elaborations within the field. Her recent publications include articles in Frontline Learning Research, Studies in Continuing Education, and Professional Development in Education. She acts as an associate editor of Vocations and Learning journal.

Chapter 19
Video-Based Interaction Analysis: A Research and Training Method to Understand Workplace Learning and Professional Development

Laurent Filliettaz, Stéphanie Garcia, and Marianne Zogmal

Abstract This chapter presents video-based interaction analysis and discusses its contributions to research on workplace learning and professional development. Interaction analysis is a multidisciplinary qualitative approach, borrowing principles from the micro-sociology of everyday life, ethnomethodology, conversational analysis and the ethnography of communication. It aims to provide a detailed description of how individuals coordinate their actions when experiencing *social encounters* and engaging in goal-directed actions collectively. Over the past two decades, the use of interaction analysis has expanded significantly into the field of workplace practices in institutional or professional contexts, particularly thanks to the influence of Workplace Studies or applied conversation analysis. More recently, video-based interaction analysis has also been applied to the field of initial and continuing vocational education. The theoretical principles on which interaction analysis is based have been transposed to training activities and are now considered significant contributors to workplace learning and professional development. An increasing number of experiments have attempted to train professionals using a video-based interactive analysis of their work. After presenting the theoretical principles and methodological procedures of video-based interaction analysis, this chapter illustrates how the approach might be implemented in the specific empirical context of early childhood educators reflecting on their interactional competencies when encountering parents as part of their work. Data collected during collective analysis sessions illustrate the sorts of learning experiences made possible by video-based interaction analysis when it is used in continuing education programmes for qualified workers.

Keywords Interaction analysis, Video, Multimodality, Reflexivity, Continuing education, Interactional competences

L. Filliettaz (✉) · S. Garcia · M. Zogmal
Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland
e-mail: laurent.filliettaz@unige.ch

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_19

19.1 Introduction

In service-oriented economies, characterised by complex problem-solving tasks and constant interdependencies between workers, organisations are making increasing demands on their employees' communication skills and their capacity to engage effectively in verbal and non-verbal interactions with others. Such demands have translated into specific research and training methods that pay particular attention to talking's role in interactions and its place in workplace learning and professional development.

This chapter presents video-based interaction analysis, a research method inspired by the field of video-ethnography that focuses on how language and communication practices take place in work environments. Interaction analysis is a multidisciplinary field, borrowing principles from the micro-sociology of everyday life, ethnomethodology, conversational analysis and the ethnography of communication. Its objective is to describe in detail how individuals coordinate their actions when experiencing *social encounters* and use semiotic resources to accomplish goal-directed actions in a collaborative way.

Over the past two decades, the field of interaction analysis has expanded significantly in the domain of workplace practices and institutional or professional contexts, particularly under the influence of Workplace Studies or applied conversation analysis. More recently, video-based interaction analysis has also been applied in the field of initial and continuing vocational education. It has been used to investigate how novice workers are guided through internships or work placements (Filliettaz, 2014a, b; Koskela & Palukka, 2011), how they are taught in vocational schools (Filliettaz et al., 2010; Johansson et al., 2017; Kilbrink et al., 2021; Melander, 2017) or how they develop competencies in work contexts (Nguyen, 2017). The theoretical principles on which interactional analysis is based have been transposed into training activities and are now considered significant contributors to workplace learning and professional development. There have been an increasing number of experiences proposing to train professionals using video-based interactive analysis of their work (Stokoe, 2014). This chapter reports on these experiences and discusses the potential for this approach in advanced research and intervention methods for adult learning at work.

The chapter is divided into two main sections. The first section highlights the field of video-based interaction analysis from a conceptual perspective. It presents the epistemological, theoretical and methodological principles underlying the approach, as well as some of its contributions to understandings of work and vocational education. It goes on to show how the procedures related to collective data gathering sessions have been exploited as means of sustaining learning processes in continuing education. The chapter's second main section illustrates how such conceptual principles might be implemented in a specific empirical context, i.e. a continuing education programme for early childhood educators. The objectives and procedures related to this continuing education programme are briefly presented, as is its empirical research design. Data collected during collective analysis sessions illustrate the sorts of learning experiences made possible by video-based interaction analysis when it is implemented in continuing education programmes for qualified workers. Based on this empirical example, the potential of video-based interaction analysis for research and practice in the field of workplace learning and professional development is discussed and further elaborated.

19.2 Methodological Principles of Video-Based Interaction Analysis

19.2.1 Interaction Analysis as a Research Method

19.2.1.1 Origins and Definition

The field of interaction analysis originated in a set of new social sciences disciplines as they developed from the 1960s onwards in a number of approaches referred to as linguistic anthropology, socio-pragmatics, interactional sociolinguistics, ethnography of communication or conversation analysis. Despite the differences and controversies that characterise such frameworks, the promoters of these approaches share a common interest in the linguistic and communicational part of social practices, and they conceptualise language's use in social interactions as a constitutive component of human action in context. From there, verbal and non-verbal interactions are seen not only as subjects worthy of investigation in themselves but also more broadly as research *methods* through which social order and situated human actions can be investigated.

Following Erving Goffman's seminal work, face-to-face interactions can be defined as *social encounters*, specifically as "the reciprocal influence of individuals upon one another's action when in one another's immediate physical presence" (Goffman, 1956, p. 8). More precisely, *social encounters* refer to sequentially ordered processes through which participants accomplish joint actions and perform their reciprocal contributions in a coordinated way by using a range of semiotic resources, such as talk, gaze, gestures, body orientation or material objects. Face-to-face interactions are highly situated mechanisms in the sense that they are contingent on local material and practical arrangements and occur in social and cultural environments. Defined in this way, face-to-face interactions are usually assigned specific theoretical characteristics: (a) their ordered nature, (b) their sequential unfolding, (c) their multimodal accomplishment, and (d) their reliance on specific competencies.

19.2.1.2 Theoretical Principles

Following the principles of ethnomethodology (Garfinkel, 1967), face-to-face interactions can be conceptualised as the *methods* that groups of participants develop to address the practical problems they face when sharing the same interactional space. Such (ethno-)methods consist in producing organised ways of behaving that are made visible through observable lines of conducts and can be recognised by other participants as legitimate, relevant contributions. Social order is produced, negotiated and established by members of a community through situated, local behaviour and not because of pre-existing normative systems.

Among the various methods available for producing social order in interactions, temporality, progression and sequentiality can be regarded as powerful means of coordination in action. The founders of conversation analysis (Sacks et al., 1978; Schegloff, 2007) developed this idea in line with those of Garfinkel. Social interactions are not usually just static forms of reality. They unfold over time, step by step, and are built using principles such as turn-taking, overlapping talk, micro-pauses or the sequential order between recognisable actions (e.g. questions and answers, offers and acceptances). At a more macroscopic level, social encounters are also characterised by *opening* and *closing* procedures that stress the dynamic nature of verbal and non-verbal interactions. By ordering their contributions over time, participants accomplish coordination in an organised way.

Actions sequentially and collectively constructed by the machinery of situated interaction are not entirely based on speech and therefore do not rely exclusively on linguistic forms. If we accept, as Goffman (1964) stated, that "the natural home of speech is one in which speech is not always present", then we must also acknowledge that "many of the properties of talk will have to be seen as alternatives to, or functional equivalents of, extra-linguistic acts" (pp. 135–136). In line with multimodal approaches to discourse and interaction analysis (Kress et al., 2001; Goodwin, 2000; Mondada, 2016), the field of interaction analysis itself now commonly accepts that the meaning-making processes at work during interactions are based on a wide range of resources (e.g. speech, prosody, gestures, body postures, material objects or scriptural practices). Thus, within the context of these dynamic meaning-making processes, participants in interactions use a variety of resources that constitute multiple semiotic systems, referred to as *modes* (Kress et al., 2001). Hence, the meaning constructed in a context rarely results from the mobilisation of one single mode. On the contrary, it is frequently based on a combination of modes that are used simultaneously and complementarily, according to their own specificities, potentials and opportunities.

To mobilise such multimodal resources for coordination purposes, participants must develop highly contextual skills and abilities that are contingent on the situations in which their encounters take place. In the literature, these abilities have sometimes been referred to as "interactional competences" (Mondada, 2006; Nguyen, 2017; Pekarek Doehler et al., 2017). According to Young and Miller (2004, p. 520), interactional competence can be defined as the set of knowledge and skills that participants in interactions deploy to collectively configure the resources they need to engage in specific social practices. These competencies include the ways in which participants collectively accomplish actions in society, how they configure and delineate units of actions, and how they manage turn-taking rules, direct their attention, introduce new topics, take on and negotiate social roles, or use specific categories for referring to participants. Interactional competencies should not be thought of as an exhaustive repertoire of skills associated with individuals who are isolated from each other; on the contrary, they should be conceptualised as situated resources distributed among the participants involved in an interaction and made visible through the circumstances in which they are being enacted. With this in mind, one of interaction analysis's objectives is to identify the sorts of interactional competencies mobilised by participants when addressing the practical problems that they face in their social encounters.

19.2.1.3 Methodological Requirements

The theoretical principles mentioned above lead to specific methodological requirements. Analysing face-to-face interactions involves particular types of data, collected in the field, captured through audio-video recordings, and serving as the basis for multimodal transcripts. These methodological requirements are briefly summarised below.

First, interaction analysis methodology is fundamentally ethnographic in that it focuses on situated actions carried out as they are observable in naturally occurring situations. Specific requirements govern this so-called 'naturalistic' research approach (Mondada, 2016), and they concern the roles of space and time, the nature of the data collected, and how the researcher is positioned in relation to the members of the community being observed. Adopting an ethnographic approach to work and learning means going out into the field, visiting workplaces, vocational schools and training institutions, and being present on site. Research is not carried out at a distance from the practices being investigated. On the contrary, it requires direct, close contact with the people and places being observed. Adopting an ethnographic approach means spending time with the participants involved in these fields—sufficient time to develop a deep understanding of the practices at stake. From there, the ethnographic perspective requires more than an ability to look at or observe. As Winkin (2001) mentioned in a particularly appropriate manner, it also requires the researcher to "know how to be with": a willingness to meet others and develop relationships with them. Consequently, interaction analysis does not begin with making recordings, nor can it be simplified to the systematic scrutiny of those recordings. It is part of a more comprehensive procedure involving the progressive construction of relationships within a field of observation and with the individuals who engage in everyday practices within that field.

Interaction analysis's perspective differs from that of general ethnography in that it requires the collection of audio/video data (Grosjean & Matte, 2021; Heath et al., 2010). Informing the analysis of verbal and non-verbal face-to-face interactions using audio/video recordings of work or training situations has several advantages. Firstly, these recordings provide a broad yet fine-grained description of observable behaviours, as they occur in time and space (Mondada, 2016). Audio/video recordings capture not only the content of verbal exchanges but also their prosodic properties and the non-verbal parts of these interactions. The production of non-verbal gestures and actions, bodily and visual orientations in space, the manipulation of material objects, and movements within the environment all become available to the analyst. Secondly, another significant contribution of audio/video is the fine-grained recording of the dynamic and temporally ordered character of the interactions observed. Video recordings show how the actions accomplished are sequentially linked and mutually synchronised. These are necessary elements of the information required for a detailed study of the coordination processes occurring during social encounters. Finally, another advantage of using audio/video data is that it enables the observer's experience to be replayed infinite times. It also makes it possible to share observations with others across different scales of time and space. Nevertheless, video recordings should not be considered immediate and full access to the data on an activity. On the contrary, recordings focus on specific moments in time, they are framed from a specific perspective, and they are contingent on the technical and environmental conditions in which they are produced.

Collecting audio/video recordings is not an end in itself but merely the starting point for an analytical practice that requires the production and use of multimodal transcripts. Based on a long-standing tradition within conversation analysis, the activity of creating transcripts consists of putting a subset of information available on audio/video data into a written form. The objective of creating a transcript is to make data available for a more detailed analysis of the properties of situated interactions. This relates to the content of speech and to the paraverbal and non-verbal dimensions of the activities observed (e.g. pauses, intonation, overlapping talk, gaze, gestures or non-verbal actions). Producing transcripts allows to allows us to take snapshots of an audio/video recording as situated interactions unfold sequentially. Transcripts make a range of dynamic processes available for analysis that would otherwise remain difficult to apprehend, reflect on and share. In this way, researchers can access the details of the interaction, focus on various properties of verbal and non-verbal behaviour, and share their observations with others. The practice of writing transcripts is not neutral, arbitrary or objective. As many authors have pointed out (Ochs, 1979; Ten Have, 2007), the process appears to be theoretically oriented insofar as it simultaneously selects, organises, interprets and categorises the different properties of interaction. Transcripts do not claim to exhaustively describe all the properties observable in a video recording, but they do so for a subset of them that present a form of relevance consistent with the analytical issues addressed. As such, multimodal transcripts do not reflect the actions recorded in a direct and transparent way. Instead, they should be seen as means through which *some* of the properties of observed interactions are made available for analysis. Video data and multimodal transcripts are two complementary tools that can be used simultaneously to carry out video-based interaction analysis.

19.2.2 *Collective Analysis of Interactional Data*

19.2.2.1 Data Sessions as a Research Method

Collective forms of video-based interaction analysis began development in the early 1990s, particularly in the field of ethnomethodological conversational analysis (Harris et al., 2012). Known as *data sessions*, these methods have gradually become objects of investigation in their own right, and the different ways in which they have been performed have, in turn, been studied as a product of a professional practice specific to a particular scientific community.

Doing a data session can be defined as a situated, collective analytical exploration of audio/video data focusing on interactional processes and collected by researchers in various original, uncontrived institutional contexts. This analytical exploration covers both the recorded video data and its transcript, often paying particular attention to the verbal components of the interaction. However, data sessions tend to systematically resituate talk-in-interaction into the practical context of the actions in which it takes place, in constant relation to the other multimodal resources used and embodied by the participants (Tutt & Hindmarsh, 2011).

19.2.2.2 Objectives and Principles of Collective Data Sessions

Collective analysis of audio/video data provides a means for researchers to confront their analytical insights with those of other researchers at an early stage of the enquiry process. This is part of a process that Durkheim (2001 [1912]) referred to as "collective effervescence", but it is an effervescence that is itself orderly. Indeed, even if the community of researchers engaged in conversational analysis has not explicitly normalised the use of *data sessions*, there are, in practice, recurrent organisational principles. In their typical form, data sessions unfold as follows. Each member of the group present during a data session can contribute analytical comments about the observed data. The observation generally progresses sequentially. In stage one of the process, the researcher who collected the data briefly introduces and presents them. The group then views the audio/video recordings together, usually several times. In stage two, the group members individually and silently explore the video's transcript for approximately 10 min. At the end of this individual analysis, in stage three, a structured round of discussion takes place in which each member of the group is invited to make one or more analytical comments about the sequence observed. Finally, the session ends with a collective discussion.

As an interactional activity, a data session is made up of tangible actions, recognisable by the members of the group and the process itself can be the subject of an interactional analysis. Recent work in the field of conversational analysis has tried to methodically describe some of the actions that constitute the collective analysis of interaction. One recurrent practical problem faced by the participants in data analysis sessions is how to share an observation and make it "noticeable" to the other participants (Harris et al., 2012). Regarding this problem of sharing noticeable analytical observations, Tutt and Hindmarsh (2011) emphasised that the actions of noticing were not limited to verbal utterances but could become physical actions—reenactments—embodied by members of the group, particularly through gestures, gaze and bodily movements.

Another feature characterising the collective analysis of video data is participants' preference for non-normative analytical comments. Consistent with the *analytical mentality* dominant within the paradigm of conversational analysis, primacy is given to an 'unmotivated' description of the data (Psathas, 1995), which avoids *taking sides* (Sacks, 1984, p. 27) or producing moral judgements about the practices observed (Antaki et al., 2008).

Highlighting these practices makes it clear that collective data analysis processes are marked by institutional and epistemic expectations. As such, they also act as social spaces for the individuals taking part. Bushnell (2012) described how participants in data sessions involving students and senior researchers in Japan used a specialised vocabulary during their analytic activities. His work showed that participants' tendency to use terminology specific to the field of conversational analysis allowed them to accept and attribute different categories of participation. More generally, an analysis group is constituted as a community of practice in reference to anthropological approaches to learning (Lave & Wenger, 1991).

Following on from this observation, it appears that, depending on the contexts in which they are carried out, collective data analysis sessions also constitute 'pedagogical institutions' in the sense that they allow for a guided, shared experience between participants with different levels of expertise in the analysis process itself. This property was particularly emphasised by Harris et al. (2012) and, more recently, by Stevanovic and Weiste (2017), whose work showed that data sessions often involved pedagogical practices based on recognisable actions rather than on predetermined social roles. As such, they could be seen as alternative ways of teaching and learning research methods, ways that differ from an explicit, formal transmission of knowledge but that take shape through observation and participation in communities of (analytic) practices.

19.2.3 Interaction Analysis as a Training Method

19.2.3.1 Empirical Exploration of Interaction Analysis in Training Settings

In recent years, collective video-based interaction analysis has found numerous extensions and applications in the field of vocational and professional education and training. It has been used outside the field of academic research and applied as a training method. In other words, it has been assumed that the ability to perform analytic moves, based on a methodical observation of video data, is no longer a privilege restricted to researchers but can also be performed by practitioners as means of expanding the possibilities for learning experiences and professional development related to their work.

For instance, Trébert and Durand (Durand & Trébert, 2018; Trébert & Durand, 2019) applied the principles of collective data sessions to develop mentoring skills among trainers in the field of early childhood education. Working with a small group of educators in charge of guiding students, they used video-based methods to encourage trainers to reflect on their tutoring role when interacting with students in the workplace. In the same context of early childhood education, Zogmal and Durand (2020) organised collective data sessions with professionals to enhance their sense of belonging to a group and to share their experiences within that group. Filliettaz and Zogmal (2021) also used data session methodology to teach educators how to implement a programme for fostering early language acquisition in childcare facilities.

In a different empirical context, that of health, Nguyen recently explored how to implement collective data analysis methods in initial and continuing vocational education and training. Data sessions were used to develop two sorts of competencies during the initial vocational training of student nurses: a clinical competence in the field of therapeutic practice in psychiatry and a methodological competence in the field of observing interactional work with patients (Nguyen et al., 2020). Similar approaches have been applied to continuing education in an institutional setting, with the objective of training qualified nurses in the practice of self-disclosure in psychiatry (Nguyen et al., 2021). During supervision sessions, video data were analysed by groups of health professionals so as to develop interactional competencies for exchanging with patients suffering from schizophrenia.

19.2.3.2 Methodological Principles of Interaction Analysis in Training

Beyond their specificities and the diverse contexts in which they have been implemented, these experiences shared a set of methodological principles related to their epistemic, procedural and analytic aspects.

First, there are the *epistemic conditions* that characterise training methods based on the principles of interaction analysis. It is assumed that the concepts and analytic procedures associated with video-based interaction analysis can be taught, learnt and appropriated by professionals in training sessions. These concepts and procedures constitute an epistemic domain that can be shared between researchers and practitioners. Researchers usually take on mediational training roles towards practitioners, and they guide participants towards discovering a specific way of looking at their actions and interactions at work. However, the sorts of relationships taken on by researchers and practitioners are not firmly asymmetrical and are not conceptualised in a top-down manner. Professionals are also recognised as knowledgeable participants with an epistemic authority based on their occupational expertise. From there, a collaborative-type relationship emerges between the researchers, who share their methodological expertise, and the professionals, who encounter such methods during training sessions.

This collaborative process of mutually sharing expertise unfolds in an organised way and is associated with specific *procedural conditions*. To start with, a data analysis group is set up under the guidance of one or more researchers and composed of a limited number of practitioners willing to participate in training. In a preliminary phase of the training, typical practical problems inherent to specific occupations and related to face-to-face interactions are identified and discussed within the group. Audio/video recordings of a range of typical interactions identified as problematic are collected to provide empirical data and evidence of the problems initially discussed. These recordings are then made available to the group's participants, who can screen them and identify specific sequences that they wish to analyse in detail. In the next phase of training, participants transcribe the selected video clips individually and share them with the group in a collective analysis data session.

Collective data sessions in training programmes unfold sequentially and follow specific *analytic principles*. When analysing audio/video data collectively in training sessions, distinct participant roles are explicitly assigned, such as the *presenter* who has selected the excerpt and prepared the transcript, the *observer* from the analysis group and the other members of the group, who also contribute analytic input. As with the typical format of data sessions in conversational analysis, the analytic procedure begins with a brief introduction by the presenter, who contextualises the selected sequence and the practical problem associated with it. The group then watches the video recording for the first time before letting the presenter share their preliminary observations. The group then takes the time to explore the transcript and watches the video several times before collectively sharing their views on specific moments in the video. At the end of this co-analysis, the floor is given to the observer and the presenter, who summarise the salient outcomes of the collective analysis and comment on the conditions in which it took place. Different sorts of analytic moves can be accomplished during this organised procedure's different steps. Participants can produce descriptive accounts of the behaviours seen in the video and interpret what those behaviours mean to them. In some instances, it is not rare for groups to make judgements about the actions seen (Lussi Borer & Ria, 2015). The role of researchers often consists in guiding participants towards an interactional perspective on the observed data and orienting their attention towards interdependencies and the sequential connections between observable actions. By looking through the lens of this analytic procedure, participants learn to identify the ingredients of their interactional competencies and reconsider the practical problems they encounter at work as tangible coordination issues.

19.3 Implementing Video-Based Interaction Analysis in a Continuing Education Programme for Early Childhood Educators

After presenting the conceptual ingredients of video-based interaction analysis and its application in the field of vocational education and training, this section of the chapter provides an illustration of how such principles and methods can be implemented in empirical contexts and how they can contribute to highlighting and fostering adult learning as it occurs in work environments.

As our example, we will refer to an ongoing research programme being conducted in the Swiss canton of Geneva[1] and focusing on the interactional competencies required and enacted by early childhood educators when they meet with parents in everyday social encounters. The following subsections present the project's context, objectives and general research design before describing in more detail how collective forms of interaction analysis have been implemented in a continuing education programme for qualified educators. We provide a sample of the data resulting from this training programme and analyse how the participants in collective data sessions used video-based material to identify and comment on the sorts of interactional competencies they mobilise when interacting with parents in naturally occurring work situations.

19.3.1 Objectives and Empirical Research Design

Our example research programme was developed within the framework of our broader interest in the work of early childhood educators and the sorts of professional competencies they require to satisfy the numerous, complex institutional demands associated with this field of education (Filliettaz et al., 2015; Filliettaz & Zogmal, 2020). After several years of using an interactional perspective to investigate the educational activities of students studying to become early childhood educators during their internships, we chose their encounters with parents as a new subject for investigation.

This new research programme focuses on encounters between parents, early childhood educators and children during pick-ups, drop-offs and yearly parents–educators meetings at early childhood education centres. At an institutional level, the importance of creating a *partnership* with parents is increasingly a stated objective. Indeed, from a policy perspective, building that partnership is now recognised as an integral part of the work of early childhood education, established in an increasing number of rules, norms and expectations for workers in the field (OECD, 2006). The

[1] The research programme is funded by the Swiss National Science Foundation (SNF), under grant reference number 100019_182160. The programme is conducted in partnership with Prof. Stephen Billett and Prof. Beverley Flückiger from Griffith University.

concept of partnership refers to collaborative relationships, including shared decision-making, between the roles of parents and educators. However, when looking more closely at the policy documents, they propose assigning educators with a prominent role in establishing relationships with parents: professionals are to “support parenting” (VDG, 2016), “develop” the partnership and “identify” parents’ needs (PEC, 2015). Parents are, therefore, positioned symmetrically and involved in a process of “co-education”—a role that has been conferred on them by educational institutions and framed under the responsibility of others.

Establishing a partnership with parents is not easy and can be associated with numerous challenges. The conditions within which such relationships evolve are often complex and characterised by multiple practical contingencies. Encounters between parents and educators occur daily, particularly at morning drop-offs, when parents bring their children to the education centre, and at pick-ups in the afternoon. These encounters are often very brief, although they may involve multiple participants and activities that are not necessarily compatible with the *co-education* project. These encounters may also materialise in formal meetings, but these tend to be rare and usually only occur yearly. Relations with parents are not always necessarily smooth and collaborative. As evidenced in the literature, they may also include power relations and conflictual educational norms or a sense of legitimacy (Bouve, 1999; Cheatham & Ostrosky, 2009). When families come from cultural backgrounds different from the dominant local one, epistemic asymmetries and cultural misunderstandings may occur (Nunez Moscoso & Ogay, 2016; Scalambrin & Ogay, 2014).

Very little is known about how relations between educators and parents are enacted in practice or how a partnership might be created in observable social encounters and interactions. Our programme’s objective was to address these issues by investigating two main avenues of research. The first objective was to identify and recognise the sorts of interactional competencies mobilised by early childhood educators when they encounter parents. The research questions developed here were the following: What are the typical interactional patterns and characteristics of parent–educator encounters in early childhood education? What kinds of challenges do educators face when interacting with parents? What interactional competencies are required and mobilised to respond to these challenges? The research programme’s second objective was to assist early childhood educators in the development of their interactional competencies for encountering parents. The research questions developed here can be formulated as follows: How can interactional competencies be supported and developed through continuing education and training programmes? What can video-based interaction analysis contribute to such training?

To answer these questions, we used video-based interaction analysis in an empirical research design comprising two consecutive phases. The first phase consisted of a video-ethnographic inquiry focusing on the encounters between parents, early childhood educators and children in early childhood education centres

Table 19.1 Audio/video data available, in hours and minutes

	Video-ethnography phase			Training phase			
	Drop-offs	Pick-ups	Meetings	Input	Selection	Co-analysis	Feed-back
Institution A	38:59	37:03	5:58	3:15	3:54	7:54	5:56
Institution B	41:20	61:09	2:42	3:59	3:33	5:52	5:30
Total (hours)	**80:19**	**98:12**	**8:40**	**7:14**	**7:27**	**13:46**	**11:26**

in Geneva, in French-speaking Switzerland. These interactions were videotaped in standard work situations over two consecutive weeks. Our observations focused on three typical interactions: (a) morning drop-offs when parents bring their children to the educational centre, (b) afternoon pick-ups when parents collect their children again, and (c) formal yearly meetings with parents, when educators provide feedback on children's development and progress. The second phase of the study used an intervention-training design with video-based interaction analysis to assist qualified childhood education professionals reflect on their interactional skills and competencies. Data from the video-ethnographic phase served as training material for small groups of volunteer educators, with the aim of expanding their interactional competencies in relation to parent and family interactions. Educators were introduced to interaction analysis and the methodological principles associated with its *analytical mentality*. They were also trained to select and transcribe video data from their work and to perform a collective analysis of these data with the group. Finally, they prepared and delivered feedback about their training to a larger group of colleagues working in the same institutions.

This empirical research took place in two childcare facilities in the canton of Geneva between Spring 2018 and Spring 2020. Data consisted of video recordings of typical multimodal interactions that took place during the two consecutive phases of the project.

As Table 19.1 indicates, a total of 187 h of video were recorded for the video-ethnographic phase of the project, of which 80 h focused on drop-offs, 98 h looked at pick-ups and 8 h showed yearly meetings. The project's entire training phase was also video-recorded, with almost 40 h of data from the different steps of training design, including 7 h of content-based training on interaction analysis, 7 h of video sequence selection by participants, 13 h of collective video-data analysis by the groups and 11 h about the preparation of dissemination activities within larger institutions. Video recordings were organised in a database, transcribed and coded using Transana Multi-User qualitative analysis software.

Fig. 19.1 Setting for the data analysis session during the training phase

19.3.2 Illustration and Case Study

To illustrate how the methodological principles of video-based interaction analysis can be implemented in training sessions, we now turn to a small sample of data related to the training phase of the project mentioned above. We present this case study as a way of understanding how a group of professionals can experience the methodology of interaction analysis and accomplish specific analytic moves when scrutinising the video data documenting multimodal interactions between educators, parents and their children. We also wish to underline the role of the trainers and researchers involved in the collective analysis of this video data and reflect on the sorts of learning that can arise from such collaborative data analysis experiences.

This sample of empirical data comes from the training programme's co-analysis phase, a sequence of training during which the participants were conducting data analysis sessions featuring typical work situations in which they encountered parents. The data sample presented below comes from the second data session performed by the group. Alison (ALI), one of the educators enrolled in the training programme, selected and transcribed a video sequence related to a pick-up interaction, and she shared it with a group of five colleagues (KAR, LOR, SAR, MEL and DAN) under the guidance of a pair of trainers who were also researchers in the context of this programme (CH1 and CH2) (Fig. 19.1).

The video recording analysed by the group showed a situation in which Alison was interacting with a small boy named Pedro and his mother. Pedro's mother had come to pick him up at the end of the day, but Pedro did not want to leave the centre. He kept running away from his mother and continued playing with children of his

Fig. 19.2 Video sequence of Pedro's pick-up by his mother

age group. Instead of lasting just a couple of minutes, this pick-up encounter took almost 15 min, and it put Alison in an uncomfortable position because the mother refused the educator's help and expected Pedro to come to her of his own free will (Fig. 19.2).

During the data session, the group made several observations about the video clip selected by Alison. Alison emphasised the unusually long duration of this pick-up interaction and the mother's personal characteristics—a lady who often takes too much time when simply picking up her son. The following excerpt comes from the middle of the analysis session—a point when the group was observing how participants in the video were addressing each other. On several occasions during Pedro's pick-up interaction, they addressed each other indirectly. Alison talked to the child to deliver information to his mother. And the mother addressed other children to capture Pedro's attention. The excerpt below reveals the sorts of deliberations made by the groups attending data sessions when expressing their observations:

Excerpt (1): Excerpt of the data session analysing Pedro's pick-up interaction.[2]

[2]Transcript conventions are presented in the Appendix at this end of the chapter.

1. KAR but from my point of view it's easier to talk to the child than to the parent\
2. DAN yeah that's right\
3. KAR to say 'Yeah, Pedro, I think it's time to go now, mummy's waiting for you\'
4. DAN yeah well here that's not working at all actually\
5. ALI but in the end it always goes through the child . the parents talk through the child/ we talk through the child and what impact does always being uh always being in between have on the child/ *((hand gestures))*
6. KAR uh-huh
7. CH1 but it's funny because who do you speak to when you catch the child/ *((winds the film back to a few moments before))*
8. ALI well I speak to Pedro so I could've-
9. CH1 *((nods))* $es but is it just to Pedro/ *((laughter))*
10. ALI well... no... indir- so directly it's to Pedro and indirectly I make the mother understand that she should come over and get him\
11. DAN yeah\
12. KAR yeah\
13. ALI I think so\
14. CH1 exactly\ so to avoid compromising parental authority you can also use the strategy of speaking to the child\

At the beginning of the excerpt transcribed above, Karina (KAR), one of Alison's colleagues attending the data session, confessed that she found it difficult to address parents directly, particularly when educational problems might emerge from the situation: "But from my point of view, it's easier to talk to the child than to the parent" (l. 1). She gave an example of how to address parents through their children and how this pattern of interaction could be enacted in Pedro's very specific context: "To say, 'Yeah, Pedro, I think it's time to go now. Mummy's waiting for you'" (l. 3). From there, a discussion emerged between the analysis session participants in which they shared their views about the efficacy and legitimacy of such a strategy. Daniela (DAN) observed that this form of indirect communication was inefficient in that context ("Yeah, well, here that's not working at all, actually" l. 4), and Alison wondered how this indirect form of communication affected the children (l. 5). At this point, one of the trainers, who was in control of the computer, brought Alison's attention to the characteristics of her way of talking to Pedro and his mother. By rewinding the video sequence a little bit and replaying it, she asked Alison whom she was really addressing (l. 7). Alison came to the realisation that although she seemed to be speaking to Pedro, she was indirectly addressing the mother and asking her to come over and get her child (l. 10). The trainer had managed to make Alison understand that she had also been deploying specific interactional resources to avoid explicitly compromising the mother's parental authority (l. 14).

This short sequence of data analysis reveals different analytic moves. First, it is noticeable that educators seemed to have the capacity to orient their attention towards the fine-grained characteristics of the sorts of interactions they had with parents and children during pick-ups. Not only did they observe that they talked or spoke to each other, but they also noted the different strategies used in doing so: just as parents sometimes talked to their children through other participants, educators chose to address the children as a way of intentionally communicating with parents. When commenting on these ongoing interactions, educators also used specific, explicit conceptual constructs, such as the distinction between direct and indirect forms of communication. These categories of communication had been introduced to them earlier in the training programme to serve as analytical tools with which to describe noticeable characteristics in the empirical material that the group was scrutinising.

It is also important to note that the programme trainers were not fully detached from these analytic moves and conceptual explorations. In the present case, it was CH1 who identified an interactional phenomenon worthy of observation and guided Alison towards the understanding that she had addressed Pedro's mother indirectly through him: "Well... no... indir... so directly, it's to Pedro, and indirectly, I make the mother understand that she should come over and get him" (l. 10). Thanks to the trainer's scaffolding questions (l. 7, l. 10) and a replay of the video excerpt, this observation was made and shared collectively within the data session.

Finally, it is also interesting to observe that the fine-grained mechanisms associated with interaction analysis in the present context are not categorised exclusively as mere verbal behaviour but are connected to broader social and professional norms. When referring to the situation under analysis, Alison questioned the potential impact of such indirect communication strategies on children ("But in the end, it always goes through the child. The parents talk through the child. We talk through the child, and what impact does always being, uh, always being in between have on the child?" l. 5). Daniela, for her part, seemed critical of this strategy of indirect communication (l. 4). Finally, the trainer established a connection between this type of indirect interactional pattern and childhood educators' preference for avoiding confrontations with parents over challenging educational situations (l. 14). In other words, analytic moves can be seen as potential connections between observable situated actions and broader professional dilemmas or controversies. It is through these sorts of empirical observations that such dilemmas and controversies can be discussed and negotiated within larger groups of participants who share the same interests and profession.

19.4 Conclusion

Based on the excerpt of empirical data analysed above, what have we learnt about work, and what have we learnt about adult learning in practical terms?

From what we can see in the small sample of data analysed here, interactional competencies play an essential role in contemporary work environments and service-oriented occupations. Early childhood educators are constantly interacting with each other, children, parents or other persons. A large part of their professional skills is mediated by their capacity to engage in interaction processes collaboratively and to coordinate with others efficiently and legitimately. These capacities are neither natural nor self-evident. In many circumstances, they are not even taught explicitly in their formal training. Instead, they are acquired through practice and are learnt in working environments and settings.

To assist qualified professionals in the development of their interactional competencies in the workplace, video-based interaction analysis could be a promising resource. As evidenced in the empirical section of this chapter, collective data sessions mediated by trainers can lead to a variety of learning outcomes connected with various dimensions of professional practices: praxis, knowledge creation and identity formation. At the praxeological level, engaging in collective forms of interactions analysis at work helps professionals to interpret the different sorts of actions they engage in. Through a descriptive account of the data they discussed, our data session group shared views about their intentions and motives and assigned meaning to something that is often difficult to interpret in work circumstances, namely, what participants *mean* to do or say when they behave how they do. From an epistemic perspective, collective data sessions also contribute to establishing, sharing and disseminating specific sorts of knowledge associated with professional practices. In our example, trainers introduced concepts associated with direct and indirect forms of communication, and participants subsequently reused and recycled those concepts during their analytic experiences. These epistemic categories can be introduced by researchers or other professionals. However, in most cases, they are collectively elaborated within a group during their successive data sessions (Garcia & Filliettaz, 2020; Garcia, 2020). Lastly, video-based interaction analysis seems to provide productive outcomes for participants in terms of their professional development and identity formation. As indicated in the brief case study analysed in this chapter, observing how interactions unfold can lead to participants discussing broader professional dilemmas and the social norms shared within their communities of practice. This may contribute to establishing or renegotiating those norms within groups and communities.

From our own practical experience and based on several other experiments mentioned in this chapter, it seems that the concepts and methods that define the principles of video-based interaction analysis can be applied fruitfully to the fields of vocational and continuing education. Not only is it a method through which researchers can investigate how interactions unfold in work situations and settings but it can also be used by professionals as a way to reflect on their own work practices. From what we have observed in the small sample of data extracted from our training sessions in the field of early childhood education, professionals seem to have the capacity not only to enact and mobilise interactional competencies but also to recognise and identify these competencies when describing and commenting on what they did in video recordings of their work. The opportunities for reflection provided by video-based interaction analysis can be seen as a promising avenue for the recognition and development of adult learning at work.

Appendix

Transcription Conventions

/ \	rising and falling intonation
(.)	micro-pause
(2.1)	pauses in seconds
XXX	inaudible segment
exTRA	accentuated segment
((pointing))	non-verbal behaviour

References

Antaki, C., Biazzi, M., Nissen, A., & Wagner, J. (2008). Accounting for moral judgments in academic talk: The case of a conversation analysis data session. *Text & Talk, 28*(1), 1–30.

Bouve, C. (1999). Confrontation des pratiques éducatives entre familles et crèches collectives. *Recherches et prévisions, 57*(1), 45–58.

Bushnell, C. (2012). Talking the talk: The interactional construction of community and identity at conversation analytic data sessions in Japan. *Human Studies, 35*(4), 583–605.

Cheatham, G. A., & Ostrosky, M. M. (2009). Listening for details of talk: Early childhood parent-teacher conference communication facilitators. *Young Exceptional Children, 13*(1), 36–49.

Durand, I., & Trébert, D. (2018). Quand l'outil des chercheurs devient celui des praticiens: Un Groupe d'Analyse des Interactions en formation de formateurs. In I. Vinatier, L. Filliettaz, & M. Laforest (Eds.), *L'analyse des interactions dans le travail: outil de formation professionnelle et de recherche* (pp. 51–76). Éditions Raison et Passions.

Durkheim, É. (2001 [1912]). *The elementary forms of religious life* (C. Cosman, Trans.). Oxford University Press.

Filliettaz, L. (2014a). Learning through interactional participatory configurations: Contributions from video analysis. In A. Rausch, C. Harteis, & J. Seifried (Eds.), *Discourses on professional learning: On the boundary between learning and working* (pp. 317–339). Springer.

Filliettaz, L. (2014b). Understanding learning for work: Interaction and discourse analysis. In S. Billett, H. Gruber, & C. Harteis (Eds.), *International handbook of research in professional and practice-based learning*. Springer.

Filliettaz, L., & Zogmal, M. (Eds.). (2020). *Mobiliser et développer des compétences interactionnelles en situation de travail éducatif*. Editions Octarès.

Filliettaz, L., & Zogmal, M. (2021). « Vraiment c'était un moment très riche » : Apprendre à analyser des interactions langagières en contexte de travail et de formation. *Travaux Neuchâtelois de Linguistique, 74*(1), 15–30.

Filliettaz, L., de Saint-Georges, I., & Duc, B. (2010). Skiing, cheese fondue and Swiss watches: Analogical discourse in vocational training interactions. *Vocations and Learning, 3*(2), 117–140.

Filliettaz, L., Durand, I., & Trébert, D. (2015). Learning through interactional participatory configurations. In L. Filliettaz & S. Billett (Eds.), *Francophone perspectives of learning through work: Conceptions, traditions and practices* (pp. 279–301). Springer.

Garcia, S. (2020). Penser les continuités entre recherche et formation professionnelle dans les métiers socio-éducatifs: le cas des compétences interactionnelles des éducatrices de l'enfance dans les relations avec les parents. *Recherche & Formation, 94*(2), 33–51.

Garcia, S., & Filliettaz, L. (2020). Compétences interactionnelles et relations des éducateurs de l'enfance avec les parents: la formation comme ressource pour la recherche. *Phronesis, 9*(2), 123–138.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Polity Press.

Goffman, E. (1956). *The presentation of self in everyday life*. University of Edinburgh, Social Sciences Research Centre.

Goffman, E. (1964). The neglected situation. *American Anthropologist, 66*, 133–136.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522.

Grosjean, S., & Matte, F. (Eds.). (2021). *Organizational video-ethnography revisited: Making visible material, embodied and sensory practices*. Springer Nature.

Harris, J., Theobald, M., Danby, S. J., Reynolds, E., & Rintel, S. (2012). "What's going on here?": The pedagogy of a data analysis session. In A. Lee & S. J. Danby (Eds.), *Reshaping doctoral education: International approaches and pedagogies* (pp. 83–96). Routledge.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage.

Johansson, E., Lindwall, O., & Rystedt, H. (2017). Experiences, appearances, and interprofessional training: The instructional use of video in post-simulation debriefings. *International Journal of Computer-Supported Collaborative Learning, 12*(1), 91–112.

Kilbrink, N., Asplund, S. B., & Asghari, H. (2021). Introducing the object of learning in interaction: Vocational teaching and learning in a plumbing workshop session. *Journal of Vocational Education & Training*, 1–26. https://www.tandfonline.com/doi/full/10.1080/13636820.2020.1850512

Koskela, I., & Palukka, H. (2011). Trainer interventions as instructional strategies in air traffic control training. *Journal of Workplace Learning, 23*(5), 293–314.

Kress, G., et al. (2001). *Multimodal teaching and learning. The rhetorics of the science classroom*. Continuum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lussi Borer, V., & Ria, L. (2015). Concevoir et expérimenter un laboratoire d'analyse vidéo de l'activité enseignante au sein d'un établissement scolaire. In V. Lussi Borer, M. Durand, & F. Yvon (Eds.), *Analyse du travail et formation dans les métiers de l'éducation* (pp. 219–238). De Boeck.

Melander, H. (2017). Becoming a "Good Nurse": Social norms of conduct and the management of interpersonal relations. In S. Pekarek Doehler et al. (Eds.), *Interactional competences in institutional settings* (pp. 171–196). Palgrave Macmillan.

Mondada, L. (2006). La compétence comme dimension située et contingente, localement évaluée par les participants. *Bulletin suisse de linguistique appliquée, 84*, 83–119.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of SocioLinguistics, 20*(3), 336–366.

Nguyen, H. T. (2017). Toward a conversation analytic framework for tracking interactional competence development from school to work. In S. Pekarek Doehler et al. (Eds.), *Interactional competences in institutional settings* (pp. 197–225). Palgrave Macmillan.

Nguyen, A., Frobert, L., & Favrod, J. (2020). L'analyse interactionnelle comme modalité pédagogique: l'exemple d'un dispositif pour la formation des infirmiers à la relation thérapeutique en psychiatrie. *Phronesis, 9*(2), 93–107.

Nguyen, A., Frobert, L., Pellet, J., & Favrod, J. (2021). La supervision à la compétence du développement de soi professionnel en psychiatrie: perspective interactionnelle sur un dispositif de formation et d'une recherche. In V. Rémery, F. Chrétien, & C. Chatigny (Eds.), *Apprentissage et transmission de l'expérience en situation de travail: dialogue entre formation d'adultes et ergonomie*. Presses Universitaires de Rouen et du Havre.

Nunez Moscoso, J., & Ogay, T. (2016). L'entretien enseignant-famille à l'école maternelle: quelles dynamiques organisationnelles, quels savoirs mobilisés? *Revue suisse des sciences de l'éducation, 42*(1), 147–181.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43–72). Academic.

OECD. (2006). *Starting strong II. Early childhood education and care*. Organisation for Economic Cooperation Development.

PEC. (2015). *Plan d'études cadre 'Education de l'enfance ES'*. SEFRI.

Pekarek Doehler, S., et al. (Eds.). (2017). *Interactional competences in institutional settings: From school to the workplace*. Palgrave Macmillan.

Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction* (Vol. 35). Sage.

Sacks, H. (1984). Notes on methodology. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 21–27). Cambridge University Press.

Sacks, H., Schegloff, E., & Jefferson, G. (1978). A simplest systematics of the organization of turn taking for conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 7–55). Academic.

Scalambrin, L., & Ogay, T. (2014). « Votre enfant dans ma classe ». Quel partenariat parents-enseignante à l'issue du premier entretien? *Education et sociétés, 34*(2), 23–38.

Schegloff, E. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.

Stevanovic, M., & Weiste, E. (2017). Conversation-analytic data session as a pedagogical institution. *Learning, Culture and Social Interaction, 15*, 1–17.

Stokoe, E. (2014). The Conversation Analytic Role-play Method (CARM): A method for training communication skills as an alternative to simulated role-play. *Research on Language and Social Interaction, 47*(3), 255–265.

Ten Have, P. (2007). *Doing conversation analysis*. Sage.

Trébert, D., & Durand, I. (2019). L'analyse multimodale des interactions comme ressource pour la formation: Le cas du tutorat dans le champ de l'éducation de la petite enfance. In N. Blanc & V. Rivière (Eds.), *Observer la multimodalité en situations éducatives: circulation entre recherche et formation* (pp. 173–198). Éditions ENS.

Tutt, D., & Hindmarsh, J. (2011). Reenactments at work: Demonstrating conduct in data sessions. *Research on Language and Social Interaction, 44*(3), 211–236.

VDG (Ville de Genève). (2016). *Les droits de l'enfant en ville de Genève*. Département de la cohésion sociale et de la solidarité.

Winkin, Y. (2001). *Anthropologie de la communication*. Editions du Seuil.

Young, R. F., & Miller, E. R. (2004). Learning as changing participation: Discourse roles in ESL writing conferences. *The Modern Language Journal, 88*(4), 519–535.

Zogmal, M., & Durand, I. (2020). De l'observation des enfants à l'analyse interactionnelle: contributions de la recherche à la formation continue des éducateurs et éducatrices de l'enfance. *Phronesis, 9*(2), 108–122.

Laurent Filliettaz (PhD) is a Professor of Adult and Vocational Education at the University of Geneva (Switzerland). He is a specialist in discourse analysis and multimodal approaches to interaction. His research is strongly interdisciplinary and combines contributions from applied linguistics and the educational sciences. Over the years, he has conducted several empirical research programmes on topics such as language use in professional contexts, apprenticeship, and interactional competencies in vocational education and training.

Stéphanie Garcia (MA) is a doctoral student at the University of Geneva and a member of the Interaction and Training research group at the Faculty of Psychology and Educational Sciences. She is conducting research on interactional competencies in the field of early childhood education and, more specifically, on partnership relations between educators and parents.

Marianne Zogmal (PhD) is a Senior Researcher at the University of Geneva and the University of Applied Sciences in Social Work in Lausanne. She has completed a doctoral dissertation on the vocational education and training of early childhood educators and is currently involved in several research programmes dedicated to interactional competencies in early childhood education, language acquisition, and fostering and mentoring practices in vocational education.

Chapter 20
Q Method: Assessing Subjectivity Through Structured Ranking of Items

Susann Leidig, Hanna Köhler, Carina Caruso, and Michael Goller

Abstract Q is an exploratory method used to elicit discrete patterns of subjectivity (e.g., subjective theories, beliefs). To be more concrete, Q method aims to identify homogeneous latent clusters of viewpoints towards a certain topic from a larger more heterogeneous set of (a priori unknown) different viewpoints held by respondents in the population. For this purpose, study participants are asked to sort qualitative statements concerning a certain topic (e.g., their beliefs toward something) into a grid roughly following a normal distribution indicating whether they agree or disagree with the given propositions. This results in a so-called Q-sort for each participant that then can be fed into a factor analysis that compares the different Q-sorts and groups them by similarity. In the first part of the chapter, Q method is introduced. The second part of the chapter describes a small study focusing on student teachers and their perspectives concerning upcoming long-term school internships.

Keywords Q method · Subjectivity · Internships · Teacher education · Q sort

S. Leidig (✉) · H. Köhler
Institute of Educational Science, Friedrich Schiller University Jena, Jena, Germany
e-mail: susann-julia.leidig@uni-jena.de; hanna.koehler@uni-jena.de

C. Caruso
Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: ccaruso@mail.uni-paderborn.de

M. Goller
Institute of Educational Science, Friedrich Schiller University Jena, Jena, Germany

Institute of Educational Science, Paderborn University, Paderborn, Germany
e-mail: michael.goller@upb.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_20

20.1 Introduction

Q is a comprehensive research method that aims at investigating human subjectivity. In the context of this method, subjectivity is broadly defined as an individual's "point of view on any matter of personal or social importance" (McKeown & Thomas, 2013, p. IX) that is based on assumptions, beliefs, conceptions, opinions, prejudices, or feelings of a person with regard to a given topic (Harteis et al., 2006; Lundberg et al., 2020; Rieber, 2020).[1] In this sense, the subjectivity of an individual encompasses emotionally charged ideas of the nature of the world, which are held to be true or valuable and which provide structure, support, security, and orientation for their holders' thinking and actions. They can be explicit or intuitive, fragmentary and even contradictory, or they can combine to form personalised subjective theories. Such subjective theories contain ideas about how the world works and what actions are appropriate to meet certain goals in the face of particular circumstances (Groeben, 1988). It follows that subjectivity plays a significant role in why and how individuals act since it affects the selection of goals and action plans as well as the perception and interpretation of situations (e.g., Billett, 2006; Reusser & Pauli, 2014).

Although such subjectivity is idiosyncratic in nature, patterns of shared values and beliefs are often formed between groups of individuals (e.g., political opinions, preferences). Knowledge about individual as well as shared subjectivity is of great relevance from a scientific and especially from an educational perspective. On the one hand, subjectivities of educators affect how they plan and organise instructional arrangements or learning environments (Reusser & Pauli, 2014). It is therefore not surprising that subjectivity in the form of beliefs has been identified as a highly relevant component of educators' professional competence (e.g., OECD, 2009). On the other hand, there is a long tradition within educational science to argue that "[a]n important element of good design, instructional or otherwise, is awareness of and being empathetic to the needs, wants, interests, values, and opinions of the intended audience" (Rieber, 2020, p. 2529). After all, such needs, wants, interests, values and opinions are nothing other than subjective. They are relevant as they have the power to affect learning behaviours and eventually competence development (e.g., Elen & Lowyck, 2000; Entwistle & Peterson, 2004; Oosterheert & Vermunt, 2001). A good understanding of learners' subjectivities that explains how and why they engage or disengage with certain learning affordances might therefore help to maximise learning outcomes. Such knowledge enables educators to take relevant aspects of

[1] It should be noted that the concept of subjectivity is discussed from a range of research perspectives that are rooted in different disciplinary discourses (e.g., González Rey, 2019; Luhrmann, 2006; Ortner, 2005; Watts & Stenner, 2012). Within this contribution, the concept of subjectivity is meant to encompass all kinds of states of mind of individuals that affect how they encounter, construe, and engage with both their social and natural world (see for a similar but much more detailed conception: Billett, 2006).

their students' subjectivity into account and consequently to tailor learning environments around learners' needs, wants and so on.

A highly suitable research method directly designed to identify patterns of subjectivity is Q method (hereafter Q). Q is particularly relevant to researchers interested in understanding the diversity of opinions or views of a number of people on a topic that is relevant and important to this group (Rieber, 2020). It is an exploratory method used to elicit discrete patterns of subjectivity—that is, to identify homogeneous latent clusters of viewpoints towards a certain topic from a larger more heterogeneous set of (a priori unknown) different viewpoints held by respondents in the population. Therefore, Q has explicitly been designed to uncover the diversity of views, regardless of whether these views are common in a population or not (Watts & Stenner, 2012; Zabala et al., 2018). In addition, Q aims to elicit the importance of viewpoints in relationship to all identified viewpoints on a particular subject matter (Watts & Stenner, 2012), as different aspects towards the topic at hand have to be evaluated in relation to each other. This distinguishes Q from other empirical methods often used to investigate subjectivity, such as questionnaires, in which each item is typically assessed independently from all other statements.

Q has recently become increasingly important in a range of research domains including political sciences, environmental psychology, marketing, media studies, and gender studies (Müller & Kals, 2004). In educational science, Q has, for example, been used in research on the following: teachers' epistemologies affecting science instruction practices (Barnes et al., 2015), the subjective role of students with disabilities in the implementation of education policy (Salaj & Kiš-Glavaš, 2017), perspectives of teachers concerning their professional development experiences (Cooper et al., 2018), decision-making strategies of school principals when dealing with challenges in the change process at schools (Summak & Kalman, 2020), and medical students' attitude development (Schick et al., 2021). However, although several prominent researchers in the field of Q emphasise the method as a promising tool in educational science, it is still seldom used (Lundberg et al., 2020). This is especially true for research on professional development and professional learning, although it could make important contributions to the field.

This chapter aims at introducing Q as a research method, comprising an inherent data collection procedure, the so-called Q sorting, the Q pattern analysis based on multivariate data-reduction techniques for data analysis and an interpretation based on specific parameters. In the first part of the chapter Q will be explained on a conceptual level. For this purpose, a short historical overview on the development of Q is given before the method itself is explained and discussed in more detail. In the second part of the chapter Q will be illustrated with a small study that focuses on student teachers in the context of their 5-month school internship. The study investigates whether students perceive different aspects of these internships as relevant for their ongoing professional development.

20.2 A Primer on Q

20.2.1 A Brief Historical Note

In 1935 Q was first introduced by William Stephenson in a letter to the journal *Nature*, in which he outlined the general concept of the method. Almost 20 years later, Stephenson presented his ideas fully in the book *The Study of Behaviour: Q-Technique and its Methodology* (Stephenson, 1953; Rieber, 2020). Since then, Q has been applied for decades in various disciplines in order to identify patterns in different facets of subjectivity (Watts & Stenner, 2012; Zabala et al., 2018). Political opinion and attitudes research, research into subjective theories in the clinical-psychological field, media research, market research, environmental psychological research, and gender research can be named as typical examples of areas that have applied Q (Müller & Kals, 2004).

Although Q has been around for decades and a few scholars have used it in their research, it seems that many (educational) researchers are not familiar with the method. Fortunately, at least two well-written introductory resources have been available for some time: (a) the book by McKeown and Thomas (2013) has now been published in its second edition and gives a short but comprehensive overview about Q, and (b) the book by Watts and Stenner (2012) gives an even more comprehensive introduction to Q by focussing on every step of a Q study in much detail. In addition, accessible software packages for data collection as well as data analysis are available (e.g., PQMethod: Schmolck, 2008; qmethod for R: Zabala, 2014; qfactor for Stata: Akhtar-Danesh, 2021; Ken-Q: Banasick, 2019).

20.2.2 The Why and How of Q Method

As foreshadowed, Q is used to identify patterns of subjectivity of individuals, including beliefs, attitudes, opinions, perspectives or subjective theories, concerning a certain topic or subject matter (Kamal et al., 2014; Müller & Kals, 2004; Zabala et al., 2018; Rieber, 2020). In other words, Q aims to identify a priori unknown homogeneous latent clusters of perspectives concerning a specific subject matter from a larger heterogeneous set of different viewpoints held by respondents in the population. It follows that Q can be used to reveal the diversity of views in a larger group of individuals. Due to its methodological approach, this can be done with relatively small sample sizes and irrespective of whether the different views are common in the population or not (Rieber, 2020; Watts & Stenner, 2012; Zabala et al., 2018).

For this purpose, Q presents study participants with a large range of different statements representing the subject matter of interest. The participants are then asked to rate the statements in a very specific form, meaning that all statements must be evaluated in relation to each other. As such, and in contrast to other data collection methods like questionnaires, the importance of a certain statement relative to all other statements is included in any further investigation. The resulting rankings are subsequently used in a factor analysis to identify similarities and differences in participants' views concerning the subject matter. The next section will explain the procedure behind Q in more detail.

20.2.3 Four Steps of Using Q Method

Following Zabala et al. (2018), a Q study can be separated into four distinct but interrelated steps that are discussed below: (a) decisions on the research design and sampling, (b) data collection, (c) data analysis, and (d) interpretation. See also Fig. 20.1 for a summary of the four steps.

1. *First*, the topic that provides the guiding framework for the study needs to be identified. In other words, researchers must narrow down the specific subject matter they are interested in and that they want to scrutinize to identify patterns of subjectivity. Once this is done a central stimulus that will be used to present to the study participants is derived (e.g., "Use the following statements in order to describe your opinion on subject matter A. Sort the items according to those with which you most agree to the ones you most disagree."; see also McKeown & Thomas, 2013, pp. 25–27). The wording of this stimulus is then used to collect a comprehensive list of statements that represent (ideally) the whole range of viewpoints regarding the subject matter that will be later used in data collection (Rieber, 2020). The final list of statements is usually called the Q-set. Researchers are free to consult different resources such as scientific publications, everyday

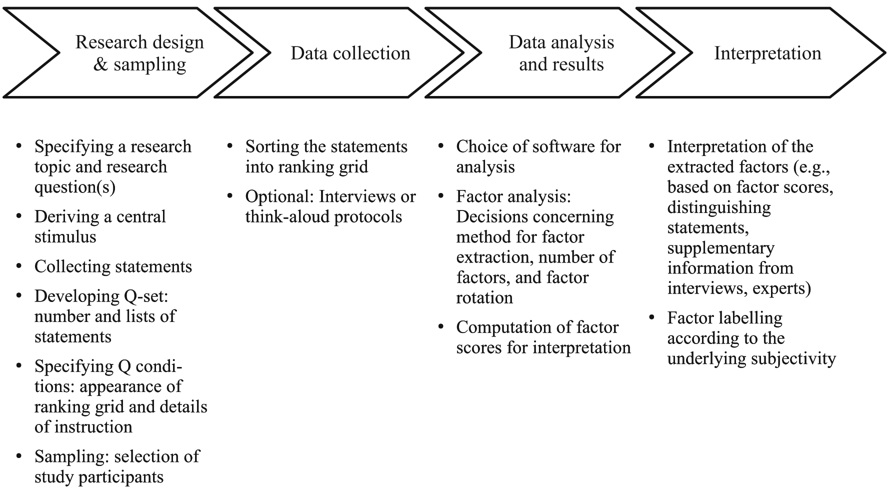

Fig. 20.1 Four steps of using Q method

literature, expert opinions or even interviews with individuals who are part of the target group in order to derive the Q-set. The statements in the Q-set should reflect all possible perspectives on the subject matter such that the phenomenon of interest is covered in its full range, as Q aims to expose the diversity of all opinions regardless of whether an opinion prevails in the population or not. It follows that the sampling strategy aims at selecting participants in such a way that a preferably heterogeneous sample is created that includes a wide spectrum of viewpoints regarding the subject matter (Zabala et al., 2018). However, Q studies can be conducted with relatively few participants. According to Rieber (2020), approximately half as many participants as there are Q statements are required. Other sources speak of a sample of 30–60 participants as sufficient (McKeown & Thomas, 2013; Watts & Stenner, 2012).

2. In the *second* step, participants are invited to react towards the statements included in the Q-set. More specifically, they are presented with the central stimulus and then asked whether they agree or disagree with each single statement by sorting it into a predefined grid (see Fig. 20.2). The possible number of statements per column in this ranking grid roughly follows a normal distribution. As only a limited number of statements can be assigned to the same ranking value, participants have to rank the items in relation to each other and as such are forced to prioritize their agreement and disagreement. Items placed in the same column later obtain the same agreement score, resulting in a subjective statement ranking for each participant. These so-called Q-sorts reflect the perspectives concerning the subject matter of every single respondent. Participants are usually asked to sort the statements into the grid by starting with items with which they subjectively agree. After that, they sort items with which they do not agree, followed by the remaining statements (Rieber, 2020). It is possible to combine the sorting procedure to obtain the individual Q-sorts with qualitative interviews

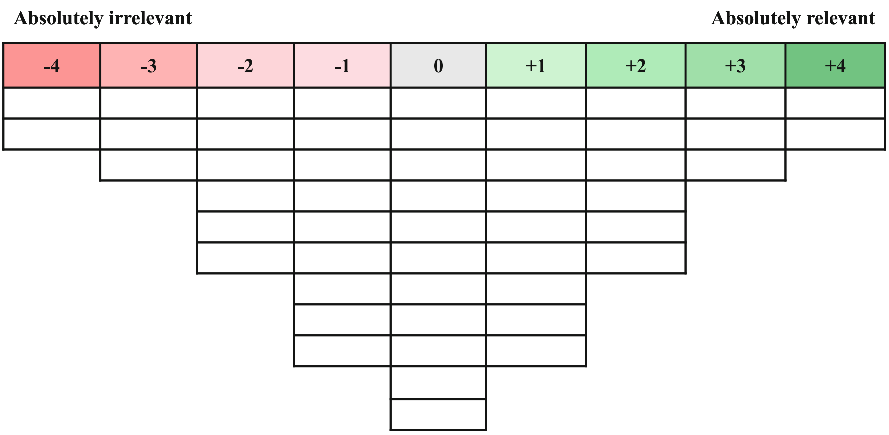

Fig. 20.2 Ranking-Grid used in the Q-sorting procedure

or think-aloud methods to obtain even richer insights into the subjectivity of participants.

3. *Third*, using multivariate data-reduction techniques (e.g., principal component or factor analysis), the Q-sorts are grouped according to similarities and summarised to identify factors or clusters of subjectivity, which can be used to sort participants into particular clusters (e.g., Watts & Stenner, 2012). Additional parameters indicate how single statements contribute to certain factors and how a statement distinguishes between single factors and therefore between different views on the topic. A more detailed description of the statistics applied can be found in the next section.
4. In the *fourth* and last step of Q studies, factors are interpreted based on estimated parameters. Based on this interpretation, each factor is given a label according to its most distinguishable attribute(s)–that is, different statements. The outcome of the study, therefore, comprises a distinct set of perspectives on the subject matter which can be used in future studies (e.g., a set of delimitable political opinions or expectations concerning an educational setting). In other words, the method allows the extraction of clusters of subjectivity existing in the population (i.e., similar viewpoints on a shared topic). In addition, Q allows classification of study participants into the clusters identified. This classification can then be used for further analysis (e.g., to test how cluster membership relates to particular dependent variables of interest or how cluster membership can be explained).

20.2.4 Analysing and Interpreting Q-Sets

Multivariate data-reduction techniques (e.g., principal component or factor analysis) are mathematical approaches for the identification of underlying patterns in datasets. As Q aims at the identification of patterns of subjectivity, such methods are used to group the discrete viewpoints of individuals, meaning the Q-sorts, and to summarise them into a circumscribed number of factors or clusters representing a shared perspective on the topic (e.g., Watts & Stenner, 2012). However, in Q, compared to classical approaches, the variables of interest are not the statements but the participants' Q-sorts, which each contain the relative placements of all statements on the grid of one person. Therefore, in Q, factors are extracted across participants rather than from statements (Webler et al., 2009). Hence, researchers are not concerned with the relation between the statements themselves but look instead for patterns in the relative placement of the statements on the grid (Young & Shepardson, 2018). More specifically, in classical survey studies, factor analysis is conducted across items wherein items that are rated similarly are thought to describe the same underlying construct (e.g., the importance of teaching methods, collaboration possibilities in classrooms or additional materials for students' satisfaction) and are therefore summarised into factors. However, in Q, analysis is conducted across persons such that participants who placed statements in similar positions in the grid, resulting in similar Q-sets, are grouped into one factor representing the same underlying viewpoint. For example, participants who attach higher importance to additional materials compared to collaboration possibilities in a learning environment are thought to share the same viewpoint and differ from other groups whose satisfaction might depend more on participation possibilities and not on materials. Data analysis in Q typically involves three steps: (a) correlation, (b) factor analysis, and (c) computation of factor scores (Brown, 2004).

1. In the *first step*, a correlation matrix is computed among all Q-sorts to identify similarities and differences. Highly correlated Q-sorts are considered to inherit a similar viewpoint on the topic whereas no correlation indicates different opinions; negative correlations show opposing beliefs.
2. Based on the correlation matrix, in the *second step*, similar Q-sorts are summarised into factors using factor extraction methods. The correlations among Q-sorts are represented by factor loadings, such that similar Q-sorts load on the same factor. A variety of different methods for factor extraction exists (see Akhtar-Danesh, 2016 for more detailed description of methods) to condense information (Zabala & Pascual, 2016).

   However, in all approaches, the researcher decides on the number of factors to extract, a decision for which no firm rules can be given (Brown, 1993; Coogan & Herrington, 2011). Decisions can be made similar to classic approaches—that is, based on eigenvalues, variability explained by factors, loadings of Q-sorts, parallel analysis, visual inspection of screeplots or theoretical importance (Zabala & Pascual, 2016), which can lead to different recommendations. The final decision on the number of factors can be based on practical criteria including the distinguishability and interpretability of factors (Coogan & Herrington, 2011).

   The remaining factors are then rotated to maximise the number of Q-sorts that can, based on their loadings, clearly be assigned to one factor (Rodl et al., 2020), therefore making the structure clearer (Zabala & Pascual, 2016). Several different methods of rotation exist, which can basically be classified into orthogonal and oblique rotation. The former presumes factors to be independent from each other, whereas the latter allows factors to be correlated (see Akhtar-Danesh, 2016 for more information on rotation methods). Sometimes, rotation is conducted manually. However, this should only be done when considerable knowledge on a given respondent is at hand (Zabala & Pascual, 2016) or when the Q-sorts of participants are compared with the sorting of one specific person of interest (Coogan & Herrington, 2011). Most studies in the field of educational research use principal component analysis (PCA) with varimax rotation (Lundberg et al., 2020).
3. In a *third step*, the extracted factor solution is interpreted. In classical approaches, the resulting factor loadings are the foundation for factor interpretation; however, in Q the meaning of factors is based on the relative position of a statement within the factor: the so-called factor scores (Brown, 2004; Zabala & Pascual, 2016). Therefore, in contrast to factor loadings, which describe the relationship between factors and Q-sorts, factor scores indicate the relation between factors and statements (Zabala & Pascual, 2016) which can be seen as the score that is

typically given to the statement by participants loading high on the factor (Rodl et al., 2020). Statements with high (positive or negative) factor scores can be understood as a prototypical profile for the underlying point of view which is represented by the factor and are therefore usually used as the basis of interpretation.

Typically, a strong emphasis is placed on so-called distinguishing statements for which the scores are statistically different across factors, meaning that their relative position in some factors differs substantially from their position in other factors (Zabala & Pascual, 2016). To describe it more practically, participants with high loadings on a factor have placed the distinguishing statement in a significantly different position than participants who have low loadings on that factor (Coogan & Herrington, 2011; Rodl et al., 2020).

However, it is not only factor scores that can be included in the interpretation of factors. If participants are asked for explanations during the sorting procedure, this information can also be used as a basis to give meaning to factors (Brown, 1993). Other supplementary sources, for example field notes, demographics, knowledge from previous studies or experts, as well as interviews with persons which load high on factors, can additionally be used as supplementary information (see Albright et al., 2019 for a detailed guide on factor interpretation). Each factor is typically given a label which describes the underlying point of view.

All factors taken as a whole constitute a distinct set of perspectives on the subject matter which ideally comprises all possible opinions in the population on the topic at hand and reflects the complexity and abundance of viewpoints. The extracted clusters of subjectivity can themselves be used in future studies. Additionally, groups of people with conflicting viewpoints can be extracted and can guide research to understand the differences.

20.2.5 Assessing Advantages and Disadvantages of Q

Q offers several benefits compared to other social research methods used for similar purposes (e.g., Zabala et al., 2018). This is mainly due to the fact that Q not only depicts a data collection method but also is based on a holistic research approach which “focuses on the subjective or first-person viewpoints of its participants” (Watts & Stenner, 2012, p. 4), explicitly aiming at identifying a priori unknown and distinct sets of subjectivity existing in the population (e.g., Müller & Kals, 2004; Kamal et al., 2014). In comparison to other methods, it is exactly this strong focus on a particular purpose of the method as well as the fitting combination of data collection and analysis that makes Q an interesting and useful method for researchers interested in all facets of subjectivity.

Due to its nature, Q combines advantages of quantitative and qualitative research approaches. After all, it provides numerical results to interpret the relationship between the clustered statements included in the Q-set. This allows researchers to interpret the meaning of the identified set of subjectivity using both the wording of the different statements (qualitative aspect) as well as statistical parameter estimates that indicate how strongly a statement describes a cluster and how well it discriminates between different clusters (quantitative aspect). In addition, Q can easily be combined with think-aloud or interview techniques, resulting in even more qualitative information that can be used to interpret the statistics provided. Other benefits emerge as a result of how data are collected in a Q study. First, sorting statements into a grid differs from answering Likert-style rating scales. This can have a motivating effect on participants as Q is perceived as something unfamiliar and potentially interesting. Second, Q uncovers how divergent but related issues are interconnected by asking respondents to consider these issues simultaneously. This is because participants are required to assess all statements together in order to sort them into the grid. In survey methods or interviews, single topics are usually presented, evaluated or discussed in a much more sequential logic without taking into account how they relate to each other. Third, Q's data collection might help to mitigate certain response biases. Since participants need to sort all statements into the grid, they are bound to deal constructively also with statements that they perceive as unfitting, inappropriate or irrelevant (Zabala et al., 2018); it is impossible to ignore single statements. As a result, reactions to all statements can be collected and used for data analysis. Fourth, only relatively small sample sets are required when conducting a Q study (McKeown & Thomas, 2013; Watts & Stenner, 2012).

However, besides those advantages, Q also comes with a set of disadvantages that should be considered when choosing a suitable research strategy to answer particular research questions. A highly relevant disadvantage is the time required to collect data using Q. In fact, collecting the Q-sorts takes a lot of time compared to conventional studies, for example, questionnaire surveys with Likert-style items. Participants will mostly not be familiar with the general design of the study that focuses on sorting statements into a grid. It is therefore necessary to instruct participants well at the beginning with a personal introduction, a video or a lengthy text description. In addition, sorting 30–50 statements into a grid takes participants easily more than 15 min as all statements must be evaluated in contrast to the others, which might be experienced as cumbersome by the participants. This can be a possible reason for low participation rates or even participant drop-out in studies employing Q even though the method is perceived as interesting in the first place, which is especially relevant when conducting the study online (Dairon et al., 2017). Accordingly, even though the required sample size is small, considerable effort is needed to obtain the corresponding number of study participants.

Concerning the validity of results, it remains open as to how they can be generalised to the broader population (Kamal et al., 2014). Even if relevant sampling strategies (random, theoretical) are used, it is unclear whether the extracted viewpoints are representative of all perspectives on the subject matter or whether relevant subgroups with a certain subjectivity concerning the topic are missing. With regard to data evaluation, Q leaves less freedom of interpretation than, for example, qualitative discourse analysis and interviews, as the perspectives in Q are limited to a number of statements that are presented to the respondents and, to some extent,

Table 20.1 Advantages and Disadvantages of Q

Advantages	Disadvantages
Combines quantitative and qualitative aspects of data analysis.	Takes a long time compared to traditional studies such as questionnaire surveys.
Represents a different type of survey compared to Likert scales. This might motivate participants to take part in a Q study.	Unclear how well the results can be generalised to a larger population.
Uncovers how divergent but related issues are interconnected.	With regard to data analysis, Q leaves less freedom of interpretation than, for example, qualitative discourse analyses and interviews.
Mitigates certain response biases through data collection mode.	It is not possible to reflect on all possible perspectives on the subject both a priori and a posteriori.
Only small sample sizes are required.	No strict criteria exist that determine the final number of clusters to extract.

to the requirements of quantitative data interpretation (Zabala et al., 2018). However, this disadvantage might be compensated for when participants are interviewed in parallel or when think-aloud techniques are employed. Another relevant limitation is seen in the process of creating the Q-set. It is impossible, both a priori and a posteriori, to reflect all possible perspectives on the topic in the developed statements. That is why sufficient time and effort needs to be put into the development of different statements to include as many perspectives on the subjectivity in question as possible. As described before, researchers should draw on all kinds of resources for this purpose (e.g., prior research, interviews, group discussions, magazines, web forums). And finally, similar to other methods that employ multivariate data-reduction techniques, no strict criteria exist of how many cluster solutions should be extracted. However, researchers should follow accepted guidelines here (see Preacher et al., 2013, for general aspects or Akhtar-Danesh, 2016, for specific considerations when using Q).

Scholars interested in identifying patterns of subjectivity in a group of individuals should be aware of both the advantages and disadvantages that come with Q. Table 20.1 summarises them again, providing a quick overview. To further illustrate the use of Q, a small study focusing on the expectations of student teachers in terms of their long-term school internship will be presented in the next section.

20.3 Illustrating Q Method with an Empirical Study

20.3.1 Background and Purpose of the Study

Long-term internships are an integral part of many university degrees. They are curricular components that provide students with domain-specific experience during a limited time frame in the course of their academic studies (Goller et al., 2020;

Maertz et al., 2014). In other words, internships are opportunities for workplace learning and professional development that go beyond more formal learning affordances typically offered by higher education institutions (Goller et al., 2020).

Within the context of teacher education, long-term internships mainly aim at helping student teachers to make connections between knowledge learned in university courses and vocational practices that constitute workplaces at schools (Maertz et al., 2014; Resnick, 1987; Sides & Mrvica, 2017; Ulbricht & Schubarth, 2016). However, they are also thought to allow students to gain initial practical experience as teachers, to gather insights how schools work, to develop initial routines in lesson planning and teaching, to understand their current strengths and weaknesses for future development, and to reflect whether they really want to become a teacher in the long run (e.g., Gröschner & Schmitt, 2010; Schubarth et al., 2014). Unfortunately, there is no agreement on how these goals relate to each other or on how they are hierarchically structured. In fact, evidence exists that different stakeholders responsible for organisation of the internships during teacher education have very different perspectives about what internships should achieve and how this can be done (Caruso & Goller, 2021a, b). For instance, while university staff usually perceive that it is most important that students productively connect theoretical knowledge with practical experience (Caruso & Harteis, 2020), this is not seen relevant by many teachers who act as mentors. Mentors often describe certain assignments that students get from the university during their internship time, like conducting a small research project, as a rather unnecessary task that does not contribute to professional development (Caruso & Goller, 2021b).

In a similar vein, evidence exists that students approach internships with different expectations and ideas. Endedijk et al. (2014), based on prior research of Oosterheert and Vermunt (2001), identified four distinct student groups that differ in terms of their beliefs of how teaching can be learned: (a) students who do not worry too much about their teaching capabilities and who understand learning as a result of practice alone and for whom neither feedback nor help of others is wanted, (b) students who worry about their teaching capabilities but use practice only to refine already acquired competences, (c) students who are interested in developing competences and who heavily rely on others (like mentors), and (d) students who use all available chances to develop their teaching competences and who are open for all kinds of learning affordances. Other research has found that students' expectations often do not seem to correspond to those defined by the educational administration or the university's objectives: Compared to practice, training at the university and theoretical knowledge are often of little relevance for students. Instead, some students ask for recipes for action and just want to gain experience (Caruso, 2019; Caruso & Goller, 2021a, b).

The goal of this study was to further investigate student teachers' subjective perspectives on their long-term school internship. To be more concrete, the study aims to identify whether homogeneous patterns exist on what elements of their internships they perceive as relevant for their professional development as teachers. Such clusters of different subjectivities concerning internships will be extracted by the application of Q.

20.3.2 Methods

20.3.2.1 Participants

Student teachers from two German universities who were about to start their 5-month school internship were invited to participate in this study using mailing lists. The students were offered an expense allowance in the form of a gift card over 10 €. A total of 60 students took part in the study (43 female, 17 male, median age ranged from 20 to 22 years).

20.3.2.2 Development of the Q-Set

In the present study, a theoretical approach was used for the development of the Q-set. Based on research literature on long-term internships in the context of teacher education, student teachers' beliefs about how teaching can be learned (e.g., Caruso, 2019; Caruso & Goller, 2021a, b; Endedijk et al., 2014; Gröschner & Schmitt, 2010; Schubarth et al., 2014; Ulrich & Gröschner, 2020), the goals laid down in the universities' curricula concerning the long-term internships (HRK, 2016), as well as prior knowledge of the authors, an initial set of potential statements was generated. These statements were critically discussed by all authors to select the most suitable. Inclusion criteria encompassed mainly (a) variety of perspectives, (b) understandable wording, (c) consistency of wording, and (d) clarity of proposition. The final Q-set included 51 statements in four main categories (see Table 20.2 for categories and examples, and Table 20.3 for a list of all statements).

20.3.2.3 Procedure

Data collection was carried out online. HtmlQ, an open-source software developed in HTML5, was used to conduct the Q-sorting and for the collection of additional sociodemographic information (Aproxima, 2015). To protect privacy, contact data for gift card delivery were gathered separately using LimeSurvey (LimeSurvey Project Team, 2012).

Since Q studies differ from classical questionnaires, the method was first explained using a short video which guided the participants through the different stages of the survey. In the Q-sorting, participants were asked to rank the given statements on the different aspects of their internship according to their subjective relevance for their professional development. First, subjects should sort each statement into one of three groups: absolutely irrelevant, neutral, and absolutely relevant. From these initial groups, students were asked in a second step to refine their rating by sorting the items into a predefined grid (see Fig. 20.2 in Sect. 20.2.3) roughly following a normal distribution with ranking values from −4 (*absolutely irrelevant*) to +4 (*absolutely relevant*). For this purpose, participants were instructed to reread

Table 20.2 Overview of the Q-sample

Category	Subcategory	Example statement	Number of statements
Practical teaching	Own teaching experience	Developing routines in teaching.	9
	Support from experienced teachers	Observing lessons from various teachers and, on this foundation, to develop one's own ideas of good teaching.	3
	Coping with heterogeneity	Developing and implementing individually tailored learning affordances for heterogeneous learning groups (e.g., with regard to age, gender, socio-economic status).	4
Future vision	Personal development as a teacher	Using potential mistakes as a learning opportunity for one's own development.	5
	Career orientation	Trying out to be a teacher and thereby reconsidering one's own career choice.	4
	Long-term planning	Introducing oneself to the teaching staff for future applications.	4
Research and theory	Application of knowledge	Applying the knowledge acquired throughout one's studies.	4
	Research-based learning	Understanding school and teaching by conducting own research projects.	3
	Theory-practice relationship	Becoming aware of the potentials and limits of scientific theories and models.	5
Extracurricular aspects of teaching	Students' phenomenology	Understanding what learners are concerned with these days.	3
	Insight into the school system	Getting to know school structures (e.g., tasks of teacher conferences, school conferences, class conferences, staff meetings).	7

all statements previously defined as "absolutely relevant". They then decided which two of these statements they thought to be most relevant (+4, positively labeled) and placed them in the column on the far right. Next, participants were requested to do the same for the group "absolutely irrelevant" and to place their two most irrelevant statements (−4, negatively labeled) in the column on the far left. The procedure was then repeated for the next columns (+3 and −3) until all statements were ranked and the grid was completed. In both ranking stages, participants were able to change the position of all statements as often as necessary until they were satisfied with their ranking. In a third step, subjects were requested to explain why they rated the items in the outer columns as the most relevant or irrelevant, respectively. Lastly, participants were asked to provide sociodemographic information about themselves.

Table 20.3 Factor scores for all statements

Statements	Perspective pattern 1 $N = 26$	Perspective pattern 2 $N = 19$	Perspective pattern 3 $N = 8$
Consensus statements			
Gathering experience that will help in the future to link theories, models, and concepts dealt with at the university directly to practice.	−1	−1	−1
Recognising one's own strengths and weaknesses through feedback (from experienced teachers) in order to personally develop.	3	3	2
Gaining insights into what the different staff groups at school do (e.g., school social workers, school attendants, school psychologists, integration assistants, department heads, etc.).	−2	−1	−2
Understanding school and teaching by conducting own research projects.	−3	−3	−3
Observing lessons from various teachers and, on this foundation, developing one's own ideas of good teaching.	2	1	2
Statements significantly distinguishing among all perspective patterns			
Collecting teaching materials in order to prepare for the second phase of teacher training.	−1	1	2
Identifying potential support options for low-performance learners.	−1	1	0
Trying out to be a teacher and thereby reconsidering one's own career choice.	3	−3	0
Gaining initial experience on issues of classroom disruptions and learning how to deal with them.	2	1	1
Trying out to be a teacher and thereby checking whether one is suited for the job.	4	−1	2
Recognising how one as a teacher affects the competence development of learners.	0	0	4
Identifying potential support options for high-performance learners.	−1	0	1
Recognising how knowledge learned in one's studies can be applied.	−1	0	−2
Gaining insights on what it means to be a teacher.	2	0	0
Theoretically reflecting one's own teaching experiences in the university seminars.	−2	0	−4
Getting acquainted with the process of class observations (e.g., by principals, other teachers) in order to be prepared for the second phase of teacher education.	0	0	−2
Receiving feedback on whether one is suited for school service.	3	−1	0
Planning and implementing as many different lesson topics as possible.	0	1	0
Statements significantly distinguishing Pattern 1 from Patterns 2 and 3			
Recognising one's own strengths and weaknesses with regard to my prospective job as a teacher.	4	2	1
Developing routines in teaching.	0	2	2
Recognising previously acquired didactic approaches to the conception of teaching processes in practice.	0	−1	−1

(continued)

Table 20.3 (continued)

Statements	Perspective pattern 1 $N = 26$	2 $N = 19$	3 $N = 8$
Improving one's own research methodological skills through research-based learning.	−2	−4	−3
Statement significantly distinguishing Pattern 1 from Pattern 2			
Getting to know school structures (e.g., tasks of teacher conferences, school conferences, class conferences, staff meetings).	−1	0	−1
Statements significantly distinguishing Pattern 2 from Patterns 1 and 3			
Familiarising myself with teaching methods by observing other teachers.	0	1	−1
Learning to diagnose students' weaknesses and strengths.	0	1	−1
Developing concrete ideas about learners in a certain grade and how they behave.	0	0	−1
Developing and implementing individually tailored learning affordances for heterogeneous learning groups (e.g., with regard to age, gender, socio-economic status).	0	1	−1
Reviewing one's own ideas about schools and teaching.	1	−1	1
Reflecting one's own abilities and skills on the basis of practical experiences.	1	0	2
Creating one's own first lesson drafts and gaining experience by putting them into practice.	1	4	1
Trying things out without getting graded on.	1	2	0
Becoming aware of the potentials and limits of scientific theories and models.	−3	−2	−2
Understanding curricular and extracurricular problems and difficulties of learners.	0	0	−1
Learning through the guidance of (experienced) teachers how lessons are planned and conducted in practice.	1	4	1
Checking if the knowledge acquired during the studies is useful in practice.	−1	−2	−1
Statements significantly distinguishing Pattern 3 from Patterns 1 and 2			
Reflecting practical experience on the basis of scientific theories, models, and concepts.	−2	−2	−3
Reviewing one's own abilities for teacher activities outside of teaching.	−2	−2	3
Understanding what learners are concerned with these days.	−1	−1	0
Bringing previously learned theories, models and concepts to life.	−1	−1	−2
Using potential mistakes as a learning opportunity for one's own development.	2	2	1
Getting to know the full range of tasks connected to the teaching profession (teaching, advising, organising, innovating, educating, etc.)	1	2	0
Reflecting one's own abilities and skills with the help of (experienced) teachers.	2	3	0
Getting insights into school structures and administration.	0	−1	4

(continued)

Table 20.3 (continued)

Statements	Perspective pattern 1 $N = 26$	2 $N = 19$	3 $N = 8$
Trying out different teaching methods.	2	3	1
Applying the knowledge acquired throughout one's studies.	1	0	−2
Teaching the same lesson in different classes.	−2	−2	0
Expanding one's own professional network (e.g., teachers, principal).	−3	−2	3
Assessing the quality of teaching by researching lessons empirically.	−4	−3	−4
Further developing one's own strengths and compensating deficits by reflecting on one's own teaching experiences.	1	1	3
Planning and conducting a comprehensive sequence of lessons on a particular topic.	1	2	0
Introducing oneself to the teaching staff for future applications.	−4	−4	1

20.3.2.4 Data Analysis

After sorting, each item was given a number according to the rank value it was assigned to on the grid to obtain the individual rankings. Using the R package qmethod (Zabala, 2014) these Q-sorts were factor analysed using principal component analysis with varimax rotation. Factor scores for each factor were computed incorporating the underlying viewpoint and therefore were the basis for interpretation.

20.3.3 Results

Factor analysis revealed three different factors or, in other words, three different patterns of perspectives that accounted for 45.08% of the variance in the Q-sorts. Of the 60 participants, 26 were assigned to the first factor, 19 to the second factor, and eight to the third factor. The viewpoints of the remaining seven participants were not covered by the extracted factors.

In the following, the factors are described based on their characterising statements (i.e., statements ranked at both ends, −4 or +4, of the average Q-sort for the respective factor). If statements were ranked as +3/−3 and +2/−2 and significantly distinguished the factor from other viewpoints, those were also included. Factor scores for all statements are displayed in Table 20.3. At this stage, factors will be described purely on statistical grounds. An interpretation including the assignment of factor labels will be part of the discussion in Sect. 20.3.4.

20.3.3.1 Perspective Pattern 1

The first perspective pattern accounts for 19.43% of the variance among the Q-sorts. According to the characterising statements, students of this group agreed that "*Trying out the role of the teacher and thereby checking whether I am suited for the job*" (+4) and "*Recognising one's own strengths and weaknesses with regard to my prospective job as a teacher*" (+4) are the most relevant aspects of their internship. On the other hand, "*Assessing the quality of teaching by researching lessons empirically*" (−4) and "*Introducing oneself to the teaching staff for future applications*" (−4) were ranked as the most irrelevant aspects. Compared to the other groups, students assigned to this group ranked the statements "*Trying out the role of the teacher and thereby reconsidering one's own career choice*" (+3), "*Receiving feedback on whether one is suited for school service*" (+3) and "*Improving one's own research methodological skills through research-based learning*" (−2) as significantly more relevant. Given that 26 of the 60 participants were assigned to the first group, this pattern represents the most prominent perspective in the sample.

20.3.3.2 Perspective Pattern 2

The students assigned to this group rated the statements "*Creating one's own first lesson drafts and gaining experience by putting them into practice*" (+4) and "*Learning through the guidance of (experienced) teachers how lessons are planned and conducted in practice*" (+4) as the most relevant, and the statements "*Improving one's own research methodological skills through research-based learning*" (−4) and "*Introducing oneself to the teaching staff for future applications*" (−4) as the most irrelevant aspects of their internship. According to the distinguishing statements, "*Trying things out without getting graded on*" (+2) and "*Becoming aware of the potentials and limits of scientific theories and models*" (−2) were ranked as significantly more relevant than in the other groups, whereas "*Checking if the knowledge acquired during the studies is useful in practice*" (−2) was seen as a significantly less relevant aspect of their internship. This perspective pattern constitutes 16.64% of the total variance in the Q-sorts. There were 19 students included in this pattern.

20.3.3.3 Perspective Pattern 3

For the students sharing this perspective pattern, the most relevant aspects of their internship were "*Recognising how one as a teacher affect the competence development of learners*" (+4) and "*Gaining insights into the external conditions and organization inside of school*" (+4), whereas the most irrelevant aspects were "*Theoretically reflecting one's own teaching experiences in the university seminars*"

(−4) and "*Assessing the quality of teaching by researching lessons empirically*" (−4). Distinguishing from the other groups, they rated the statements "*Reviewing one's own abilities for teacher activities outside of teaching*" (+3), "*Expanding one's own professional network (e.g., teachers, principal)*" (+3) and "*Further developing one's own strengths and compensating deficits by reflecting on one's own teaching experiences*" (+3) as significantly more relevant. The statements "*Reflecting practical experience on the basis of scientific theories, models, and concepts*" (−3), "*Bringing previously learned theories, models and concepts to life*" (−2) and "*Applying the knowledge acquired throughout one's studies*" (−2), on the other hand, were ranked as significantly more irrelevant; 9.01% of the variance in the Q-sorts was based on the third perspective pattern to which eight students were assigned.

20.3.3.4 Common Perspective Across All Participants

Five statements were shared by all three described factors and thus can be seen as representing the consensus of all perspective patterns in the assessment of the importance of various aspects of the internship. All three groups appreciate the exchange with experienced teachers as they all ranked the statements "*Familiarising myself with teaching methods by observing other teachers*" and "*Recognising one's own strengths and weaknesses through feedback (from experienced teachers) in order to personally develop*" of high relevance in their internship. On the other hand, the statements "*Gathering experience that will help in the future to link theories, models, and concepts dealt with at the university directly to practice*" and "*Gaining insights into what the different staff groups at school do (e.g., school social workers, school attendants, school psychologists, integration assistants, department heads, etc.)*" as well as "*Understanding school and teaching by conducting own research projects*" were ranked as irrelevant by all groups.

20.3.4 Discussion

Using Q, the present study identified three different viewpoints shared by students, summarising aspects of long-term school internships that are perceived as relevant for their professional development. A 3-factor solution was chosen due to good interpretability as well as a high number of assigned study participants. In the following paragraphs the three extracted perspective patterns will be characterised and briefly discussed.

Evaluating Career Choice (Perspective Pattern 1) This cluster represents the viewpoints of students who are unsure whether they want to become teachers or not. For these students, the internship is perceived mainly as providing opportunities for trying themselves out as teachers, for checking their initial career choice, and also for exploring strengths and weaknesses in regard to a prospective career in teaching. These students also hope to receive external feedback from their school mentors on their suitability to become teachers. Accordingly, these students expect the internship to be some kind of testing arena in which they can check what it takes to work as a teacher and whether they are suited for this profession. However, the students consider it strongly irrelevant to make themselves known in their particular internship school for future applications. This might further indicate that students who share this pattern do not yet see themselves as future teachers.

Gaining Experience in Teaching (Perspective Pattern 2) Students who share this perspective pattern can be characterised by their shared desire for practical work. They see the internship as an opportunity to create their own lesson drafts and consequently to implement them in their own teaching. Through practical work, students in this group want to gain hands-on experience. In addition, they want to get feedback from more experienced teachers on how well they are able to teach. Similar to students from Perspective pattern 1, students in this pattern consider it irrelevant to make themselves known in the internship school for future applications. However, as they show no signs of insecurity about their career choice, this finding might indicate that students from Perspective pattern 2 are highly confident that they will find an appropriate school for the second phase of teacher education and that they do not need familiar acquaintances for this purpose.

Gaining practical experience in teaching is the most relevant aspect of their internship for students in this group. Similar findings have been reported by other studies. For instance, Oosterheert and Vermunt (2001) described five orientations to learning to teach; students in three of these distinct orientations understand professional development mainly as a function of learning by doing and referring to prior experiences. Endedijk et al. (2014) as well as Festner et al. (2020) identified similar patterns in which students aim to develop their teaching competences mainly by relying on practice and the feedback of mentors. Students who share this particular kind of subjectivity have only a low interest in research-based learning and theoretical discussions.

According to previous interpretations (e.g., Oosterheert & Vermunt, 2001; Festner et al., 2020), this could indicate that students in this group will not base their lesson preparation in the future on scientific theories and models, but exclusively on practical experience. Compared to the other two clusters, subjects sharing this perspective might even regard their university education as unnecessary for their practical work, as they do not consider knowledge acquired in academic courses as relevant for practical use.

Going Beyond Teaching (Perspective Pattern 3) Students who were classified in this perspective pattern share an interest in aspects of a teaching career that are not directly connected with teaching as such. They see their internship as an opportunity to gain insights into structures, processes, and tasks beyond teaching. Schools are seen as a larger system and they want to understand it as such. They follow a strong development mode and expect their internship to be an opportunity for both testing their own skills and expanding their professional network. Taken together, this student group seems to adopt a more future-oriented perception.

A similar pattern of so-called open-meaning-oriented or versatile students could also be identified in earlier research (Oosterheert & Vermunt, 2001; Festner et al., 2020). These students use different kinds of learning opportunities to expand their frame of reference for professional development (Festner et al., 2020). In other words, they use all available opportunities to develop their teaching skills and are open to all types of learning opportunities (Endedijk et al., 2014). However, students who share Perspective pattern 3 differ from the groups identified in the other studies: both research-based learning and relating theory to practice are still aspects of the internship that are not perceived as relevant as other aspects for their professional development.

Consensus Over All Perspective Patterns The three patterns have in common that they consider research-based learning to be least relevant for their professional development compared to the other aspects of the internship. In addition, the internship is not perceived as an opportunity to make cognitive connections between knowledge learned in university courses and vocational practices that constitute workplaces at schools. In fact, it seems that the majority of students in this sample consider theoretical knowledge as irrelevant for their later profession (see also Hascher, 2006). This may be due to the fact that knowledge taught at university is understood as not being relevant for practice by many students. If internships explicitly aim at students making connections between knowledge learned in academia and practice contexts, then professional guidance is needed. In fact, students need to receive support from lecturers and experienced teachers who actively help them to relate practical experiences and theories in such a way as to generate knowledge that contributes to their professional development (Caruso & Harteis, 2020).

20.4 Conclusion

Subjectivities play an important role in many educational contexts. They have the power to determine how learners interact with learning environments or how instructors plan interventions that aim at learning and development. Although subjectivities, by definition, are idiosyncratic in nature, it is not uncommon that groups of individuals share common patterns of values, beliefs, preferences, opinions, subjective theories, and so on. In particular, knowledge about such shared subjectivities is highly relevant from both a scientific but also a practical perspective as it allows for (a) explaining why groups of learners act and learn differently, as well as (b) tailoring specific learning affordances that better meet learners' demands, which could result in more efficient or efficacious learning and development processes.

Having adequate methods to investigate subjectivities and to identify patterns of shared perspectives is an important foundation to further understand and improve learning and development processes. In this chapter, Q method was introduced as a holistic research method that has specifically been developed for the assessment of such subjectivities and that has several advantages compared to other methods, making it a valuable tool in this research field. Q allows for identification of homogeneous patterns of subjectivity from a larger, more heterogeneous set of (a priori unknown) subjectivities existing in the population. This is done by asking study participants to sort a number of statements concerning a certain topic (e.g., their values, beliefs, preferences, opinions, subjective theories toward something) into a grid roughly following a normal distribution indicating, for example, whether they see a given proposition as more important or if they agree with it more in relation to other statements. Based on multivariate data-reduction techniques, these data are then used to extract subjectivity patterns shared among study participants. The resulting patterns can be used for phenomenological interpretation as well as either independent or dependent variables for further analyses.

Although Q has been in existence for more than 50 years, it seems to be a less well-known research method in educational science. Moreover, almost no studies exist that use Q to investigate professional learning and development (see, however, Schick et al., 2021). The main aim of this chapter was to encourage scholars who are interested in professional learning and development familiar with Q. For instance, Q could be used to further investigate student teachers' beliefs about how to learn teaching, employees' preferences and priorities concerning professional development opportunities, educators' self-perceived roles in planning and teaching professional development courses, or human resource development practitioners' perceptions of different learning resources. As long as the goal of a research project is the identification of patterns of subjectivity, Q could be an appropriate research method to use.

As book chapters are limited in their scope, interested readers are referred to existing textbook literature (e.g., McKeown & Thomas, 2013; Watts & Stenner, 2012) as well as *Operant Subjectivity*, a journal dedicated to Q method and research conducted with Q, for more detailed instruction as well as further topics. Besides the conceptual introduction of Q, a small study was presented in this chapter to illustrate the method for researching professional learning and development. It is hoped that this study shows on an exemplary level how Q is applied in research projects and what can be done with it.

References

Akhtar-Danesh, N. (2016). Rejoinder: An overview of the statistical techniques in Q methodology: Is there a better way of doing Q analysis? *Operant Subjectivity: The International Journal of Q Methodology, 38*(3/4), 29–36. https://doi.org/10.15133/j.os.2016.007

Akhtar-Danesh, N. (2021). *QFACTOR: Stata module to perform Q-analysis on Q-sorts using different factor extraction and factor rotation techniques* [Computer software]. https://EconPapers.repec.org/RePEc:boc:bocode:s458326

Albright, E. A., Christofferson, K., McCabe, A., & Montgomery, D. (2019). Lessons learned: Some guidelines to factor interpretation. *Operant Subjectivity: The International Journal of Q Methodology, 41*, 134–146. https://doi.org/10.15133/j.os.2019.001

Aproxima. (2015). *HtmlQ (Version 1.0.4)* [Computer software]. GitHub. https://github.com/aproxima/htmlq

Banasick, S. (2019). *Ken-Q analysis (Version 1.0.6)* [Computer software]. https://shawnbanasick.github.io/ken-q-analysis/

Barnes, C., Angle, J., & Montgomery, D. (2015). Epistemologies of science instruction. *School Science and Mathematics, 115*(3), 141–150. https://doi.org/10.1111/ssm.12111

Billett, S. (2006). Work, subjectivity and learning. In S. Billett, T. Fenwick, & M. Somerville (Eds.), *Work, subjectivity and learning* (pp. 1–20). Springer.

Brown, S. R. (1993). A primer on Q methodology. *Operant Subjectivity, 16*(3/4), 91–138.

Brown, M. M. (2004). *Illuminating patterns of perception: An overview of Q methodology*. Carnegie Mellon University. https://doi.org/10.1184/R1/6574184

Caruso, C. (2019). *Das Praxissemester von angehenden Lehrkräften. Ein Mixed-Methods-Ansatz zur Exploration ausgewählter Effekte* [Long-term internships of prospective teachers. A mixed-methods study exploring selected effects]. Springer VS.

Caruso, C., & Goller, M. (2021a). Aufgaben, Ziele und Betreuungsstrategien von am Praxissemester beteiligten Lehrkräften. Zur Exploration der Perspektive und der Erfahrungen von Mentor*innen [Tasks, goals, and mentoring strategies of teachers mentoring during long-term internships. An exploration of perspectives and experiences from mentors]. *PFLB - PraxisForschungLehrer*innenBildung, 3*(1), 104–113. https://doi.org/10.11576/pflb-4541

Caruso, C., & Goller, M. (2021b). Das Praxissemester aus der Perspektive von Mentor*innen: Aufgaben der Lernorte und wahrgenommene Veränderungsbedarfe [Mentors' perspectives on long-term school internships: The role of different stakeholders in teacher students' professional development as well as perceived needs for change]. *Lehrerbildung auf dem Prüfstand, 14*(2), 5–9.

Caruso, C. & Harteis, C. (2020). Inwiefern können Praxisphasen im Studium zu einer Theorie-Praxis-Relationierung beitragen? Implikationen für die professionelle Entwicklung angehender Lehrkräfte [How can practise phases affect how students relate theory to practice: Implication for the professional development of prospective teachers]. In K. Rheinländer, & D. Scholl (Eds.), *Verlängerte Praxisphasen in der universitären Lehrerbildung. Spannungsfelder zwischen Theorie, Praxis und der Bestimmung von Professionalisierung* (pp. 58–73). Klinkhardt.

Coogan, J., & Herrington, N. (2011). Q methodology: An overview. *Research in Secondary Teacher Education, 1*(2), 24–28. https://doi.org/10.15123/uel.8604v

Cooper, L., Baron, C., Grim, L., & Sandling, G. (2018). Teaching teachers onsite: Using evaluation to develop effective professional development programs. *Journal of Museum Education, 43*(3), 274–282. https://doi.org/10.1080/10598650.2018.1489196

Dairon, M., Clare, S., & Parkins, J. R. (2017). Participant engagement and data reliability with internet-based Q methodology: A cautionary tale. *Operant Subjectivity: The International Journal of Q Methodology, 39*(3/4), 46–59. https://doi.org/10.15133/j.os.2017.011

Elen, J., & Lowyck, J. (2000). Instructional metacognitive knowledge: A qualitative study on conceptions of freshmen about instruction. *Journal of Curriculum Studies, 32*, 421–444. https://doi.org/10.1080/002202700182637

Endedijk, M. D., Vermunt, J. D., Meijer, P. C., & Brekelmans, M. (2014). Students' development in self-regulated learning in postgraduate professional education: A longitudinal study. *Studies in Higher Education, 39*(7), 1116–1138. https://doi.org/10.1080/03075079.2013.777402

Entwistle, N. J., & Peterson, E. R. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments.

International Journal of Educational Research, 41(6), 407–428. https://doi.org/10.1016/j.ijer.2005.08.009

Festner, D., Gröschner, A., Goller, M., & Hascher, T. (2020). Lernen zu Unterrichten—Veränderungen in den Einstellungsmustern von Lehramtsstudierenden während des Praxissemesters im Zusammenhang mit mentorieller Lernbegleitungund Kompetenzeinschätzung [Learning to teach – Changes in learning-to-teach patterns of student teachers during long-term internships in connection with mentoring and self-perceived competence assessment]. In I. Ulrich & A. Gröschner (Eds.), *Praxissemester im Lehramtsstudium in Deutschland: Wirkungen und Effekte auf Studierende. Edition der Zeitschrift für Erziehungswissenschaft* (pp. 209–241). Springer VS. https://doi.org/10.1007/978-3-658-24209-1_7

Goller, M., Harteis, C., Gijbels, D., & Donche, V. (2020). Engineering students' learning during internships: Exploring the explanatory power of the job demands-control-support model. *Journal of Engineering Education, 109*(2), 307–324. https://doi.org/10.1002/jee.20308

González Rey, F. L. (2019). Subjectivity and discourse: Complementary topics for a critical psychology. *Culture & Psychology, 25*(2), 178–194. https://doi.org/10.1177/1354067X18754338

Groeben, N. (Ed.). (1988). *Das Forschungsprogramm Subjektive Theorien: Eine Einführung in die Psychologie des reflexiven Subjekts* [Research programme subjective theories: An introduction into the psychology of reflexive subjects]. Francke.

Gröschner, A., & Schmitt, C. (2010). Wirkt, was wir bewegen?—Ansätze zur Untersuchung der Qualität universitärer Praxisphasen im Kontext der Reform der Lehrerbildung [Do the things work that we move? – Approaches to investigate the quality of practice phases in higher education in terms of reforms of teacher education]. *Erziehungswissenschaft, 21*(40), 89–97. https://doi.org/10.25656/01:2738

Harteis, C., Gruber, H., & Lehner, F. (2006). Epistemological beliefs and their impact on work, subjectivity and learning. In S. Billett, M. Somerville, & T. Fenwick (Eds.), *Work, subjectivity and learning* (pp. 123–140). Springer.

Hascher, T. (2006). Veränderungen durch das Praktikum: Eine empirische Untersuchung zur Wirkung von schulpraktischen Studien in der Lehrerbildung [Development due to internships—An empirical investigation about the efficacy of long-term internships in teacher education]. In C. Allemann-Ghionda & E. Terhart (Eds.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (pp. 130–148). Beltz.

HRK. (2016). *Praktika im Studium. Praxis integrieren und Qualität von Praktika erhöhen* [Internships as part of a university degree: Integrating practice and increasing the quality of internships]. Bonn. https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/Impuls_Praktika__2_pdf

Kamal, S., Kocór, M., & Grodzinska-Jurczak, M. (2014). Quantifying human subjectivity using Q method: When quality meets quantity. *Qualitative Sociology Review, 10*(3), 60–79.

LimeSurvey Project Team. (2012). *LimeSurvey: An open source survey tool*. LimeSurvey Project. Hamburg. http://www.limesurvey.org

Luhrmann, T. M. (2006). Subjectivity. *Anthropological Theory, 6*(3), 345–361. https://doi.org/10.1177/1463499606066892

Lundberg, A., de Leeuw, R., & Aliani, R. (2020). Using Q methodology: Sorting out subjectivity in educational research. *Educational Research Review, 31*, 100361. https://doi.org/10.1016/j.edurev.2020.100361

Maertz, C. P., Stoeberl, P. A., & Marks, J. (2014). Building successful internships: Lessons from the research for interns, schools, and employers. *Career Development International, 19*(1), 123–142. https://doi.org/10.1108/CDI-03-2013-0025

McKeown, B., & Thomas, D. (2013). *Q methodology* (2nd ed.). SAGE.

Müller, F. H., & Kals, E. (2004). Die Q-Methode. Ein innovatives Verfahren zur Erhebung subjektiver Einstellungen und Meinungen [Q method: An innovative approach to capture subjective attitudes and opinions]. *Forum: Qualitative Sozialforschung, 5*(2), Art. 34.

OECD. (2009). *Creating effective teaching and learning environments: First results from TALIS*. OECD. https://www.oecd.org/education/school/43023606.pdf

Oosterheert, I. E., & Vermunt, J. D. (2001). Individual differences in learning to teach: Relating cognition, regulation and affect. *Learning and Instruction, 11*(2), 133–156.

Ortner, S. B. (2005). Subjectivity and cultural critique. *Anthropological Theory, 5*(1), 31–52. https://doi.org/10.1177/1463499605050867

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research, 48*(1), 28–56. https://doi.org/10.1080/00273171.2012.710386

Resnick, L. B. (1987). The 1987 presidential address learning in school and out. *Educational Researcher, 16*(9), 13–54. https://doi.org/10.3102/0013189X016009013

Reusser, K., & Pauli, C. (2014). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern [Job-related beliefs of teachers]. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 642–661). Waxmann.

Rieber, L. P. (2020). Q methodology in learning, design, and technology: An introduction. *Educational Technology Research and Development, 68*(5), 2529–2549.

Rodl, J. E., Cruz, R. A., & Knollman, G. A. (2020). Applying Q methodology to teacher evaluation research. *Studies in Educational Evaluation, 65*, 100844. https://doi.org/10.1016/j.stueduc.2020.100844

Salaj, I., & Kiš-Glavaš, L. (2017). Perceptions of students with disabilities regarding their role in the implementation of education policy: A Q method study. *Hrvatska Revija Za Rehabilitacijska Istrazivanja, 53*, 47–62.

Schick, K., Gartmeier, M., & Berberat, P. O. (2021). Senior medical student attitudes towards patient communication and their development across the clinical elective year—A Q-methodology study. *Frontline Learning Research, 9*(1), 1–29. https://doi.org/10.14786/flr.v9i1.583

Schmolck, P. (2008, October). *Common and specific approaches in the analysis of Q-sort data with PQMethod*. Keynote speech presented at the 24th Annual Q Conference. Hamilton, Ontario. Available http://schmolck.org/qmethod/Schmolck_Hamilton-2008_Talk.ppt

Schubarth, W., Gottmann, C., & Krohn, M. (2014). Wahrgenommene Kompetenzentwicklung im Praxissemester und dessen berufsorientierende Wirkung. Ergebnisse der ProPrax-Studie [Perceived competence development during long-term school internship as well as career-related effects. Evidence from ProPrax study]. In K.-H. Arnold, A. Gröschner, & T. Hascher (Eds.), *Schulpraktika in der Lehrerbildung. Theoretische Grundlagen, Konzeptionen, Prozesse und Effekte* (pp. 201–219). Waxmann.

Sides, C. H., & Mrvica, A. (2017). *Internships: Theory and practice*. Routledge.

Stephenson, W. (1953). *The study of behaviour: Q-technique and its methodology*. University of Chicago Press.

Summak, S. M., & Kalman, M. (2020). A Q-methodological analysis of school principals' decision-making strategies during the change process at schools. *CEPS Journal, 10*(2), 123–144. https://doi.org/10.26529/cepsj.527

Ulbricht, J., & Schubarth, W. (2016). *Nexus Impulse für die Praxis. Praktika im Studium: Praxis integrieren und Qualität von Praktika erhöhen* [Nexus impulses for practice. Internships as part of higher education: Integrating practice and increasing the quality of internships]. Hochschulrektorenkonferenz.

Ulrich, I., & Gröschner, A. (2020). *Praxissemester im Lehramtsstudium in Deutschland: Wirkungen auf Studierende* [Long-term school internships as part of German teacher education: Effects on students]. Springer VS.

Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method & interpretation*. SAGE.

Webler, T., Danielson, S., & Tuler, S. (2009). *Using Q method to reveal social perspectives in environmental research* (Vol. 54, pp. 1–45). Social and Environmental Research Institute.

Young, J. M., & Shepardson, D. P. (2018). Using Q methodology to investigate undergraduate students' attitudes toward the geosciences. *Science Education, 102*(1), 195–214. https://doi.org/10.1002/sce.21320

Zabala, A. (2014). qmethod: A package to explore human perspectives using Q methodology. *The R Journal, 6*(2), 163–173.

Zabala, A., & Pascual, U. (2016). Bootstrapping Q methodology to improve the understanding of human perspectives. *PLoS One, 11*(2), 1–19. https://doi.org/10.1371/journal.pone.0148087

Zabala, A., Sandbrook, C., & Mukherjee, N. (2018). When and how to use Q methodology to understand perspectives in conservation research: The Q methodology. *Conservation Biology, 32*(5), 1185–1194. https://doi.org/10.1111/cobi.13123

Susann Leidig is a doctoral candidate at the Institute for Educational Science at the Friedrich Schiller University Jena. Her research interests include social and educational inequality, teaching processes, and digital learning environments.

Hanna Köhler (MSc) is a doctoral candidate at the Institute for Psychology and research associate at the Institute of Educational Science at the Friedrich Schiller University Jena. Her research interests encompass criterion-referenced computerized adaptive testing, digital learning, and difficulties in the transition from school to university in terms of mathematical competencies.

Carina Caruso (PhD) is a post-doc at the Institute of Educational Science at the Paderborn University in Germany focussing on empirical educational research. Her research interests comprise students' professional learning in internships, teachers' professional development, and expertise development within the context of digitalised work environments.

Michael Goller (PhD) is a research associate at the Institute of Educational Science at Paderborn University. Between October 2018 and March 2020, he served as visiting professor at the University of Bamberg and between April and September 2021 as a visiting professor at the University of Jena. His current research interests comprise expertise development and workplace learning in general, including in the particular context of digitalisation, as well as learning and development of students at the boundary of institutionalised and less institutionalised learning environments such as internships or particular learning arrangements in VET.

Chapter 21
Eye Tracking in Professional Learning and Development: Uncovering Expertise Development Among Residents in Radiology

Helen Jossberger

Abstract Eye tracking is a particularly interesting technology for investigating professional learning and development in vision-intensive professions. At the workplace, professionals are often confronted with complex visual tasks that they must solve quickly. From a psychological and educational point of view, it is interesting to examine professionals' attentional behaviours during work activities and to understand how they analyse and interpret visual input. Common to vision-intensive professions is the notion that professionals need the ability to perceive the relevant from the irrelevant and correctly interpret it. A radiologist, for instance, needs to make correct medical diagnoses based on complex visual material. Eye tracking enables the measurement of eye movements. By tracking the movements of the eyeball(s), we can learn where a person is looking, the duration of his or her gaze, and the order of the eye movements. Eye-tracking technology does not explain the underlying motives of looking; it only visualises gaze behaviour. The focus of this chapter is the meaning of eye tracking, the purposes of its application and the aspects of eye tracking that warrant attention. To illustrate the challenges and benefits of using eye-tracking technology in workplace learning, an empirical study in the medical domain is presented. A longitudinal quasi-experimental study with three measurement points was designed with the aim of investigating expertise development among eight residents of a radiology department and to identify changes in their way of analysing and diagnosing medical X-ray images during their residency.

Keywords Eye tracking · Workplace learning · Diagnostic reasoning · Visual expertise professional development · Residency in radiology

---

H. Jossberger (✉)
Faculty of Human Sciences, Department of Educational Science, University of Regensburg, Regensburg, Germany
e-mail: helen.jossberger@ur.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_21

21.1 Introduction

Eye tracking is a particularly interesting technology for investigating professional learning and development in vision-intensive professions (such as medicine, air traffic control, or teaching). Common to vision-intensive professions is the notion that professionals need the ability to perceive the relevant from the irrelevant and correctly interpret it. A radiologist, for instance, needs to make correct medical diagnoses based on complex visual material. Such visual material poses challenges to the cognitive and perceptual system; especially inexperienced individuals face difficulties as they are distracted by visually salient information instead of attending to thematically relevant areas (Jarodzka et al., 2010). Research on how the perceptual and cognitive processes advance and adapt with increasing expertise during professional development and how the learning process can be improved is scarce. Eye tracking enables the measurement of eye movements while a person processes visual information. By tracking the movements of the eyeball(s), we can learn where a person is looking, the duration of his or her gaze, and the order of the eye movements.

It is impossible to cover the full complexity of eye tracking within this chapter; therefore, readers are encouraged to consult comprehensive, introductory books that explain eye tracking usage and methodology in-depth (Duchowski, 2017; Holmqvist et al., 2011; Holmqvist & Andersson, 2017). The history of eye tracking is described by Wade and Tatler (2005). Recently, Carter and Luke (2020) have published an article about best practices in eye tracking research, which provides an overview and practical guide for researchers who want to employ this rich experimental method. In this chapter, the focus is the meaning of eye tracking, the purposes of its application and the aspects of eye tracking that warrant attention. Moreover, a small, longitudinal, eye-tracking study in the medical domain is employed as an empirical illustration.

21.2 Eye Tracking in Workplace Learning

In a review study, Lai et al. (2013) showed an increasing use of eye-tracking technology to explore learning in particular patterns of information processing. Especially from the year 2009 onwards, the increase in publications is notable. This trend can also be observed in an international networking organisation for junior and senior researchers in education – the European Association for Research on Learning and Instruction (EARLI) community. For instance, in 2014, the most recent EARLI Special Interest Group (SIG 27) was founded to focus on online measures of learning processes. Within the EARLI, a SIG is a smaller community with a shared interest in advancing knowledge in a particular field; members meet regularly, exchange ideas, learn from one another, and collaborate. Furthermore, special issues published in journals such as Frontline Learning Research (Harteis et al., 2018) or Educational Psychology Review (Jarodzka et al., 2021) show that capturing learning and professional development with online measures such as eye tracking is considered important. While eye tracking used to be expensive and effortful, nowadays, software methodology and hardware technology, are more accessible, applicable, and affordable, which has also increased the rate of research (Harteis et al., 2018).

Many professional fields are vision-intensive, and when researchers want to study professionals' visual perception, eye tracking is the method of choice (Holmqvist et al., 2011). Eye tracking allows the measurement and recording of eye movements in relation to an external stimulus to learn what a person saw (Jarodzka et al., 2021). Thus, eye tracking is a reliable tool for exploring questions concerning the allocation of visual attention (Carter & Luke, 2020). The eyes are considered windows to the mind, as humans are usually motivated to move the eyes to stimuli that are cognitively processed. For instance, when reading text, researchers presumed that the eyes fixate a word as long as it is processed supporting the assumption that the eyes reflect mental processing (Just & Carpenter, 1980; Rayner & Reingold, 2015). Although fixating something usually means attending to it, it is important to note that it need not be the case. The assumption by Just and Carpenter has been disproved and research has indicated that visual attention precedes gaze somewhat (Holmqvist & Andersson, 2017). Nevertheless, attending to something is required for further processing and eye tracking can capture these moment-by-moment insights due to its high temporal sensitivity making real-time cognitive processes observable in an objective and nonintrusive manner (Carter & Luke, 2020; Kok & Jarodzka, 2017a). Moreover, as individuals do not have full control of their eye movements and have difficulties remembering where they looked, tracking their eyes can make nonconscious processing visible (Kok et al., 2017). By tracking a person's eye movements, it is possible to follow along the path of attention, which can provide insight into what the person found interesting or what drew his or her attention (Duchowski, 2017). Still, caution is advised, because eye tracking does not directly measure what information is processed; instead, it indicates what can be processed (Kok & Jarodzka, 2017b). As researchers do not uncover underlying motives of an individual when solemnly looking at the tracked gaze behaviour, eye movement recordings are often enriched with verbal protocols (Ericsson, 2018) to examine how a person interprets what was seen and visually perceived.

At the workplace, professionals are often confronted with complex visual tasks that they must solve quickly. However, the information processing capacity of human beings is limited. Thus, professionals need attention to focus their mental capacities on selections of the sensory input so that a particular stimulus of interest is processed successfully. The brain needs to integrate all sensory input to construct a coherent representation of a whole scene (Duchowski, 2017). From a psychological and educational point of view, it is interesting to examine professionals' attentional behaviours during work activities and to understand how they analyse and interpret visual input.

Applied eye-tracking research has been conducted in many different professional disciplines to examine expertise-related differences in visual information processing.

The medical domain is one of the fields that has been investigated extensively as errors at the workplace can have serious consequences for patient safety. A study that caught media attention was titled "The invisible gorilla strikes again", in which Drew et al. (2013) showed that even experts are vulnerable to inattentional blindness. The results revealed that 83% of the radiologists who participated in the empirical study did not see a gorilla in a chest X-ray image when focussing on a task to detect lung nodules. Thus, eye tracking research can help identify challenges in professional practice (such as missing abnormalities under certain circumstances) and can point to inherent limitations of human attention and perception. Gaining a better understanding of these limits can provide valuable insights into how to design tasks and work activities to reduce mistakes at the workplace. For other professional domains, eye-tracking research also revealed valuable insights. In the domain of teaching, classroom management is a highly complex task, and teachers need to integrate a multifaceted set of sensory inputs to ensure a stimulating learning environment. Eye-tracking technology has been applied to identify differences in the perception of problematic classroom management scenes between expert teachers and novice teachers (Wolff et al., 2016) or of judgment accuracy (Seidel et al., 2021). Other vision-intensive professional domains in which perceptual processes were investigated with eye tracking are biology (Jarodzka et al., 2010), air traffic control (Van Meeuwen et al., 2014), music (Puurtinen, 2018), sport (Kishita et al., 2020), and visual art (Puppe et al., 2021).

21.3 Basic Understanding of Measuring Eye Movements

Four large classes of eye-tracking measures are described in a book by Holmqvist et al. (2011). These researchers differentiate between movement measures (address eye movements through space), position measures (address where a person is looking), numerosity measures (address countable eye-movement events), and latency measures (address duration from the onset of one event to the onset of another event). As each measure is further subdivided in the introductory book (e.g., movement direction measures, movement amplitude measures, movement duration measures), it is beyond the scope of this chapter to introduce all possible dependent variables. Readers are advised consulting Part III of the comprehensive guide for a detailed overview. Depending on the research question, particular measures can be selected.

Holmqvist et al. (2011) recommend using a theory-driven approach and operationalisation through traditions and paradigms to benefit and build upon an accepted experimental set-up and measures. The chosen theory needs to fit the domain of investigation (Brams et al., 2019). The better specified a theory is, the easier it is to derive predictions that can be empirically tested. Subsequently, another advantage is that the kind of data that is relevant to answer your question(s) and how to analyse it is clearly specified at the onset of an investigation (Holmqvist et al., 2011). Otherwise, a researcher might easily become overwhelmed by the multitude of different dependent measures for analysis.

To gain insight into visual perception processes, fixations, saccades, and smooth pursuit are important due to their functionality (Duchowski, 2017). Referring to the four classes of eye-tracking measures mentioned previously, fixations can either belong to position measures (e.g., fixation duration or dwell time) or numerosity measures (e.g., number of fixations or number of revisits), while saccades and smooth pursuit are categorised as movement measures.

Fixations are the most reported eye-tracking data. The eye remains relatively still for tens of milliseconds to several seconds while focused on a particular visual target to obtain information (Holmqvist et al., 2011). Often researchers are interested in specific regions or areas in stimulus material (such as a pathology in an X-ray image or an object in a scene). Most eye-tracking software allows these areas of interest (AOIs) to be defined, for instance, by drawing a frame around a particular object. Researchers decide about the number of AOIs and the respective size. Importantly, AOIs have consequences for the outcome of an experiment and increase the number of possible dependent measures. For example, researchers might want to know the number of fixations within an AOI (*fixation count*), when participants first fixate on an AOI (*time to first hit*), how long they fixated within an AOI (*dwell time*) or how often they return to the AOI (revisits) (Carter & Luke, 2020). A recent systematic review (Brams et al., 2019) found that experts across domains can selectively allocate attention toward important task-related information, which is consistent with the information-reduction hypothesis (Haider & Frensch, 1999). Thus, with practice, people learn to distinguish task-relevant from task-redundant information and therefore can ignore task-irrelevant information. Moreover, their review revealed that experts in the medical domain extract visual information from distal and para-foveal regions, which are areas that are more in the periphery of the retina compared to the centre of vision (foveal region) where we have full acuity and can see visual detail sharply (see Holmqvist et al., 2011; Rayner & Reingold, 2015). In other words, experts in the medical domain have demonstrated a more efficient global-local processing or an extended visual span (cf. holistic model of image processing by Kundel et al., 2007).

Saccade refers to a rapid motion of the eye from one fixation to another fixation, typically taking 30–80 milliseconds to complete. During most saccades, an individual is considered blind. *Smooth pursuit* requires a moving object that is followed with the eyes (such as a plane), and these eye movements are slower than saccades (Duchowski, 2017; Holmqvist et al., 2011). According to Carter and Luke (2020), fixations and saccades are the basic unit of data for most analyses. In Fig. 21.1, fixations and saccades are visualised.

To record these eye movements, a variety of commercially available eye trackers exist. The eye tracker needs to be well chosen as each system has specific properties that are more or less suitable for particular experiments. Trackers can be stationary or portable and they vary in their speed of data acquisition. The sampling rate of an eye tracker is measured in hertz (Hz) and varies between 25 Hz for the slowest system and 2000 Hz for high-speed machines (Holmqvist et al., 2011). The higher the sampling rate, the more data are collected. An eye tracker with a sampling rate of 50 Hz records the eye position as much as 50 times per second; such a system is

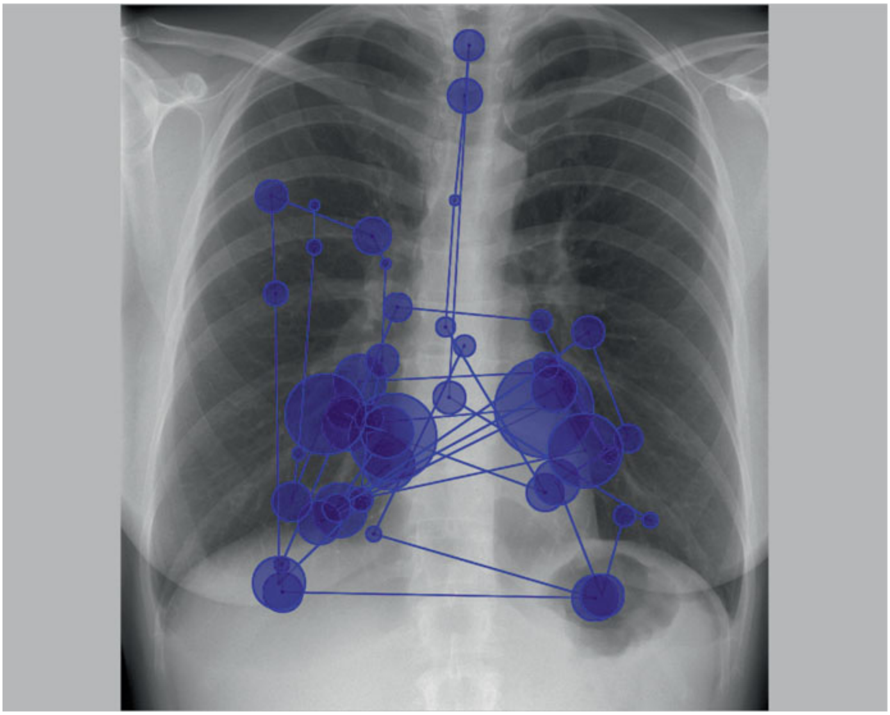

Fig. 21.1 Fixations and saccades in a posterior-anterior chest x-ray image. (*Note*: Circles represent fixations and the larger the circle, the longer the fixation duration. The lines between the fixations are the saccades)

usually appropriate for examining where a person looks. Mobile eye trackers, such as eye-tracking glasses, have these lower sampling rates, but their great advantage is that they are portable and can be utilised in many different settings, such as a classroom or a museum, increasing the ecological validity. The fastest commercial eye tracker has 2000 Hz and allows millisecond accuracy recordings (Carter & Luke, 2020). This type of system is stationary, and the head of participants is stabilised so that no head movement is possible. The advantage of such a system is the accuracy, and thus, the possibility of recording fast eye movements. Another stationary machine is the remote eye tracker. Typically, a camera and infrared source are integrated with the computer screen (below the stimulus area) monitoring the user's eyes from a certain distance. These systems often have a sampling rate of 250 Hz or 500 Hz and are less restrictive for participants compared to high-speed machines.

When the eye tracker is selected, the appropriate use is essential for collecting reliable and valid data. Orquin and Holmqvist (2018) discuss the validity of eye-tracking studies in detail; their article is recommended for further reading. Most modern eye trackers are video-based and use infrared light to produce a reflection on the cornea that is identified by the eye tracking software. Additionally, this software registers the pupil as the darkest point of the eye. The distance between the two centres of the areas in combination with the screen gives the exact position where a person is looking on the screen (Carter & Luke, 2020). As eyes are very different, individual calibration of participants is necessary. Calibration is the process whereby the geometric characteristics of a subject's eyes are estimated as the basis for a fully customised and accurate gaze point calculation. The smaller the difference between the true gaze position and the recorded gaze position the more accurate is the eye tracker. Therefore, researchers should strive for a maximum average deviation of 0.5°. Calibrating can be quite tricky because many factors can influence the quality of the calibration, such as mascara, droopy eyelids, downward eyelashes, glasses, and contact lenses (Holmqvist et al., 2011). Without a good calibration, no accurate eye movement measures can be expected, and therefore, no valuable findings are gained. Therefore, researchers should invest enough time and training so that they can make necessary adjustments. Another challenge in eye tracking research is that eye movements are influenced by a variety of visual and cognitive factors (Carter & Luke, 2020; Holmqvist et al., 2011). These factors can be top-down influences such as prior knowledge or task instruction. For instance, the influential study of Yarbus (1967) showed that eye movements differ tremendously depending on the provided instruction. Systematic tendencies, such as central bias, influence the measurement of eye movements (Tatler, 2007) as well as bottom-up processes. For instance, some visual aspects are more salient than other aspects due to colour, contrast, brightness, alignment, and therefore, attract attention (Itti & Koch, 2000). For a researcher, this finding means that setting up an eye tracking experiment requires an intense preparation phase, in which the stimulus material is carefully selected to control the above-mentioned factors and to answer the research questions of interest. Ideally, the researchers choose the variables they will analyse before the study is conducted (Carter & Luke, 2020).

21.4 Empirical Illustration

To illustrate the challenges and benefits of using eye-tracking technology in workplace learning, an empirical study in the medical domain is presented. The clinical diagnostic is a difficult and highly complex skill. Fundamental theoretical knowledge and experienced-based knowledge gained through clinical cases are required (Boshuizen et al., 2020; Lesgold et al., 1988). Croskerry (2009) proposed a model of diagnostic reasoning in which the dual-process theory is used to explain clinical judgment. The dual-process theory distinguishes two systems of decision making: System 1 (heuristic, intuitive) and System 2 (systematic, analytical). If salient features are recognised immediately, System 1 starts automatically. The recognised visual characteristics of an illness, injury, or combinations of salient traits trigger unconscious pattern recognition responses in System 1. If the illness or injury is unfamiliar or the visual information is ambiguous and there is uncertainty, System

2 processes start up a more systematic analysis based on reasoning and logic. Consequently, System 2 is slower than System 1, more costly in terms of resources, but less error-prone. The diversity and far-reaching relevance for effective and accurate medical care render professional learning and expertise in radiology an interesting subject of research (Norman et al., 2018).

The peculiarity of the perception-related component distinguishes radiology from other medical domains. Good perceptual and cognitive skills are essential when diagnosing. Perceptual skills encompass visual search, visual information processing, and visual discrimination and differentiation. Cognitive skills help develop rule-based schemas that guide perception, evaluation or interpretation, and decision-making (Nodine & Mello-Thoms, 2018). Studies have investigated how novices and experts differ when analysing complex visual material, identifying abnormalities and how these processes can be influenced through practice (Kok et al., 2012; Kundel et al., 2007; Manning et al., 2006; Nodine & Mello-Thoms, 2018). However, information about the process of expertise development between academic graduation and qualification as a specialist remains scarce. The aim of this study is to investigate expertise development among residents in radiology and to identify changes in their way of analysing and diagnosing medical X-ray images during their residency. The following hypotheses were addressed:

H1: With the advancement of further training as a specialist, residents' visual information processing during clinical reasoning improves.
H2: With the advancement of further training as a specialist, residents' diagnostic performance improves.

The underlying theoretical considerations for the present study derived from a rich body of prior research. An important basis for the operationalisation were the assumptions of the following four theories of expertise: Encapsulation (Boshuizen & Schmidt, 1992), the theory of long-term working memory (Ericsson & Kintsch, 1995), the information-reduction hypothesis (Haider & Frensch, 1999), and the holistic model of image perception (Kundel et al., 2007). Supportive findings in line with these theories were revealed in a meta-analysis (Gegenfurtner et al., 2011) that examined differences in expertise in various professional domains showing that a higher degree of expertise was associated with higher performance accuracy and a lower reaction time. Experts' rapid information processing was reflected in shorter fixation durations, a higher fixation count on task-relevant areas, and lower fixation count on task-redundant information. Moreover, it took experts less time to first fixate relevant information and they had longer saccades indicating their superiority in parafoveal processing and selective attention allocation. In addition, research has demonstrated that dwell time is related to the detection of abnormalities. A short dwell time is seen as indicator for fast visual processing and by revisiting the abnormal area again the diagnostic assumption can be verified (Bruno et al., 2015; Mallett et al., 2014).

21.5 Method

21.5.1 Design

A longitudinal, quasi-experimental study with three measurement points was designed to investigate expertise development among residents of a radiology department. The experiment consisted of two parts. In the first part of the experiment, residents had to silently analyse and diagnose posterior-anterior, chest X-ray images of real patient cases (see Fig. 21.1) while their eye movements were recorded. In the second part of the experiment, residents had to think aloud while analysing posterior-anterior, chest X-ray images of real patient cases and while their eye movements were tracked. As the focus of this chapter is eye tracking, only the first part of the experiment is reported here. The following dependent variables were collected with an eye tracker to capture residents' visual processing during clinical reasoning: fixation count, fixation duration, fixation count in AOI, time-to-first hit AOI, revisits to AOI, dwell time in AOI, and saccadic amplitude. Self-assessment of radiological skills, confidence on presence or absence of abnormality and accuracy of diagnostic decisions were the dependent variables to examine residents' diagnostic performance. Table 21.1 presents the hypotheses, dependent variables, and predictions for the three measurement points.

Table 21.1 Overview of hypotheses, dependent variables, and predictions

Hypotheses	Dependent variables	Predictions
H1: With the advancement of further training as a specialist, residents' visual information processing during clinical reasoning improves over the three measurement points	Fixation count	$T1 > T2 > T3$
	Fixation duration	$T1 > T2 > T3$
	Fixation count in AOI	$T1 < T2 < T3$
	Time-to-first hit AOI	$T1 > T2 > T3$
	Revisits to AOI	$T1 < T2 < T3$
	Dwell time in AOI	$T1 > T2 > T3$
	Saccadic amplitude	$T1 < T2 < T3$
H2: With the advancement of further training as a specialist, residents' diagnostic performance improves over the three measurement points	Self-assessment radiological skills	$T1 < T2 < T3$
	Accuracy of diagnostic decisions:	
	Correct	$T1 < T2 < T3$
	Incorrect	$T1 > T2 > T3$
	Mixed	$T1 > T2 > T3$
	Confidence on abnormality	$T1 = T2 = T3$

Note: T1 = first measurement point, T2 = second measurement point, T3 = third measurement point

21.5.2 Sample

Residents at one radiology department were recruited to participate in the study: at the first measurement point (T1), $N_1 = 15$; at the second measurement point (T2), $N_2 = 15$; and at the third measurement point (T3), $N_3 = 14$. Between each measurement point, there was a time-lag of approximately 6 months. Since the residents started their training at different points in time and thus completed it over time, not all the 15 participants at the first measurement point could be recorded at all three measurement times. Eight residents participated in all three measurement points and are the focus of this chapter. Due to the Coronavirus 2019 (COVID 19) pandemic, the data collection ended earlier, and four residents could not take part at the third measurement point as planned. The eight participants (3 females and 5 males) were aged between 26 and 33 years ($M = 28.88$, $SD = 2.42$) at T1. At the beginning of the longitudinal study, the participants had an average training time of 29.27 months ($SD = 18.96$) and worked on average 44.69 h a week ($SD = 3.79$). Table 21.2 shows their training time at the three measurement points. All participants reported that they conducted further training in addition to professional practice, such as reading specialist literature, attending advanced training or participating in conferences. Three of the eight residents wore glasses or soft contact lenses during the recording.

21.5.3 Apparatus and Stimulus Material

A remote eye tracker (SMI GmbH, Teltow, Germany) with a sampling rate of 250 Hz and a spatial resolution (dispersion) of 1.0° visual angle, including the manufactured software SMI Experiment Center™, SMI iViewX™ and SMI BeGaze™, was employed. SMI Experiment Center™ was selected to create the

Table 21.2 Overview of the longitudinal sample

		T1	T2	T3
Participant	Age[a]	Training time[b]	Training time[b]	Training time[b]
P01	33	33.79	43.17	50.10
P02	28	22.26	31.20	37.52
P03	31	46.07	55.13	60.52
P04	26	0.098	08.18	14.62
P05	30	44.13	52.24	58.90
P06	29	36.33	44.25	51.61
P07	28	48.33	56.02	62.59
P08	26	3.23	11.97	18.60

Note: T1 = first measurement point, T2 = second measurement point, T3 = third measurement point
[a]At the first measurement point; [b]Training time in months

experiment; it functioned as the planning and experiment execution environment. SMI iView X™ was required for the gaze-tracking data acquisition, and SMI BeGaze™ was selected for the gaze-tracking data analysis (e.g., editing AOIs). The stimulus material was presented on a screen that had a display resolution of 1680×1050 pixels. A chinrest restricted the head movements of the participants and increased the accuracy of the eye movement data.

At T1, the participants had to analyse 24 posterior-anterior, chest X-ray images in the first part of the experiment, of which eight images were normal, eight images showed one abnormality, and eight images had two abnormalities. Due to the feasibility of the experiment, the number of posterior-anterior, chest X-ray images was reduced to 12 at T2 and T3 with four normal images, four images with one abnormality and four images showing two abnormalities. At each measurement point, the X-ray images were similar but not identical.

21.5.4 Instrument

At each measurement point, the participants completed a demographic questionnaire that collected data such as age, experience, training time, perceived alertness, well-being, and radiological skills. The participants were asked to self-assess their radiological skills (analytical ability, discrimination ability, and diagnostic performance) on a 7-point Likert scale. During the experiment, each chest X-ray image was followed by three questions: (a) Did you have enough time to analyse the chest X-ray? (b) Is there a pathological finding? and (c) What is your diagnosis? The first question was assessed with a 5-point Likert scale with options from “insufficient” (1) to “enough time” (5). The second question addressed participants’ confidence and was also assessed with a 5-point Likert scale ranging from “definitely not pathological” (1) to “definitely pathological” (5). The residents had to type their diagnoses into a text field on the screen.

21.5.5 Procedure

Data collection was supported by the radiology department, and residents were encouraged to participate in the study. Participation was on a voluntary basis, and anonymity was ensured. Participants gave their written informed consent at each measurement point. For residents’ convenience, the experiment was conducted in a room at the hospital, allowing residents to participate during working hours. The experimental set-up at all three measurement points was identical. The participants received information about the experiment on paper and on screen. To familiarise them with the experimental process, the residents were shown an example chest X-ray first followed by three questions. If participants had no further questions, instructions about how to behave during the experiment (e.g., to sit still and remain in the same position) were provided. To ensure a precise recording of the eye movements, a calibration was performed to ensure values below or near .50°. Data with values above 1.0° were excluded, which was the case for one participant. Before each chest X-ray image was displayed, information about the age and gender of the patient was provided. The stimulus material was presented for a maximum of 30 s. Two considerations led to this speed condition. On the one hand, experts are known for making fast decisions (Kundel & Nodine, 1975) and therefore, it was expected that the time restriction would reveal performance differences among the residents. On the other hand, the overall length of the experiment was restricted in time making it more feasible for residents to participate during working hours. If participants needed less time, they could click on the space bar and continue the experiment. It was not possible to return to a previous image nor could participants zoom in or out on an image. After the stimulus presentation, the three questions about time, confidence and diagnosis followed, and participants used a mouse and keyboard. In total, the whole experiment lasted 60.71 min ($SD = 17.04$) at T1, 30.27 min ($SD = 9.88$) at T2 and 23.75 min ($SD = 9.00$) at T3.

21.5.6 Analysis

To analyse the data, different methods were utilised. First, the diagnoses residents provided in open answer format were categorised to statistically analyse them. The correctness of the diagnoses was assessed with the help of an experienced radiologist. It was important to categorise the diagnoses as correct or incorrect. Furthermore, a fine-grained assessment was performed to distinguish among correct-negative, correct-positive, false-negative, false-positive (Kim & Mansfield, 2014; Šimundić, 2009) and mixed diagnoses (refer to Table 21.3).

The software R (R Core Team, 2019) was applied to statistically analyse the data. Some extra packages were installed: *psych* (Revelle, 2015), *car* (Fox et al., 2020) *lme4* (Bates et al., 2020), *lmerTest* (Kuznetsova et al., 2020), *rmarkdown* (Xie et al., 2020), *ez* (Lawrence, 2016) and *sjstats* (Lüdecke, 2019). Descriptive statistics were calculated for the data of the demographic questionnaire, diagnoses, and eye movements. An analysis of variance (ANOVA) with repeated measures was performed to compare the means for the dependent variables self-assessment of radiological skills, accuracy of diagnostic decisions, and confidence on presence or absence of abnormality over the three measurement points. In case sphericity was violated, the Greenhouse-Geiser correction procedure was applied to correct the degrees of freedom. When the F-values were significant, post-hoc pairwise t-tests were executed (Luhmann, 2015).

Since eye-tracking data are hierarchically structured data with several levels, which can be assigned to a higher-level group, a multi-level analysis was chosen for the statistical investigation of the hypotheses. Multi-level models are also known as mixed models. The term “multi-level” indicates the distinction of different levels of analysis in nested data structures. In longitudinal studies, individual data points

Table 21.3 Coding scheme for the analysis of diagnoses

Value	Description	Possible for X-rays with	Performance classification	Type of error
1	Correct-negative	0 pathology	Correct diagnose	
2	Correct-positive	1 pathology	Correct diagnose	
3	False-negative	1 pathology	Incorrect diagnose	Underreading
4	False-positive	0 pathology	Incorrect diagnose	Overreading
5	Correct-positive + false-negative	2 abnormalities	Mixed diagnose	
6	Correct-positive + false-positive	1 pathology	Mixed diagnose	Overreading
7	False-negative + false-positive	1 pathology	Incorrect diagnose	
8	2x correct-positive + false-positive	2 pathologies	Mixed diagnose	Overreading
9	2x correct-positive	2 pathologies	Correct diagnose	
10	2x false-negative	2 pathologies	Incorrect diagnose	Underreading
11	Correct-positive + false-negative + false-positive	2 pathologies	Mixed diagnose	
12	2x false-negative + false-positive	2 pathologies	Incorrect diagnose	Underreading

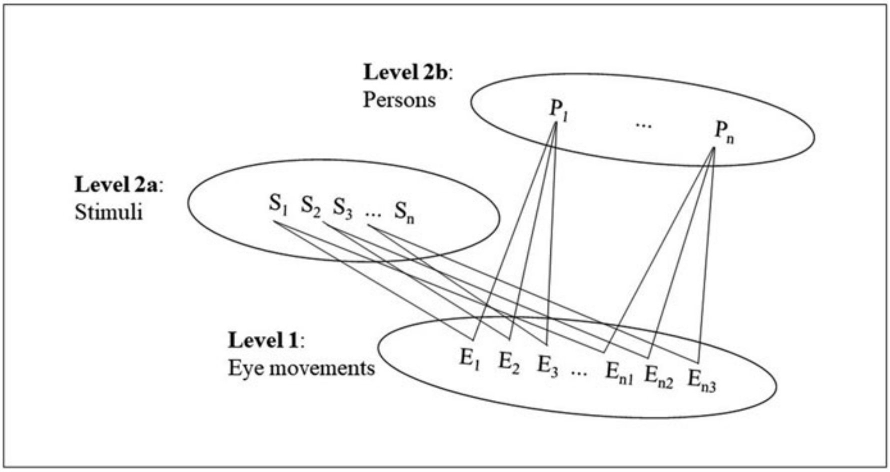

Fig. 21.2 Exemplary graphical visualisation of hierarchical data structure

are nested within persons. For further reading about analysing longitudinal studies with linear mixed models, the interested reader is referred to the recent articles by Hilbert et al. (2019) and Lindl et al. (2020). Figure 21.2 visualises how eye movements are nested in stimulus material and stimulus material is nested in persons. There were several interdependent values per participant, and these form a hierarchy level: fixation count, fixation duration, fixation count in AOI, time-to-first hit AOI, revisits to AOI, dwell time in AOI, and saccadic amplitude. It was expected that training time and measurement point influenced these dependent variables. Moreover, the number of pathologies displayed in a chest X-ray image can be influential. Therefore, a linear-mixed model on the individual level with measurement point and pathology as predictors was constructed for each variable.

21.6 Results

21.6.1 Visual Information Processing During Clinical Reasoning: Residents' Eye Movements

The main and interaction effects of measurement point and number of pathologies on the dependent variables are displayed in Table 21.4. The intercept describes each dependent variable for T1 without consideration of pathology and is considered a point of reference.

The fixation count in the image was predicted significantly by the time of measurement. Both at T2, $b = -24.73$, $t(211) = -5.57$, $p < .001$, and at T3, $b = -31.51$, $t(211) = -3.82$, $p < .001$, the fixation count compared to the intercept and without taking into account the number of pathologies decreased assuming a lower fixation count on task-redundant information with the advancement of further training. A small significant interaction effect was obtained for T2 and the number of pathologies, $b = -7.18$, $t(211) = -2.11$, $p = .035$, indicating that the presence or absence of pathologies matters. The fixation count was lower for images without pathology. There was no significant main effect of the number of pathologies, $b = 1.41$, $t(211) = 0.59$, $p = .560$, on the fixation count in the image.

For the fixation duration, only the measurement time was a predictor. Compared to T1 (intercept), the fixation duration at T2, $b = -28.52$, $t(211) = -2.98$, $p = .003$, and at T3, $b = -42.80$, $t(211) = -2.41$, $p = .017$, decreased significantly indicating faster information processing with the advancement of further training. There were no significant interaction effects, and the number of pathologies was not a predictor.

At T2, the fixation count in AOI increased significantly, $b = 16.53$, $t(155) = 2.62$, $p = .003$, assuming that residents fixated more task-relevant areas compared to T1. In addition, a significant effect was observed between T2 and the number of pathologies, $b = -16.75$, $t(155) = -4.21$, $p < .001$, indicating a higher fixation count in AOI when pathologies are present. At T3, there was no significant change in the fixation count in AOI compared to T1.

For the time to first hit AOI, no significant main and interaction effects were revealed. However, the number of revisits was predicted by the number of pathologies, $b = 5.58$, $t(112) = 5.04$, $p < .001$, showing that residents had a higher number of revisits when pathologies were present. Moreover, a small interaction effect

Table 21.4 Results of the mixed-linear models of the eye movements with measurement point and pathology as predictors considering the individual data structure

	b[a]	*SE*	*df*	*t*	*p*
Fixation count					
Intercept[b]	94.64	7.84	12.35	12.07***	<.001
T2	−24.73	4.44	211.52	−5.57***	<.001
T3	−31.51	8.27	211.23	−3.82***	<.001
Pathology	1.41	2.40	211.08	.59	.56
T2 x Pathology	−7.18	3.40	211.08	−2.11*	.035
T3 x Pathology	.13	5.37	211.08	.02	.98
Fixation duration					
Intercept[b]	248.29	13.85	18.4610	17.93***	<.001
T2	−28.52	9.56	211.90	−2.98**	.003
T3	−42.80	17.77	211.47	−2.41*	.017
Pathology	−.35	5.18	211.15	−.07	.95
T2 x Pathology	−2.58	7.32	211.15	−.35	.72
T3 x Pathology	7.87	11.58	211.15	.68	.50
Fixation count in AOI					
Intercept[b]	−9.40	5.18	106.92	−1.81	.07
T2	16.53	6.30	155.01	2.62**	.01
T3	.33	6.31	155.13	.05	.96
Pathology	16.79	2.81	154.71	5.97***	<.001
T2 x Pathology	−16.75	3.98	154.71	−4.21***	<.001
T3 x Pathology	−1.68	3.98	154.71	−.42	.67
Time to first hit AOI					
Intercept[b]	4123.29	1787.64	135.35	2.36*	.02
T2	2741.25	2325.63	142.64	1.18	.24
T3	2423.29	2423.29	142.89	1.08	.28
Pathology	−1735.86	1001.00	142.33	−1.73	.09
T2 x Pathology	−573.93	1430.20	142.23	−.40	.69
T3 x Pathology	−1013.60	1391.64	142.38	−.73	.47
Revisits to AOI					
Intercept[b]	1.67	2.21	79.91	−.76	.45
T2	6.11	2.51	111.93	2.44*	.02
T3	2.43	3.33	114.47	.73	.47
Pathology	5.58	1.11	111.55	5.04***	<.001
T2 x Pathology	−5.99	1.55	111.47	−3.86***	<.001
T3 x Pathology	−2.13	2.01	111.57	−1.06	.29
Dwell time in AOI					
Intercept[b]	−1879.10	1793.99	122.44	−1.05	.30
T2	3220.81	2145.84	144.97	1.50	.14
T3	−1409.40	2105.95	145.31	−.67	.50
Pathology	4454.47	958.44	145.10	4.65***	<.001
T2 x Pathology	−4188.40	1331.99	144.91	−3.14*	.01
T3 x Pathology	−240.84	1320.41	144.94	−.18	.86

(continued)

Table 21.4 (continued)

	b^{a}	*SE*	*df*	*t*	*p*
Saccadic amplitude					
Intercept[b]	3.50	.39	27.42	8.87***	<.001
T2	.02	.31	212.36	4.83***	<.001
T3	1.52	.58	211.71	2.02*	.04
Pathology	.34	.17	211.23	2.02*	.04
T2 x Pathology	.05	.24	211.23	.19	.85
T3 x Pathology	−.20	.38	211.23	−.52	.60

Note: [a]Unstandardised regression coefficient; [b]The constant, which indicates the first measurement point and does not consider the number of pathologies in X-ray image; * = p < .05, ** = p < .01, *** = p < .001

between T2 and the number of pathologies was revealed, $b = -5.99$, $t(112) = -3.86$, $p < .001$. The number of revisits increased significantly at T2, $b = 6.11$, $t(112) = 2.44$, $p = .02$, indicating that residents focused more on task-relevant information and verified their assumptions. At T3, no significant main and interaction effects were reported.

The dwell time in AOI was not significantly predicted by measurement point. A significant interaction effect was observed between T2 and number of pathologies, $b = -4188.40$, $t(145) = -3.14$, $p = .01$.

The saccadic amplitude differed significantly from the reference value at both T2 and T3 assuming a more holistic viewing with the advancement of further training. At T2, the saccadic amplitude increased by a value of $b = 0.02$, $t(212) = 4.83$, $p < .001$. At T3, the saccadic amplitude differed from the intercept with an increase of $b = 1.52$, $t(212) = 2.02$, $p = .04$. The saccadic amplitude of the participants depended on the number of pathologies in the image. A significant main effect was observed: $b = -0.34$, $t(211) = 2.012$, $p = .04$. No interaction effects were reported.

21.6.2 Residents' Diagnostic Performance

Self-assessment of radiological skills. The ANOVA revealed no differences across the three measurement points for analytical ability, discrimination, and diagnostic performance (refer to Table 21.5).

Confidence on presence or absence of abnormality. At the three measurement points, the confidence of residents was similar: $M_1 = 3.71$ ($SD = 1.58$), $M_2 = 3.88$ ($SD = 1.38$), and $M_3 = 3.75$ ($SD = 1.41$). The ANOVA revealed no significant difference: $F(2, 14) = .45$, $p = .64$ and $\eta_P{}^2 = .06$. Across all measurement times, the residents estimated the presence or absence of at least one pathology in a chest X-ray image 31 times as "definitely not pathological",

Table 21.5 Means, standard deviations, and f statistics of self-assessment radiological skills

	T1	T2	T3		
	M (SD)	*M (SD)*	*M (SD)*	*F (2, 14)*	*p*
Analytical ability	5.62 (.92)	5.88 (.35)	5.88 (.64)	.78	.48
Discrimination	5.75 (.71)	6.25 (.46)	6.00 (.00)	1.90	.19
Diagnostic performance[a]	5.25 (1.39)	6.12 (.35)	5.75 (.46)	.59	.18

Note: T1 = first measurement point, T2 = second measurement point, T3 = third measurement point
[a]Greenhouse Geisser correction was applied, as sphericity was violated

Table 21.6 Absolute and relative frequencies of diagnoses across the time

	T1		T2		T3	
Frequencies	*n*	%	*n*	%	*n*	%
Correct diagnoses	30	15.43	41	37.61	32	28.31
Correct-positive	17	14.41	22	20.18	15	13.27
Correct-negative	13	11.02	19	17.43	17	15.04
Incorrect diagnoses	25	21.18	13	11.93	16	14.16
False-positive	19	16.10	13	11.93	14	12.39
False-negative	6	5.08	0	.00	2	1.77
Mixed diagnoses	23	19.41	26	23.85	31	27.43
Type of error						
Overreading	34	28.81	26	23.85	25	22.12
Underreading	6	5.08	3	2.75	9	7.96

Note: T1 = first measurement point, T2 = second measurement point, T3 = third measurement point

44 times as “probably not pathological”, 22 times as “uncertain”, 49 times as “probably pathological”, and 140 times as “definitely pathological”.

Accuracy of diagnostic decisions. Table 21.6 shows the absolute and relative frequencies of diagnoses as well as the type of error across the three measurements. A repeated analysis of variance found no significant differences in the mean values for correct positive: $F(2, 14) = .69$, $p = .52$, $\eta_P^2 = .07$; correct negative: $F(2, 14) = .86$, $p = .44$, $\eta_P^2 = .12$; false positive: $F(2, 14) = .96$, $p = .41$, $\eta_P^2 = .12$; false negative: $F(2, 14) = 2.88$, $p = .09$, $\eta_P^2 = .29$; and mixed diagnoses: $F(2, 14) = 1.44$, $p = .27$, $\eta_P^2 = .17$. There were also no significant results of the ANOVA for the error overreading, $F(2, 14) = .79$, $p = .47$, $\eta_P^2 = .10$, and the error underreading, $F(2,14) = 1.15$, $p = .35$, $\eta_P^2 = .14$.

21.7 Conclusion and Practical Implication

The empirical illustration shows how eye tracking was used to gain more insights into expertise development among residents in radiology. More data than presented here was collected to identify changes in residents’ way of analysing and diagnosing medical X-ray images. We expected that with the advancement of further training as

Table 21.7 Overview of supported, partly supported, and not supported expectations

Hypotheses	Dependent variables	Predictions	Expectations
H1: With the advancement of further training as a specialist, residents' visual information processing during clinical reasoning improves over the three measurement points	Fixation count	T1 > T2 > T3	Supported
	Fixation duration	T1 > T2 > T3	Partly supported
	Fixation count in AOI	T1 < T2 < T3	Not supported
	Time-to-first hit AOI	T1 > T2 > T3	Not supported
	Revisits to AOI	T1 < T2 < T3	Partly supported
	Dwell time in AOI	T1 > T2 > T3	Supported
	Saccadic amplitude	T1 < T2 < T3	Partly supported
H2: With the advancement of further training as a specialist, residents' diagnostic performance improves over the three measurement points	Self-assessment radiological skills	T1 < T2 < T3	Not supported
	Accuracy of diagnostic decisions:		
	Correct	T1 < T2 < T3	Not supported
	Incorrect	T1 > T2 > T3	Not supported
	Mixed	T1 > T2 > T3	Not supported
	Confidence on abnormality	T1 = T2 = T3	Supported

Note: T1 = first measurement point, T2 = second measurement point, T3 = third measurement point

a specialist, residents' visual information processing during clinical reasoning improves (Hypothesis 1) as well as their diagnostic performance (Hypothesis 2). Table 21.7 displays which expectations were supported or not. The results of the linear mixed models were consistent with our theoretically grounded expectations. For instance, the number of fixations and dwell time decreased, which indicates that faster information processing occurs with increasing experience. Moreover, the increase in saccadic amplitude at T2 and T3 indicates a more holistic viewing pattern and therefore supports the global-focal theory (Kundel et al., 2007). The diagnostic performance of the residents did not change during the longitudinal study. Based on a descriptive level, the residents tended to make more overreading errors rather than underreading errors, which indicates an over-caution. Contrary to professional practice, residents had less information about the patients, and it was not possible to magnify the images to perform a more comprehensive review, which might explain these finding. Data analyses are still ongoing, and a more detailed examination of the visual and cognitive processes is required. Several challenges (such as the different sizes of the AOIs and various kinds of pathologies) that need a critical re-examination of the data material have also been identified. In hindsight, the study was complex, and more standardisation would have simplified the data analyses.

Eye tracking is a powerful tool that can be applied to a wide variety of research questions across many different professional fields. Although it is promising that eye tracking technology is more affordable and accessible, researchers should be aware that using eye tracking is a challenging endeavour. There is the risk of being overwhelmed by data so the experiment should be kept as simple as possible. The application should be well-considered and not employed hastily despite the popularity of the method. We still need to learn more about how cognitive processes can be deduced from eye movements and be aware that 'the observed set of eye movements reflects the combined effects of many ongoing (cognitive) processes'. (Kok & Jarodzka, 2017a, p. 116). Therefore, references to eye-tracking guides and different studies were provided for researchers interested in learning more about eye tracking. Kok & Jarodzka (2017a) recommend a theory-driven approach, a sound study design, and methodological triangulation to draw convincing conclusions from eye tracking research. Only if eye tracking is employed correctly by careful and deliberate users, can researchers reveal findings that are meaningful for professional practice (Carter & Luke, 2020).

For researchers who are interested in investigating how individuals process information during learning and professional development eye tracking offers many possibilities. Think of research investigating visual expertise to gain a better understanding of why experts can quickly and accurately interpret complex visual material and scenes (e.g., medical images, classroom actions and interactions, or flight radars). Eye tracking research can also be used to examine learning processes and attentional strategies. Subsequently, these insights can help practitioners to improve (instructional) materials, adjust work procedures or design new teaching activities. Recent studies in different professional domains tackled challenging research questions, for instance, to find out more about expert teacher priorities (McIntyre et al., 2019), effects of fatigue from overnight shifts (Hanna et al., 2018), or recognition in crime scene investigations (Watalingam et al., 2017). These few examples are meant to point to various research possibilities that can be explored with the help of eye tracking. Many more open questions are waiting to be answered.

Conflict of Interest I have no known conflict of interest to disclose.

Acknowledgments I would like to thank my cooperation partners at the radiology department for their support. A big thank you goes to the residents for their willingness to participate in the study. I would also like to thank the master students who worked with me on this challenging project.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2020). *lme4: Linear mixed-effects models using "Eigen" and S4* (R package version 1.1-23). [Computer software].

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16*(2), 153–184. https://doi.org/10.1016/0364-0213(92)90022-M

Boshuizen, H. P. A., Strasser, J., & Gruber, H. (2020). Knowledge restructuring through case processing: The key to generalise expertise development theory across domains? *Educational Research Review, 29*, 100310. https://doi.org/10.1016/j.edurev.2020.100310

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., & Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin, 145*(10), 980–1027. https://doi.org/10.1037/bul0000207

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *Radiographics, 35*(6), 1668–1676. https://doi.org/10.1148/rg.2015150023

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology, 155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022–1028. https://doi.org/10.1097/ACM.0b013e3181ace703

Drew, T., Vo, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science, 24*(9), 1848–1853. https://doi.org/10.1177/0956797613479386

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-57883-5

Ericsson, K. A. (2018). Capturing expert thought with protocol analysis: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 192–212). Cambridge University Press. https://doi.org/10.1017/9781316480748.012

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Fox, J., Weisberg, S., & Price, B. (2020). *car: Companion to applied regression* (R package version 3.0–8). [Computer software].

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied, 5*(2), 129–151. https://doi.org/10.1037/1076-898X.5.2.129

Hanna, T. N., Zygmont, M. E., Peterson, R., Theriot, D., Shekhani, H., Johnson, J.-O., & Krupinski, E. A. (2018). The effects of fatigue from overnight shifts on radiology search patterns and diagnostic performance. *Journal of the American College of Radiology, 15*(12), 1709–1716. https://doi.org/10.1016/j.jacr.2017.12.019

Harteis, C., Kok, E., & Jarodzka, H. (2018). The journey to proficiency: Exploring new objective methodologies to capture the process of learning and professional development (editorial to Special Issue). *Frontline Learning Research, 6*(3), 1–5. https://doi.org/10.14786/flr.v6i3.435

Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM – Testing, Psychometrics, Methodology in Applied Psychology, 26*, 101–119. https://doi.org/10.4473/TPM26.1.6

Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms and measures* (2nd ed.). Lund Eye-Tracking Research Institute.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning & Instruction, 20*, 145–154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1), 1–18. https://doi.org/10.16910/jemr.10.1.3

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom (editorial to Special Issue). *Educational Psychology Review, 33*, 1–10. https://doi.org/10.1007/s10648-020-09565-7

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kim, Y. W., & Mansfield, L. T. (2014). Fool me twice: Delayed diagnoses in radiology with emphasis on perpetuated errors. *American Journal of Roentgenology, 202*(3), 465–470. https://doi.org/10.2214/AJR.13.11493

Kishita, Y., Ueda, H., & Kashino, M. (2020). Eye and head movements of elite baseball players in real batting. *Frontiers in Sports and Active Living, 2*, Article 3. https://doi.org/10.3389/fspor.2020.00003

Kok, E. M., & Jarodzka, H. (2017a). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education, 51*, 114–122. https://doi.org/10.1111/medu.13066

Kok, E. M., & Jarodzka, H. (2017b). Beyond your very eyes: Eye movements are necessary, not sufficient. *Medical Education, 51*, 1189–1190. https://doi.org/10.1111/medu.13384

Kok, E. M., De Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2012). Looking in the same manner but seeing it differently: Bottom-up and expertise effects in radiology. *Applied Cognitive Psychology, 26*(6), 854–862. https://doi.org/10.1002/acp.2886

Kok, E. M., Aizenman, A. M., Võ, M. L.-H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision, 17*(2), 1–11. https://doi.org/10.1167/17.12.2

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527–532. https://doi.org/10.1148/116.3.527

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242*(2), 396–402. https://doi.org/10.1148/radiol.2422051997

Kuznetsova, A., Brockhoff, P. B., & Bojesen Christensen, R. H. (2020). *lmerTest: Tests in linear mixed effects models* (R package version 3.1-2). [Computer software]. https://github.com/runehaubo/lmerTestR

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial experiments* (R package version 4.4-0). [Computer software]. http://github.com/mike-lawrence/ez

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Lawrence Erlbaum.

Lindl, A., Krauss, S., Schilcher, A., & Hilbert, S. (2020). Statistical methods in transdisciplinary educational research. *Frontiers in Education, 5*, 97. https://doi.org/10.3389/feduc.2020.00097

Lüdecke, D. (2019). *jstats: Collection of convenient functions for common statistical computations* (R package version 0.17.6.). [Computer software].

Luhmann, M. (2015). *R für Einsteiger: Einführung in die Statistiksoftware für die Sozialwissenschaften: mit Online-Material* (4th ed.). Beltz.

Mallett, S., Phillips, P., Fanshawe, T. R., Helbren, E., Boone, D., Gale, A., Taylor, S. A., Manning, D., Altman, D. G., & Halligan, S. (2014). Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: Visual perception of experienced and inexperienced readers. *Radiology, 273*(3), 783–792. https://doi.org/10.1148/radiol.14132896

Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography, 12*(2), 134–142. https://doi.org/10.1016/j.radi.2005.02.003

McIntyre, N. A., Jarodzka, H., & Klassen, R. M. (2019). Capturing teacher priorities: Using real-world eye-tracking to investigate expert teacher priorities across two cultures. *Learning & Instruction, 60*, 215–224. https://doi.org/10.1016/j.learninstruc.2017.12.003

Nodine, C. F., & Mello-Thoms, C. (2018). Acquiring expertise in radiologic image interpretation. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (2nd ed., pp. 167–187). Cambridge University Press. https://doi.org/10.1017/9781108163781

Norman, G. R., Grierson, L. E. M., Sherbino, J., Hamstra, S. J., Schmidt, H. G., & Mamede, S. (2018). Expertise in medicine and surgery. In K. A. Ericsson, R. R. Hoffmann, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 331–355). Cambridge University Press.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods, 50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Puppe, L., Jossberger, H., & Gruber, H. (2021). Creation processes of professional artists and art students in sculpting. *Empirical Studies of the Arts, 39*(2), 171–193. https://doi.org/10.1177/0276237420942716

Puurtinen, M. (2018). Learning on the job: Rethinks and realizations about eye tracking in music-reading studies. *Frontline Learning Research, 6*(3), 148–161. https://doi.org/10.14786/flr.v6i3.380

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences, 1*, 107–112. https://doi.org/10.1016/j.cobeha.2014.10.008

Revelle, W. (2015). *psych: Procedures for psychological, psychometric, and personality research* (R package version 1.5.1). [Computer software].

Seidel, T., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2021). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychological Review, 33*, 69–89. https://doi.org/10.1007/s10648-020-09532-2

Šimundić, A.-M. (2009). Measures of diagnostic accuracy: Basic definitions. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, 19*(4), 203–211.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4. https://doi.org/10.1167/7.14.4

Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction, 32*, 10–21. https://doi.org/10.1016/j.learninstruc.2014.01.004

Wade, N., & Tatler, B. W. (2005). *The moving tablet of the eye: The origins of modern eye movement research*. Oxford University Press.

Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International, 280*, 64–80. https://doi.org/10.1016/j.forsciint.2017.08.012

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*, 243–265. https://doi.org/10.1007/s11251-016-9367-z

Xie, Y., Allaire, J. J., & Grolemund, G. (2020). *rmarkdown: Dynamic documents for R* (R package version 2.3.3.). [Computer software]. https://rmarkdown.rstudio.com

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In A. L. Yarbus (Ed.), *Eye movements and vision* (pp. 171–211). Plenum Press.

Helen Jossberger (PhD) is an Assistant Professor at the Faculty of Human Sciences, Department Educational Science, at the University of Regensburg, Germany. She teaches courses on learning and professional development, expertise research, instructional design, evaluation research, development and socialisation, and methodology. Her research focuses on self-directed and self-regulated learning in vocational education, professional learning, workplace learning and (visual) expertise development. Currently, she is assistant editor of Educational Research Review. She received her PhD at the Open Universiteit on self-regulated learning in vocational education. She obtained a Master degree in Cognitive Psychology at Maastricht University.

Chapter 22
Seeing Workplaces from a Social Network Analysis (SNA) Approach

Tuire Palonen

Abstract The aim of this chapter is to introduce Social Network Analysis (SNA) in the context of work organisations. The examples and terms are expected to show how SNA contributes to workplace learning research by helping us understand what happens, under the surface of the workplaces, in the formal or hidden informal structures. A social network is defined as a set of interacting social entities, such as individuals or groups, and the relationships among them. From a methodological point of view, this chapter targets basic-level concepts and quantitative analysis, not a qualitative approach or the newest or most advanced techniques. The framework focuses on terms of cohesion, structural equivalence, and personal (egocentric) networks. The emphasis is on the cohesion approach, which refers to measures of connectedness and togetherness within a network. Often, it means emphasising the number of relationships among colleagues and how these relationships are distributed to create structures and borders within workplaces. Some empirical examples from a work organisation are presented at the community, organisational, and individual levels.

Keywords Workplace learning · Work organisations · Social network analysis

22.1 Introduction

Social network research is a multidisciplinary research area. Consequently, a wide range of approaches to analysing network data has been developed for many domains, contexts, and purposes. In this chapter, seen from an organisational perspective, a social network is defined as a set of actors and the relationships that hold them together through resource exchange, such as expert advice, information, or social support. The essential aim behind the methodology from this point of view is to reveal the importance of repeated exchange relations that form the basis of both

T. Palonen (✉)
Teacher Education Department, University of Turku, Turku, Finland
e-mail: tuipalo@utu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_22

dyadic (between individuals) and structural (in the network) embeddedness. The continuous flow of these resources creates structures that are subsequently studied.

In this chapter, the main principles behind social network theory and methodology are described using both cohesion and structural equivalence perspectives. The cohesion approach studies how social relationships are distributed in networks. Structural equivalence refers to the extent to which two actors are connected to the same others, supposing that structurally equivalent actors might also be similar in other ways. Second, in addition to the whole network perspective, an egocentric approach is referred to, where the network is examined from the perspective of one person (ego) and his or her relationships with other people (alters). Third, the main principles regarding data gathering and statistical analysis are discussed. Fourth, an empirical case from one work organisation is introduced. Finally, some criticism, ethical concerns, and future possibilities for SNA in the organisational context are provided.

In the organisational approach, SNA connects social context to individual capacity by describing how people create, maintain, cultivate, and activate their personal ties to other humans for various aims (Brown & Duguid, 2001; Hakkarainen et al., 2004). For example, studies on expertise from a social network perspective do not merely rely on the workplaces and institutes but consider expertise to be cultivated and covered as part of social networks. Experts are shown to nurture and profile their own expertise by reactivating and strengthening relevant links depending on what kind of work they are doing (McCarty, 2002). Further, social network studies stress the importance of cross-boundary analyses of workplaces' networks and even experts' past relationships with their former networks (Nardi et al., 2000). It is also important to study network positions (i.e., What are workers' positions towards important actors in the field?).

The chapter addresses the issue of interdisciplinary understanding of social network research. More closely, organizational approach is highlighted here with an example from one organization. The aim is to introduce SNA as a systematic effort to examine the structure of social relationships at workplaces; to uncover the informal connections between people, and to examine and visualize how communications flow through an organization. This is done by using traditional network analytic methods, where participants of the study are asked to rate their relationships with one another. Using network analytic statistics to examine network tie structure, such as density, centralization and reciprocity of relationships, is only one example of how to analyze network characteristics to contribute the field of workplace learning.

22.1.1 Workplaces and Communities—Whole Network Approach

For research approaching workplace learning from an SNA perspective, it is typical to try to understand which actors (workers and experts) are central in work communities. The central position in a knowledge-exchange network has often been described as a patterned set of advice and information flows within a certain network. The nature of a tie between two persons is not insignificant. At the individual level, knowledge diffusion occurs most easily among tightly linked workers (Burt, 1999, 2000; Friedman & Podolny, 1992; Palonen & Hakkarainen, 2014). Consequently, work communities and how they are grouped play an essential role in knowledge exchange.

Second, dense networks and the importance of non-redundant sources of information have been highlighted in SNA studies. In work communities, this is relevant, as information or knowledge that is not commonly shared may be valuable. Burt's (1992) classical argument about 'structural holes' reveals how gaps between non-redundant contacts can generate information benefits for those that can operate between various groupings—both informal or formal. Some people seem to bridge diverse groups through little or no interaction of their own. These boundary-crossing workers or 'bridges' have access to more and more varying information, and they are likely to hear about more valuable information sooner than other workers. They might also be exposed to a range of interpretations and thus be more accurate in their judgements about the trustworthiness and validity of the information available (Burt, 1999). In their report, Friedman and Podolny (1992) found a correlation between a central position within the team and boundary spanning. Those who are most influential within the teams seem to be the most likely to occupy boundary-spanning roles at the organisational level. There is also evidence that individual characteristics, such as high self-monitoring (Mehra et al., 2001) or 'entrepreneurial personality' expressed with (Burt et al., 1998) terminology correlate with network agency. It has been shown that a correlation exists between individuals' cognitive and social structures (Janicik, 1997; Krackhardt, 1990). Finally, the nature of the knowledge exchanged and the strength of ties among network members are important considerations (Hansen, 1999; Uzzi, 1997). Strong ties represent reciprocal, redundant and specialised information flow, whereas weak ties guarantee an adequate number of ties, with the result that new information can also be captured in the network. Strong ties provide the best net effect in the case of complex knowledge, whereas weak ties may be more effective for transmitting well-coded knowledge.

Regarding learning at work, one of the most important study themes concerns repeated exchange relations, often expressed as strength of dyadic ties, that is, the intensity of exchange that reflects the degree to which a link is significant, stable, and mutual (Scott, 1991; Wasserman & Faust, 1995). Standard statistical methods, in contrast, provide information about the attributes of individual actors rather than the relationships among actors. Social relations can be thought of as dyadic attributes, whereas mainstream social science is concerned with monadic attributes. In workplaces, these relations (i.e. dyadic attributes) can be, for example, social roles, affective or cognitive properties, actions, flows, distance or co-occurrence, such as memberships in various groupings or one's geographical location at work.

Relational structure models can be used to describe social and other phenomena in which interactions between units are observed. One example of this might be how interactions flow inside or between teams and workgroups. In organisational settings, the structure of knowledge exchange is often nested. This can be seen when information circulates within a workgroup more than between groups, within a division more than between divisions and so on.

22.1.2 Me and My Colleagues: Egocentric Approach to SNA

According to egocentric network theory, a network is 'owned' by ego or 'focused by ego'. The network members (alters) consist of the persons who are nominated by the ego him-or herself, the ego thus being an informant. Personal networks tend to display grouping at workplaces, this is particularly true concerning work roles, work content, frequency of meetings and so on.

Although the analysis of relational structures focuses on the pattern of relationships between involved actors, the relations are often strongly affected by the monadic attributes possessed by the actors (e.g. age, gender, educational status and field). The complexity of the situation is increased by the fact that it is often a priori unclear which attributes influence the relationship patterns and whether these attributes have been measured. Along with egocentric analysis, we might be interested in whether the size, composition, structure, or stability of networks vary with individual characteristics. Further, at the relationship level, we aim to study whether alter characteristics affect the content of the relationship with the ego, the stability of the relationship or the existence of ties with other alters.

With 'ego', the person themself is referred to, whereas 'alter' indicates the other workers to whom the ego has a tie. For personal networks, four analysis units might be utilised: (1) ego attributes, (2) alter attributes, (3) ego-alter ties and (4) alter-alter ties, where ego has been dropped away. The first two are often studied from a homophily perspective, which has shown that network compositions favour similarity. People typically prefer people that are in some respect similar to them (e.g. regarding domain, age, gender, field of education or cultural background) (McPherson et al., 2001). Alter-alter ties often offer the most interesting information from a workplace learning perspective. This analysis unit simply indicates the ego's social context. If it is sparse, the ego is an important person, as most information must go through her/him. In dense alter-alter networks, the ego is not as important, but the ties overlap in any case, and there are many ways to get information from other alters. The number of alter-alter networks indicates the ego's network size and so on. Specifically, large and sparse alter-alter networks are expected to be rich in knowledge exchange. From a data-gathering perspective, alter-alter ties presuppose much data-gathering work to be done, as network ties must be gathered from all actors. However, if whole network data have been collected for any other purpose, all information is already available. In turn, ego-alter ties provide a useful tool for studies with limited resources. Along with the personal network approach, one can zoom in to individual level (Ibarra & Kilduff, 2005).

In general, personal networks are sources of social support providing instrumental, emotional, informational support or social companionship. From egocentric analysis points of view, important variables are personal network size, its composition (homogeneity, tie strength and so on) and structure, such as the density of the alter network.

In the whole network approach, cohesion indicates resource exchange flows whereas structural equivalence measures focus on its structure. Multiplexity, in turn, occurs in networks when flows interact within and across relationships. It can define how the networks function and cannot be examined by studying network structure or flows separately. Next, cohesion, structural equivalence, and multiplexity are elaborated.

22.1.3 General Terms and Analyses in SNA

In the *cohesion approach*, density and centralisation are basic concepts. Cohesion can be measured in a variety of different ways, most of which are based on dyads (i.e. between two persons). The density measure can be used as a simple way to measure resource exchange in a network: the more actors that have a relationship with one another, the denser the network will be. When studying centralisation, it is possible to focus either on the centrality of an individual actor or the centralisation of a network structure (e.g. team or workplace). The centrality of an individual shows the most popular (or, in the workplace context, most important) actors, those who stand at the centre of attention, in contrast to the isolates, who are rarely or not at all chosen. In addition, individuals' positions as mediators seem to be an important consideration (Borgatti et al., 1996). Being in between must be seen through the concept of path distance, which can be understood if we think of workplace communication as an information flow consisting of a series of connections, not only exchanging information between two people. Interactions are then seen between two nonadjacent actors (e.g. colleagues who are not directly interacting with each other but depend on the other colleagues), who 'lie on the paths' between the two. The term centralisation refers to the extent to which a network has a centralised structure. Centralisation measures are always related to individual centrality measures, with which individuals' positions in the network can be studied. The concepts of density and centralisation further focus on different aspects of the overall compactness of a network. Density describes the general level of cohesion in a graph, while centralisation describes the extent to which this cohesion is organised around particular focal points. Centralisation and density, therefore, are important complementary measures (Wasserman & Faust, 1995, pp. 169–219).

Structurally equivalent people, in turn, occupy the same position in a social structure and are proximate to the extent that they have the same pattern of relationships with occupants of other positions. Therefore, two people are structurally equivalent if they have identical relations with all other individuals in the study population. Structurally equivalent actors do not need to be in direct contact with each other. In the workplace, this might mean those workers who are in contact with the same other workers (i.e. who share the same third parties). Many methods that are concerned with this kind of notion of *social position* or *social role* translate into procedures for analysing actors' structural similarities and patterns of relations in multirelational networks. In real (work) life, it is presumably rare for two actors to have an equivalent position. In practice, this means that people share the same third parties at some approximate level, which is then used to evaluate how similar their network members are.

From the *multiplex network approach*, SNA comprises several network dimensions, such as collaboration, advice asking or social support, which may first be studied separately and later added as one network matrix, which might be called a multirelational network. Relational ties are highly diverse in nature and can represent, for example, feelings (e.g. friendship), communication or behavioural interactions (e.g. cooperation). Each type of relationship spans a social network of its own. The shape of one network influences the topologies of others, as networks of one type may act as a constraint, an inhibitor or a catalyst on networks of another type of relation, all of which may be called multiplex, multirelational, multimodal or multivariate networks. All these terms share a common goal of representing patterns in complex social network data in simplified form to reveal a subset of actors who are similarly embedded in networks of relations and to describe the associations among relations in multirelational networks (Wasserman & Faust, 1995, pp. 345–393).

To sum up, SNA provides access for such workplace learning themes where research questions are, for example, targeted to collegiality or collaboration among workers, structures of expert knowledge or communication, communication inside and between teams and workgroups, explanations for differences between workers in the same work context, questions about which individual attributes are important in workplaces, etc.

22.1.4 Data Gathering for SNA

SNA data can be gathered in many ways, where one distinction can be made between an active (i.e. surveys) and passive network data collection (i.e. metadata from e-mail, or any other electronic technology). In whole network studies within professional learning, network data is often collected using questionnaires in which interpersonal collaboration and informal discussion are addressed (i.e. active network data collection). The questionnaire is not typical; SNA surveys consist of a list of names in which rows represent names, and columns indicate different types of networking relationships, such as advice seeking, information exchange, collaboration or social support. Typical questions are as follows: To whom do you ask advice about work-related matters? With whom do you collaborate? From whom do you receive social support? In this way, data can be collected concerning networking practices, with a focus on tracing how knowledge sharing takes place in a community or an organisation. Each network dimension (questions presented in columns, such as collaboration, advice asking or social support) can be studied separately, but these can also be combined as a multiplex network if a correlation between them is found or if data can be simplified for further analysis. It would also be possible to have a sample of informants who report the information needed for data collection, although reporting ties between other workers may be challenging. If not carefully planned, data gathering for an SNA study can be very time-consuming, as network techniques are usually analysed in case-by-case matrices. For 10 network members, there are 10×10 dyads, minus the persons' own ties to oneself (i.e. altogether 90 cells in the matrix), whereas for 100 network actors there are 9900 dyads and for 200 member-networks—39,800 dyads. In data gathering, snowball sampling is often a better choice than randomised samplings as community level may be one of the analysing units. Samplings are sensitive to research design and may grow large without showing the saturation expected. In an SNA study, a response rate of 70–75% is expected to guarantee that important members are present in the data, which is quite a demanding task in many contexts.

Even if data regarding one person's ties can be taken from full network data sets for egocentric network analysis, egocentric data is often gathered via interviews using name generator techniques in which network members are free-listed. A name generator is simply a question that asks respondents to enumerate individuals with whom they share a particular type of relationship. *Name interpreters*, in turn, are questions that ask respondents about the attributes of individuals they named in response to the name generator, such as their work role. Name interpreters provide descriptive data, whereas name generators provide the essential components that define a framework for a social network (Burt et al., 2012; Eagle & Proeschold-Bell, 2015). For interviewees, this seems to be a natural way to report their personal network members. People tend to classify their collaborators into groups and often members of one group do not know members in all other groups, but the network is connected through ego. In attribute-based analyses, the data are often summaries of attributes of network members that are then compared to the same or other attributes of the respondents.

22.1.5 Density and Centralisation—Most Typical Analyses in SNA

In SNA, densities may be calculated for binary networks, which means that only the existence of the tie is reported (the value in the cell is either one or zero) or valued networks (where the frequency, importance, tie multiplexity or another property is shown by a value larger than one). The density of the binary network is the total number of ties divided by the number of possible ties. This measure can vary from zero to one and may be written as a percentage (Scott, 1991, p. 74). For a valued matrix, attention is paid to how strong (or frequent) the connections are. Consequently, for valued matrices, density values no longer range from zero to one but can be higher than that. Valued matrices are often later dichotomised as binary matrices by using suitable criteria, such as the mean value of the tie. For both valued and binary (dichotomised) matrices, it is possible to compare network density measures between various networks. However, small networks tend to be denser than larger ones, as maintaining ties requires resources.

Further, one fruitful feature in SNA concerns the symmetry of the ties (e.g. whether the tie is reciprocal or not). Some dimensions, such as collaboration, are symmetric in nature, and asymmetry in data must be seen as a kind of bias or subjective criteria for collaboration. In turn, some network dimensions are expected to be asymmetric, with advice seeking being one example. Typically, some workers are sources of expert advice for their colleagues but do not get similar benefit back for themselves, or at least not from the same people to whom they provide help. Instead, the advice to these 'knowledge gatekeepers' may come from outside the company borders. How asymmetric resources flow in networks can be made visible as part of SNA analysis.

In searching for the most active and visible key workers in a company, centrality values indicate the share of addressed and received information (or knowledge) for an individual. As opposed to the active, popular and highly chosen individuals, it is common to find 'isolates' whose location in the network is peripheral, or connections to other workers are missing. For isolates, sensitivity when reporting results is a particularly important consideration. In binary networks, Freeman's degree measure (i.e. degree of connection) is simply the number of other points to which a point (an actor) is adjacent (Scott, 1991, pp. 70–87). The measure is also known as network size in a binary network since it shows the total number of each individual's direct links with other actors (i.e. how many members they have in their network). Degree is the most basic SNA measure and can be utilised in many ways, although it is simple. It is also possible to calculate normalised centrality values. Regarding organisational studies, the advice size variable may be set up as a performance measure for an individual worker, as a rough estimate of his or her relative importance in the organisation or as *cognitive centrality* (Burt, 2000; Krackhardt, 1990). The advice size indicator is utilised later in the empirical section of the chapter.

When studying densities and the centralisation of the network, we can detect the number of relationships and whether they target some actors more than the rest of the participants in the network. Both binary and valued matrices are utilised in the analyses. SNA software typically includes tools to transform a valued matrix into a binary matrix (i.e. to dichotomise it). Furthermore, it is possible to treat information as directed or non-directed. In a non-directed matrix, the direction of a tie is not known or at least not reported; thus, it is symmetric. Regarding directed matrices, we can calculate out-ties (reported by the respondent) and in-ties (reported by other respondents towards a certain respondent). Some analyses can use either symmetric or asymmetric data, whereas for some analyses, data must be symmetrised. The decision to use symmetric or asymmetric (directed) information is usually made by the researcher and should be reported when the results are presented.

Next, an empirical example is presented to explain the kinds of phenomena that can be studied via SNA regarding workplace learning. The example emphasises the local structural properties of the case company, as all qualitative data and other content information has been removed due to the use of pseudonyms. The whole company participated in the study, and there was no problem evaluating what a relevant sample might be. The decision regarding who participates in the study may have an impact on the results and conclusions (Laumann et al., 1983; Smith, 2013).

22.2 Case EcoTec Ltd.

In this chapter, basic SNA analyses for EcoTec Ltd. (a pseudonym) are reported. The following research questions are proposed:

RQ1: How dense are communication flows in expert advice giving and collaboration in the company? How centralised are these flows around the most important workers? Are expert advice-giving and collaboration networks more centralised for frequent (strong) ties compared to all ties?

RQ2: How is expert knowledge communicated inside and between organisational units? Do advice-giving and collaboration flows circulate more inside subunits (departments) than between them? Do the organisational subunits differ in this regard?

RQ3: What kinds of personal attributes do key workers have that are at the core of expert advice flows? What do their personal networks look like and what is the size of their personal networks? Do they have boundary-spanning properties in their personal networks; i.e. are they bridges between the different organisational units and thus various domain fields?

RQ4: What does the latent organisational structure look like, as seen by the network visualisation tool? Are there informal coalitions on the organisation's map that differ from the formal organisational structure? If so, then which organisational subunits (departments) collaborate closely?

22.2.1 Participants

The case company, with its 84 workers (37 females), participated in an SNA study as part of a large organisational developmental project. The empirical data reported here have not been published previously but have been utilised for the company's own purposes. However, the relationships among workers and their attributes, such as membership in four departments, come from original data and are real. There are four departments in EcoTec Ltd.: Department 1 with 49 workers, conducting the company's core functions ('Operations'), Department 2 with 16 workers conducting product development and innovation ('Products and Innovations'), Department 3 and its 10 workers providing support services ('Support Services') and Department 4 where all 9 workers are occupied with selling and marketing ('Sales').

22.2.2 Data Gathering

The data were gathered using a networking questionnaire in which advice seeking and collaboration were addressed. Consequently, the respondents indicated their colleagues' names (roster) in response to the following questions: To whom do you go to ask advice about work-related matters? With whom do you collaborate? The questionnaire was distributed by EcoTec Ltd.'s internal mailing system, helping to obtain a high response rate, which is essential when conducting SNA. Data gathering is often the most difficult phase of SNA studies. The data are sensitive, as they include names and are thus directly related to relationships between human beings. Furthermore, the questionnaire with a roster looks atypical in the respondents' eyes if not introduced carefully. Much trust is needed between workplaces and researchers to collect enough data for SNA, while at the same time following ethical principles in workplaces and in research communities. In this case, trust was earned throughout a joint project that took several months, where a consultant company participated with university researchers to guarantee that the organisation itself got the results that it wanted from the project and that workers felt that the information they gave was treated with care. Before the project started, workers were told that individual-level data was not to be given for the company representatives. The leaders of the company supported data gathering by sending a letter to all workers. Data were collected by the researchers via an electronic form of Analystica (www.analystica.fi) with agreements related to the usage of names.

The response rate for the survey was 91.7%, meaning that 7 out of 84 participants did not respond to the survey. However, as all workers took part in the study project, their personal attributes (gender and department) are known elsewhere and missing data concerns only network questions. In this article, other workers' replies were utilised to complement the data for missing replies by symmetrising data for certain analyses. Regarding missing data, for each dyad where these seven non-responding participants were present, information has been added based on what other respondents have reported (i.e. if person A did not return the questionnaire, we still have information from his/her relationship if person B replied and so on). Yet, regarding dyads where neither A nor B replied, the possible ties are still missing. If the response rate is high, this share is tiny and insignificant. Symmetrising data is not best solution if there is much missing data. Various methods have therefore been developed for the imputation of missing network data. In SNA, it is important to try to obtain as much information as possible, particularly concerning participants who are active and central in their communities. Who they are is often unclear before the data have been collected.

22.2.3 Density and Centralisation

First, the density and centralisation of ties among workers are calculated. In both network matrices (advice giving and collaboration), which are square matrices with the same persons in rows and columns, a value of two is given for a frequent relationship between two workers, one for an occasional relationship, and zero for a missing tie. Thus, if both people in the dyad report the tie in a valued matrix, the values for each dyad vary from zero to four if we add information from both ends of the tie concerning their mutual relationship.

To examine the general structure of exchange relationships in EcoTec Ltd.'s internal network, some descriptive measures are calculated for density and centrality (see Table 22.1).

Both symmetric and asymmetric information is used and values are calculated for binary (dichotomised) and valued data. If valued data are symmetrised, one needs to report whether a smaller value (min) or higher value (max) is used inside the dyad; if matrices are dichotomised, a cut point must be indicated. All these decisions influence the results.

The results in Table 22.1 indicate that respondents have reported more collaboration ties than advice-seeking ties (see higher densities for collaboration), except for the strongest ties, where densities are at about the same level. For binary matrices, 'occasional' and 'often' ties are merged and indicate the number of all ties (see Table 22.1 Advice dich gt0 and Collaboration dich gt0). The density measure shows a 26% density of possible ties for advice seeking and 41% for collaboration concerning all ties. At the individual level, these numbers mean that each worker has, on average, 30.9 collaboration ties to other workers and 21.4 colleagues that give (or are asked for) advice. The latter individual-level information is not given in Table 22.1 but discussed later in this chapter. Whether this is an optimal result must be discussed regarding what kinds of targets and strategies are set for it. Yet, the case company EcoTec Ltd. seems to have dense collaboration and moderate advice-seeking networks. It is not possible to calculate any statistically significant tests to evaluate results or generalise them. In Table 22.1, a tendency can be seen towards

Table 22.1 Network-level density measures for advice-seeking and collaboration ties among workers ($N = 84$) at EcoTec Ltd.

	Density (directed)	Density (symm. max)	Density (symm. min)	Centralisation directed out/in	Centralisation symm. max	Centralisation symm. min
Advice dich gt0	0.26	0.36	0.16	0.51/0.48	0.49	0.38
Advice dich gt1	0.15	0.24	0.06	0.61/0.43	0.52	0.20
Collab dich gt0	0.37	0.46	0.28	0.43/0.39	0.43	0.38
Collab dich gt1	0.14	0.19	0.09	0.23/0.11	0.18	0.13
Advice valued	0.41	0.60	0.22	0.28/0.23	0.47	0.27
Collab valued	0.52	0.66	0.38	0.15/0.12	0.28	0.22

Note: In Table 22.1, 'dichotomised' equals 'dich' and cut point for the operation is given with 'greater than' (gt) criteria

Table 22.2 All ties. Group-level density values for advice seeking by department at EcoTec Ltd.

	Department 1 Operations	Department 2 Products and innovations	Department 3 Support services	Department 4 Sales
Department 1 Operations (N = 49)	**0.41**	0.09	0.06	0.29
Department 2 Products and innovations (N = 16)	0.16	**0.78**	0.09	0.12
Department 3 Support services (N = 10)	0.06	0.05	**0.62**	0.10
Department 4 Sales (N = 9)	0.27	0.04	0.09	**0.49**

Note: The diagonal presents ties inside each department. Ties are dichotomised at cut point zero

higher centralisation of advice ties. This indicates that some workers are more frequently asked for advice than most other workers in the company. Frequent collaboration ties (with 15% density) are rather evenly distributed among workers, but advice-seeking/giving ties (with 16% density) are focused more on certain workers, both regarding in-ties and out-ties.

As mentioned in the theoretical introduction section, ties in workplaces are often nested. Mostly, one works in collaboration inside one's own unit. However, ties to other subunits, such as departments, are typically wanted in organisations to prevent subunits from turning 'silos' and decreasing knowledge flow inside the company. Therefore, a balance needs to be found between groups' internal strengths and interorganisational knowledge flows. Next, we move towards group-level analyses to see how collaboration and advice-seeking ties are observed inside and between subunits (i.e. the four departments in EcoTec Ltd.). At the group level, the case example regards only advice-seeking ties, not collaboration, to simplify the results for methodological purposes.

The density values inside subunits of differing sizes are not directly comparable in resource-taking ties, such as advice networks. Although the density measure itself is the same for all size networks and can be understood as the percentage of possible ties; smaller units tend to be denser than bigger ones as workers have limited time and other capacities for relationships. In Tables 22.2 and 22.3, advice ties are indicated inside and between departments. The departments' internal ties are presented with diagonals. Asymmetry of the ties can be seen, for example, as the ties reported by the Operations Department to the Products and Innovations Department differ from the ties reported the other way around. Self-reported ties (out-degree; advice-seeking) are presented in rows and ties reported by colleagues (in-degree; advice giving) are in columns. In Table 22.2, the tie strength is not taken into account, but the values in each cell are dichotomised at cut point zero, whereas in Table 22.3, the ties represent only the most frequent (i.e. strong ties) and are dichotomised at cut point one.

Table 22.3 Strong ties. Group-level density values for advice seeking by department at EcoTec Ltd.

	Department 1 Operations	Department 2 Products and innovations	Department 3 Support services	Department 4 Sales
Department 1 Operations (N = 49)	**0.25**	0.05	0.03	0.15
Department 2 Products and innovations (N = 16)	0.08	**0.48**	0.04	0.04
Department 3 Support services (N = 10)	0.02	0.02	**0.48**	0.08
Department 4 Sales (N = 9)	0.15	0.01	0.03	**0.29**

Note: The diagonal presents ties inside each department. Ties are dichotomised at cut point one

The results in Tables 22.2 and 22.3 indicate that advice seeking takes mainly place inside departments, as expected. The internal group density of advice seeking indicates values that seem to be low for the largest department—Department 1 (Operations). However, one should not compare big departments with smaller ones. Values indicating tie density between Department 1 (Operations) and Department 4 (Sales) show that these two departments are more connected to each other than to other departments.

In work organisations, emphasis is often placed on workers who are central, mediate knowledge between organisational subunits or transform knowledge outside their workplace. They might be called key workers, knowledge brokers, bridges and so on, as referred to in the introduction. Next, we move towards individual-level analyses and take a look at the workers who are at the core of EcoTec Ltd. and targets for advice seeking by their colleagues.

SNA software typically includes analyses for studying *centralisation* at the network level and *centrality* measures at the individual level. We continue analyses to see how much variation there is at the individual level regarding advice seeking. In binary asymmetric matrices, out-degree values indicate how many colleagues have been reported as sources of advice, whereas in-degree values indicate the number of 'hits' each worker has received. On average, each worker at EcoTech Ltd. has 30.9 collaboration ties and 21.4 advice ties when the frequency of ties is ignored. Regarding strong (frequent) ties, each worker has 12.0 collaboration ties and 12.4 advice ties. Variation between individuals is often largest regarding the 'weakest' ties and less regarding resource-taking ties.

Next, we focus more on advice-seeking matrices. Regarding the valued matrix for advice seeking, the centrality values vary for out-ties from 0 to 125 (SD 23.3) and for in-ties from 2 to 109 (SD 17.6), with the average value being 33.8 for both of these. For frequent (strong) advice ties, out-ties vary from 0 to 63 (SD 13.6), and in-ties vary from 1 to 61 (SD 10.6). The results indicate the different workloads of workers in practice.

In the following step, the advice size indicator is utilised. It is calculated as Freeman's in-degree measure from the valued advice matrix. The workers with the 10 highest scores in advice size are used for further analyses; the highest value among them is 109 and the tenth highest is 58, with 8 workers being somewhere in between. The 10 most central workers come from the biggest department (Operations) (N = 8) and the smallest department (Sales) (N = 2). The Sales Department seems to be able to offer expert advice, although its internal advice exchange value is sparse (see Tables 22.2 and 22.3). Gender does not play a role here, as the 10 most central workers are almost evenly divided by gender (six males, four females). If our study had included interview methods, we probably would have asked about the participants' willingness to be interviewed. SNA has potential to provide a holistic picture of the structure of work organisations, which is helpful for many reasons, in work-related matters and for research purposes. Formal positions and roles matter in organizational contexts. Personal characteristics have not been the main research theme in organisational studies outside of leadership (Balkundi & Kilduff, 2006).

22.2.4 Egocentric Approach and EcoTec Ltd.'s Key Workers

The personal network approach connects individual-and network-level perspectives. Next, the 10 workers with the greatest advice size are analysed regarding their network tie construction at the company. Some basic measures are calculated for this purpose. If the data had not been anonymised and more detailed information was available, many interesting characteristics and attributes, such as work task, length of work experience, age, education and so on could be utilised. Using pseudonyms, we can mainly examine network variables, department information and gender.

Similar to the whole network perspective, density and centrality measures can be utilised in individual-level analyses. Centrality in a network is not only related to the number of ties at the dyad level but also to the 'mediator' or 'brokering' aspect, such as being on a path between two other workers, not in direct contact with someone. This notion of transitivity is essential for network studies. Brokering value can be calculated, for example, based on how many times an ego lies on the shortest path between two alters that are not directly connected. Brokering value is a parallel concept to betweenness, mentioned in the Introduction (Wasserman & Faust, 1995, pp. 188–192). In Table 22.4, a normalised value of the broker index is given (nBroker), where the broker value is divided by the number of brokerage possibilities.

Table 22.4 indicates some basic egocentric measures. The variation shown in Table 22.4 is not large, as only central workers are present in the analysis. However, the worker with the highest advice size value (ID 2) has the sparsest alter-alter network and the largest nBroker values. Various centrality measures correlate but also complement each other. Central workers have a broker position (i.e. a mediator's role) in addition to having many ties with their colleagues.

Table 22.4 Examples of network indicators for most central persons with high advice size (Freeman's in-degree value in advice asking)

ID	Advice size	Ego's department	Ego's gender	Network size (includes in- and out-ties)	Alter-alter density (ego not present)	nBroker
2	109	Operations	Female	69	61%	0.69
13	56	Operations	Female	59	61%	0.69
34	60	Operations	Male	34	71%	0.53
37	72	Operations	Male	42	72%	0.55
41	64	Operations	Male	49	67%	0.57
49	60	Sales	Female	48	65%	0.60
51	58	Operations	Male	30	79%	0.42
68	68	Sales	Female	51	64%	0.64
75	63	Operations	Male	50	61%	0.64
76	58	Operations	Male	42	71%	0.55

Advice size measure is in-degree on valued advice matrix; other measures are calculated on dichotomised advice matrix (cut point at zero)

22.2.5 *Visualisations*

For visualising purpose and to simplify the process, advice network and collaboration network dimensions are added. Correlations are first calculated to see how much these dimensions overlap with each other. Pearson's Quadratic Assignment Procedure (QAP) correlation compares two matrices regarding their similarities. High correlations for the same persons under various network dimensions are expected, although varying densities may decrease the correlation of the two matrices. For dichotomised matrices in advice seeking and collaboration (at cut point zero), the Pearson QAP correlation value is expectedly high ($r = 0.740$) and the same is true with valued matrices for these two network dimensions ($r = 0.750$). Both results are statistically significant ($p < 0.001$). Thus, workers often ask for work-related advice from their collaborators, even if the advice-seeking network is sparser than the collaboration network. This provides one more reason to add the two network dimensions. The new matrix is called the 'pooled knowledge-exchange network'. A graph is drawn based on the added matrix using the Netdraw programme (see Fig. 22.1).

In Fig. 22.1, nodes represent workers and lines represent the most frequent ties between them. Had we added all ties to the figure, it would have been too messy. Even now, it looks dense, although the density value for the strongest dichotomised ties (expressed as lines in the figure) is only 0.18, indicating that 18% of possible ties are present in the pooled knowledge-exchange network (at cut point two). Thus, the line is present in the figure for those dyads where the value in the cell is three or four. The node colours indicate department information. Typically, SNA visualisations are made so that the node locations on the map minimise crossing lines and the lines are as short as possible; the nodes are then placed close to those nodes with which

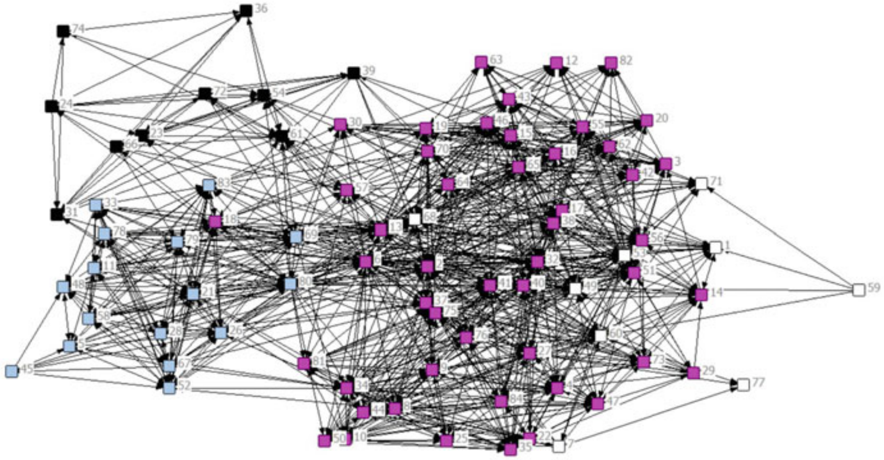

Fig. 22.1 Pooled and Valued Knowledge-Exchange Network for EcoTec Ltd. Advice and collaboration dimensions have been added. Nodes represent workers, coloured based on their department. The lines represent the most frequent ties between workers (cut point two); Violet = Operations Department (1); Blue = Products and Innovations Department (2); Black = Support Services Department (3); White = Sales Department (4). Netdraw programme

they have (most) ties. There are typically various algorithms used in drawing programmes for this purpose.

In Fig. 22.1, the results, some of which are shown by density values at the group level, indicate that workers seem to exchange knowledge mostly with colleagues from their own departments. Only a small number of ties cross department borders. Most relationships can be observed inside the biggest department (Operations), where the core expertise of the company is situated and also between Sales Department members who are stakeholders for marketing and selling. These two departments are integrated on the right side of the graph. The Products and Innovations Department and Support Service Department are on the other side of the graph, making it bipolar. There is a gap between the Products and Innovations Department and the Support Services Department. The merged coalition thus consists of Operations Department and Sales Department members. Earlier reported results in this paper have also indicated that the most central workers are all on the right side of the graph. The Support Service Department seems to be the sparsest and most peripheral part of the organisation. This might fit with the company's interests, but it is worth reflecting on EcoTec Ltd.'s strategy plans if SNA results are used for consultation purposes. In all, local interests—that is, the company's approach—differ from scientific purposes. Scientific research might, in turn, be interested in how expertise develops and how organisational structures are related to learning at work, not in an internal evaluation of one business case (Mehra et al., 2006; Starkey et al., 2000). There are many tools to create visualisations (De Laat & Schreurs, 2013; Mintzberg & Van der Heyden, 1999), for example, Gephi, which is an open-source application that can analyse social networks through a graphical user interface (Bastian et al., 2009).

To conclude, SNA can produce visualisations of which one tiny example is given here, created easily with Netdraw software integrated as a part of Ucinet. Moreover, visualisations may cover a range of topics relevant to making organisational performance and complex structures visible at various organisational levels. SNA's contribution to organisational consulting is discussed more elsewhere (Palonen & Froehlich, 2020). Expectations certainly differ in this regard between the world of work, organisational consulting and scientific communities and its publishing forums.

22.3 Discussion

SNA studies have shown strengths in their ability to operate using various units of analysis. At the network level, the focus is typically on structures and various dimensions in multiplex networks, where more or less correlating network dimensions can be compared or summarised. At the group level, information flows inside and between organisational subunits, such as departments or teams, and may be studied. At the individual level, workers who are sought out for access to certain types of information and expertise, who are engaged in routine decision making or who are consulted when dealing with problems, look interesting. In the personal (egocentric) network approach, these analysis units are intertwined. Earlier research has especially emphasised the cohesion perspective, where key workers, boundary-crossing workers and their relationships in informal and formal networks have been studied (Sparrowe & Liden, 2005). Densely linked networks are shown to be more efficient than sparsely linked groups in many situations. Brokers for knowledge or resources can become overwhelmed by their roles (Long et al., 2013). SNA studies have also claimed that a team should not be seen as an average of its members' attributes but instead emphasise team members' relationships to other workers and access to rich knowledge repositories. Balkundi and Harrison (2006) indicated in their meta-analysis that teams with denser expressive and instrumental social network ties tend to perform better and remain more viable than other teams. In particular, the centrality of leaders is important (Cross et al., 2002). Concrete interactions between workers might be utilised to examine latent constructs and concepts used in literature and consulting, such as climate or informal organisational culture, for example (Zohar & Tenne-Gazit, 2008).

To conclude, in workplaces, the aim is often focused on the individuals' personal characteristics or attributes, such as their general communication skills, whereas SNA's contribution is from a structural relational perspective. Employees' collaboration networks have been further analysed to understand how high-performing individuals and teams communicate or take part in decision making (Brass, 1984; Froehlich & Messmann, 2017; Moolenaar et al., 2010). At the same time, there is doubt about what kind of measures regarding 'high performance' or concerning indicators for learning are available. SNA might be able to provide at least one performance indicator (i.e. advice size shows how useful a person is at providing expert advice or other resources to his or her coworkers).

22.3.1 Criticism of SNA and Some Limitations

There are several concerns regarding SNA studies that need to be discussed. One of them is the contextual nature of the findings and problems generalising these to other contexts. The results often depend heavily on the underlying cases (Crossley, 2010). For instance, the effects of density are difficult to compare across various sizes of organisations in different domains. Including qualitative information can, in this regard, produce explanations for quantitative results (Bolíbar, 2015; Domínguez & Hollstein, 2014; Franke & Wald, 2006; Rienties et al., 2015). According to Hollstein (2011), qualitatively-oriented SNA might cope better with questions relating to the constitution and dynamics of social networks.

As the cohesive approach seems to be the main approach for SNA studies, the study design has often been based on frequency measures instead of on measuring the true quality of the relationships. Therefore, there is a challenge in how to indicate that some collaboration relationships are worthier than others or that tiny communities can be powerful.

A significant problem is related to ethical concerns, especially the use of names in data gathering. Therefore, data collection is an Achilles heel for SNA projects, at least regarding surveys. One related issue is the requirement of a high response rate, which is not easy to achieve, keeping participants' voluntary participation in mind. Even if the names and other attributes of respondents are anonymised, one cannot easily hide the structural embeddedness of a single person if the target organisation is known or if results are utilised by the organisation itself or given to participants. At the same time, while highly contextualised data may be useful for participating organisations, the same features sometimes make actors easy for network insiders to identify. Information about persons not participating may also be seen if it is removed from the data—at least indicating that a person is missing and did not participate in the study. One solution to solve the problems in participating organisations is to leave individual-level results out of reports and instead use group-level outcomes and indicators.

SNA researchers in the field of organisational studies collect information regarding positive ties, such as collaboration and advice seeking. Information related to negative ties is rarely collected outside of some examples (Labianca, 2014; Labianca et al., 1998). SNA might have the potential to provide information for solving conflicts if it were easier from an ethical point of view. In schools, for example, bullying is examined by using a negative tie perspective. However, for workplaces, it is hard to believe that people who are, for example, obstacles to innovation or well-being at work are nominated by name in the survey.

22.3.2 Future Directions for SNA and Workplace Learning Studies

Emphasising relationships over attributes, utilising the embeddedness of social entities and empirical research more generally has shown potential in SNA studies (Kilduff & Brass, 2010; Cross & Parker, 2004). Furthermore, SNA is increasingly becoming a component in the mixed-method perspective (Palonen & Froehlich, 2020). New and classical tools may complement the viewpoint used in the workplace learning approach or participative action in work organisations (Borgatti & Foster, 2003; Borgatti & Li, 2009; Collins & Clark, 2003; Cross et al., 2002; Tichy et al., 1979).

Regarding workplace learning and learning in organisations, the structural equivalence perspective should be utilised. A rich collection of SNA software includes several quantitative and qualitative analyses for all phases in SNA study processes—from data gathering to results presentation (e.g. visualisations). In this chapter, the classical mainstream programme Ucinet was utilised for data analysis (Borgatti et al., 2002). Recently, SNA has been conducted within the versatile and popular R environment, which contains several packages relevant to SNA. Many methodological handbooks are available, as is an international scientific community (https://www.insna.org/), to help newcomers to the field or advanced researchers with new insights, theories, and tools.

SNA is truly a multi-scientific paradigm in which researchers from various domains collaborate and organise scientific community activities. SNA is a useful tool that can be combined with other methodological tools and elements for mixed methods in SNA, or MMSNA, as it has been abbreviated in the literature (Palonen & Froehlich, 2020).

The author of the article has got funding from Joint Nordic-UK Research Programme on Migration and Integration, Nordforsk (Project no. 94935).

References

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders and time in teams: Strong inference about the effects of network structure on team viability. *Academy of Management Journal, 49*(1), 49–68. https://doi.org/10.2307/20159745

Balkundi, P., & Kilduff, M. (2006). The ties that lead: A social network approach to leadership. *The Leadership Quarterly, 17*(4), 419–439. https://doi.org/10.1016/j.leaqua.2006.01.001

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *ICWSM, 3*(1), 361–362.

Bolíbar, M. (2015). Macro, meso, micro: Broadening the 'social' of social network analysis with a mixed methods approach. *Quality & Quantity, 3*, 2217–2236. https://doi.org/10.1007/s11135-015-0259-0

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management, 29*(6), 991–1013. https://doi.org/10.1016/S0149-2063(03)00087-4

Borgatti, S. P., & Li, X. (2009). On social network analysis in a supply chain context*. *Journal of Supply Chain Management, 45*(2), 5–22. https://doi.org/10.1111/j.1745-493X.2009.03166.x
Borgatti, S., Everett, M., & Freeman, L. (1996). *UCINET IV version 1.64 reference manual*. Natick, Analytic Technologies.
Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINet for windows: Software for social network analysis*. Analytic Technologies.
Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly, 29*(4), 518–539. https://doi.org/10.2307/2392937
Brown, J., & Duguid, P. (2001). Knowledge in organization: A social practice perspective. *Organization Science, 12*, 198–213.
Burt, R. S. (1992). *Structural holes: The social structure of competition*. Harvard University Press.
Burt, R. S. (1999). Entrepeneurs, distrust, and third parties: A strategic look at the dark side of dense networks. In L. L. Thompson, J. M. Levine, & M. Messick (Eds.), *Shared cognition in organizations. The management of knowledge* (pp. 213–243). Erlbaum.
Burt, R. S. (2000). The network structure of social capital. In R. I. Sutton & B. M. Staw (Eds.), *Organizational behaviour* (pp. 345–423). JAI.
Burt, R. S., Jannotta, J. E., & Mahoney, J. T. (1998). Personality correlates of structural holes. *Social Networks, 20*(1), 63–87. https://doi.org/10.1016/S0378-8733(97)00005-1
Burt, R. S., Meltzer, D. O., Seid, M., Borgert, A., Chung, J. W., Colletti, R. B., … Margolis, P. (2012). What's in a name generator? Choosing the right name generators for social network surveys in healthcare quality and safety research. *BMJ Quality and Safety, 21*(12), 992–1000.
Collins, C. J., & Clark, K. D. (2003). Strategic human resource practices, top management team social networks, and firm performance: The role of human resource practices in creating organizational competitive advantage. *Academy of Management Journal, 46*(6), 740–751. https://doi.org/10.5465/30040665
Cross, R. L., & Parker, A. (2004). *The hidden power of social networks: Understanding how work really gets done in organizations*. (4.1.2004 edition). Harvard Business Review Press.
Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review, 44*(2), 25–47.
Crossley, N. (2010). *Towards relational sociology*. Routledge.
De Laat, M., & Schreurs, B. (2013). Visualizing informal professional development networks: Building a case for learning analytics in the workplace. *American Behavioral Scientist, 57*(10), 1421–1438.
Domínguez, S., & Hollstein, B. (2014). *Mixed methods social networks research: Design and applications*. Cambridge University Press.
Eagle, D. E., & Proeschold-Bell, R. J. (2015). Methodological considerations in the use of name generators and interpreters. *Social Networks, 40*, 75–83.
Franke, K., & Wald, A. (2006). Möglichkeiten der Triangulation quantitativer und qualitativer Methoden in der Netzwerkanalyse [Opportunities for triangulating quantitative and qualitative methods in social network analysis]. In B. Hollstein & F. Straus (Eds.), *Qualitative Netzwerkanalyse: Konzepte, Methoden, Anwendungen* [Qualitative network analysis: Concepts, methods, applications] (pp. 153–175). VS Verlag für Sozialwissenschaften.
Friedman, R. A., & Podolny, J. (1992). Differentiation of boundary spanning roles: Labour negotiations and implications for role conflict. *Administrative Science Quarterly, 37*(1), 20–48.
Froehlich, D. E., & Messmann, G. (2017). The social side of innovative work behaviour: Determinants of social interaction during organizational innovation processes. *Business Creativity and the Creative Economy, 3*(1), 31–41.
Hakkarainen, K., Palonen, T., Paavola, S., & Lehtinen, E. (2004). *Communities of networked expertise: Professional and educational perspectives*. Elsevier Science.
Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly, 44*(1), 82–102.

Hollstein, B. (2011). Qualitative approaches to social reality: The search for meaning. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 404–416). Sage.

Ibarra, H., & Kilduff, T. W. (2005). Zooming in and out: Connecting individuals and collectivities at the frontiers of organizational network research. *Organization Science, 16*(4), 359–371.

Janicik, G. A. (1997). *The social cognition of social structure: Examining the learning of relations*. University of Chicago Press.

Kilduff, M., & Brass, D. J. (2010). Organizational social network research: Core ideas and key debates. *The Academy of Management Annals, 4*(1), 317–357. https://doi.org/10.1080/19416520.2010.494827

Krackhardt, D. (1990). Assessing the political landscape: Structure, cognition and power in organizations. *Administrative Science Quarterly, 35*, 342–360.

Labianca, G. (J.). (2014). Negative ties in organizational networks. In D. J. Brass, G. (Joe) Labianca, A. Mehra, D. S. Halgin, & S. P. Borgatti (Eds.), *Research in the sociology of organizations* (Vol. 40, pp. 239–259). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2014)0000040012

Labianca, G., Brass, D. J., & Gray, B. (1998). Social networks and perceptions of intergroup conflict: The role of negative relationships and third parties. *Academy of Management Journal, 41*, 55–67.

Laumann, E. O., Marsden, P. V., & Prensky, D. (1983). The boundary specification problem in network analysis. In R. S. Burt & E. Minor (Eds.), *Applied network analysis: A methodological introduction* (pp. 18–87). Sage.

Long, J. C., Cunningham, F. C., & Braithwaite, J. (2013). Bridges, brokers and boundary spanners in collaborative networks: A systematic review. *BMC Health Services Research, 13*(1), 158. https://doi.org/10.1186/1472-6963-13-158

McCarty, C. (2002). Measuring structure in personal networks. *Journal of Social Structure, 3*(1). http://www.bebr.ufl.edu/Articles/Structure_persona_networks.pdf

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444.

Mehra, A., Kilduff, M., & Brass, D. J. (2001). The social networks of high and low self-monitors: Implications for workplace performance. *Administrative Science Quarterly, 46*, 121–146.

Mehra, A., Smith, B. R., Dixon, A. L., & Robertson, B. (2006). Distributed leadership in teams: The network of leadership perceptions and team performance. *The Leadership Quarterly, 17*(3), 232–245. https://doi.org/10.1016/j.leaqua.2006.02.003

Mintzberg, H., & Van der Heyden, L. (1999). Organigraphs: Drawing how companies really work. *Harvard Business Review, 77*(5), 87–94.

Moolenaar, N. M., Daly, A. J., & Sleegers, P. J. C. (2010). Occupying the principal position: Examining relationships between transformational leadership, social network position, and schools' innovative climate. *Educational Administration Quarterly, 46*(5), 623–670. https://doi.org/10.1177/0013161X10378689

Nardi, B., Whittaker, S., & Schwarz, H. (2000). It's not what you know, it's who you know: Work in the information age. *First Monday, 5*, 5. http://firstmonday.org/issues/issue5_5/nardi/index.html

Palonen, T., & Froehlich, D. (2020). Mixed methods social network analysis to drive organizational development. In D. E. Froehlich, M. Rehm, & B. C. Rienties (Eds.), *Mixed methods social network analysis: Theories and methodologies in learning and education*. Routledge.

Palonen, T., & Hakkarainen, K. (2014). Social network analyses of learning at workplaces. In *Discourses on professional learning* (pp. 293–315). Springer.

Rienties, B. C., Johan, N., & Jindal-Snape, D. (2015). Bridge building potential in cross-cultural learning: A mixed method study. *Asia Pacific Education Review, 16*(1), 37–48. https://doi.org/10.1007/s12564-014-9352-7

Scott, J. (1991). *Social network analysis. A handbook*. Sage.

Smith, S. S. (2013). Social network boundaries and tricky to access populations: A qualitative approach. *International Journal of Social Research Methodology, 17*(6), 613–623. https://doi.org/10.1080/13645579.2013.820076

Sparrowe, R. T., & Liden, R. C. (2005). Two routes to influence: Integrating leader-member exchange and social network perspectives. *Administrative Science Quarterly, 50*(4), 505–535. https://doi.org/10.2189/asqu.50.4.505

Starkey, K., Barnatt, C., & Tempest, S. (2000). Beyond networks and hierarchies: Latent organizations in the U.K. television industry. *Organization Science, 11*(3), 299–305. https://doi.org/10.1287/orsc.11.3.299.12500

Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *The Academy of Management Review, 4*(4), 507–519. https://doi.org/10.2307/257851

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly, 42*, 35–67.

Wasserman, S., & Faust, K. (1995). *Social network analysis: Methods and applications*. Cambridge University Press.

Zohar, D., & Tenne-Gazit, O. (2008). Transformational leadership and group interaction as climate antecedents: A social network analysis. *Journal of Applied Psychology, 93*(4), 744–757. https://doi.org/10.1037/0021-9010.93.4.744

Tuire Palonen (PhD) works as a University Research Fellow in the Teacher Education Department at University of Turku, Finland. Her interests are focused on the context of learning and networked environments, in school and out. Connections and ties between people, analysed by the tools of Social Network Analysis (SNA), make a core for her special expert area, the target group for research ranging from children at school up to senior workers at workplaces. Trying to develop visualisations to better understand complex structures of various kinds of social communities, and what happens in these, have been her main interests recently.

Chapter 23
Design-Based Research – Grounding, Understanding and Empirical Illustration in the Context of Vocational Education

Karl-Heinz Gerholz and Anne Wagner

Abstract Design-based research is an approach to develop new theories and educational practices in a context-sensitive manner. The aim of this chapter is to introduce design-based research using the example of a concrete design research project. The combination of theory building and practical design is central to this approach. The research process is carried out in iterative design cycles consisting of design, implementation and evaluation. The orientation to the respective context, theory-based procedure, collaboration of practitioners and researchers, and, related to this, interconnection of theory and practice are characteristic of design research. Building on the theoretical and methodical basics of design research, the chapter shows a possible implementation of this approach based on a project in vocational education. The research interest of this project relates to the design of vocational education lessons with tablets to prepare vocational students for the requirements in digitized working processes. Building on the theoretical framework of the project, it illustrates how design, implementation and evaluation were implemented in cooperation with teachers – the educational practice – at the vocational schools (e.g. digital consultation hours, workshops). Results of the project are presented in terms of didactical design principles. At the end of the chapter, three reflected recommendations for design research projects – documentation and structuring of time, relevance of a design research portfolio and dealing with data formats – will be carved out.

Keywords Design-based research · Vocational education · Professionalization of teachers

K.-H. Gerholz (✉) · A. Wagner
University of Bamberg, Bamberg, Germany
e-mail: Karl-heinz.Gerholz@uni-bamberg.de; Anne.wagner@uni-bamberg.de

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_23

23.1 Introduction or How Can Tablets Be Used in a Pedagogical Manner in Vocational Schools?

A main challenge in educational research can be described by the terms ‘rigor’ and ‘relevance.’ Rigor refers to scientific rules, methods and norms or scientific rigor. Relevance refers to the practical relevance of scientific insights and knowledge (Finch et al., 2018; Gerholz, 2022). Of course, science and practice are two different systems which are – depending on the position – ‘interdependent’ (Perna, 2016) or ‘unbridgeable’ (Kieser & Leiner, 2009). The idea of design research is to offer a way to overcome the rigor–relevance gap in educational research. Design research aims to develop both practical solutions or interventions and theoretical insights simultaneously in real educational contexts (McKenney & Reeves, 2012). The issue of design research has less to do with the difference between basic and applied research (Reeves, 2006), and more about building a bridge between design and research. Therefore, design research can be explained as a framework for educational research (Euler, 2012). Both design and research are important, with the implication that the design process is an important approach with its own logic (Edelson, 2002). The design process can be understood as a scientific act undertaken in a creative manner (Reinmann, 2014). Thus, design-based research will make a contribution toward understanding the relationship between educational theories, designed interventions and practical surroundings. “Because the intervention as enacted is a product of the context in which it is implemented, the intervention is the outcome (or at least an outcome) in an important sense” (Design-Based Research Collective, 2003, p. 5).

Design-based research is an emerging approach and has been developed over the last 20 years. However, different approaches can be accumulated here, such as “design experiments” (Brown, 1992), “design-based research” (Design-Based Research Collective, 2003), “development research” (van den Akker, 1999) or “educational design research” (McKenney & Reeves, 2012). These approaches can be summarized under the term “design research” (Euler, 2014, p. 16). The common mission of these approaches is to combine scientific and practical interests. This reflects the two goals, on the one hand, designing learning environments and, on the other hand, developing ‘prototheories’ of learning. Both processes are intertwined (Design-Based Research Collective, 2003, p. 5).

The approach of design-based research has becoming more widespread in the last few years. Results of design research projects are published increasingly in international journals, special issues for design research can be observed (e.g. Euler & Sloane, 2014) and new journals with an explicit focus on design research have been established (e.g. Educational Design Research, EDeR). However, depending on who is asked, the knowledge and understanding of design-based research is also elaborated differently between the subareas in educational research. Therefore, this chapter presents an overview of the design research approach. In order to understand the design research approach better, we illustrate it based on the tabletBS.dual project, a real design research project at vocational schools in Germany. It is a pilot project to implement tablets in vocational schools in a pedagogical way and foster digital competencies among vocational education and training (VET) teachers and their students. The project is anchored in the dual VET system organized in companies and vocational schools (Gerholz & Brahm, 2014). The tabletBS.dual project is structured as a design research project incorporating 50 vocational schools. The tablets can be seen as an innovation which is implemented in the pedagogical processes at the vocational schools. The starting point of a design research project is an existing educational problem in the practical field. The question is not whether there is a concept or theory which will effective solve the given educational problem. A design research process begins with the question: "How can an aspired, initially vaguely formulated goal be reached by a yet to be developed design?" (Euler, 2017, p. 4). Thus, the starting point or the central question in the tabletBS.dual project was: How to design vocational education lessons with tablets in order to prepare vocational students for the requirements in digitized working processes?

Bearing the project mentioned in mind, we describe in the following the characteristics and methodical way of the design research approach (Sects. 23.2, 23.3, and 23.4) as a framework. Based on this, we illustrate the realization of design research projects based on the context of the tabletBS.dual project presented (Sects. 23.5 and 23.6).

23.2 Characteristics of Design Research

Design research is defined as "the systematic study of designing, developing and evaluating educational interventions (such as programs, teaching-learning strategies and materials, products and systems) as solutions for complex problems in educational practice, which also aims at advancing our knowledge about the characteristics of these interventions and the processes of designing and developing them" (Plomp, 2007, p. 13).

The definition already underlines the practical perspective at the beginning of a design research process. Design research can be seen as a possible and promising way to overcome the research-practice gap. The weak link between educational research and practical problems is often mentioned in the literature (e.g. McKenney & Reeves, 2012). The reason for the weak link can be found on both sides. Practice sees little potential in research results or uses the results inappropriately. Research offers only a few clear results and few practical results (Broekkamp & van Hout-Wolters, 2007; McKenney & Reeves, 2012). Therefore, the cooperation between practitioners and researchers is a main condition in design research projects. The cooperation is important to connect both perspectives practice and science or scientific theories and daily-life theories early in the process. The cooperation at the beginning of the tabletBS.dual project, for instance, was important in order to understand the implementation problem of the tablets at the vocational schools from the perspective of the practitioners (e.g. fears about dealing with tablets in VET lessons, low digital self-efficacy or challenges in the IT infrastructure). Consequently, the context could be extensively illuminated, and this was a basis for a theoretical foundation based on scientific theories: Which conceptual aspects were relevant for the following design and research process? Moreover, collaboration is also a main condition in the course of a design research process. Having carved out the definition above, design research is characterized by a phase-oriented process – designing, developing and evaluating – to develop valuable educational environments together between practice and research. The cooperation between practice and science is important in all phases to ensure a mutual reference between scientific and practical insights.

All in all, the characteristics of design research can be described in terms of (a) context-sensitive, (b) theory-based, (c) collaborative and (d) interconnected (e.g. Euler, 2014; McKenney & Reeves, 2012; Reinmann, 2005).

(ad a) Context-Sensitive The starting point of design research processes is an educational problem in a given context. An educational prototype should be developed regarding this context to reach an aim. Thus, the practitioners and scientists act in a specific, educational context in design research projects. Consequently, the educational practice, artefacts or new and adapted theories produced within design research are context-sensitive (Barab & Squire, 2004; Reinmann, 2005). They require an effect for the given context and not beyond that context. In other words, design research can produce context-specific findings. A generalization means a generalization in the context given.

(ad b) Theory-Based Design research is theory-based. This is a basis in general for educational research. Doing research is accompanied by using theory and scientific knowledge. Theory in design research processes is not an aspect of the reproducibility of theoretical or empirical insights (Reinmann & Sesink, 2014). Theory serves as a tool utilized to understand the problem of the educational practice, understand unforeseen events in the design or implementation process and evaluate the interventions designed. Using theoretical and conceptual approaches in design research is inspired by a systematic and creative manner. Theory helps to describe activities and results in design research processes, especially the interplay between design and implementation. Theory also helps to describe what could be (Schwartz et al., 2005).

(ad c) Collaborative The collaboration of practitioners and researchers is a main element in design research. The fundamental interests of the researchers and practitioners remain preserved, but the actions of these two groups can vary during a design research project (Euler, 2014). The practitioners, for instance, have a relevant practical knowledge and are sensitized to the educational target group. The researchers can help with the selection of the most fitting theoretical approaches for the development of the educational intervention and the evaluation of the designs realized (McKenney & Reeves, 2012). The collaboration between research and practice is generally relevant in all phases – design, implementation and evaluation – of the design cycles (Gerholz, 2014). It is important that collaboration is a mutual relationship for a reciprocal development of both groups. The role can change if practitioners support the researchers or the other way around. It is not the idea that it is a hierarchal relationship in which researchers act as explainers or trainers for practice to solve the problem on their own.

(ad d) Interconnected Design and research are interconnected in design research processes. Both are elements integrated into each other (Bannan-Ritland & Baek, 2008) or run synchronously (Ejersbo et al., 2008). However, it is important for the commitment between practitioners and researchers that theoretical insights and practical solutions are developed simultaneously in the real educational context (McKenney & Reeves, 2012, p. 9). Therefore, collaboration is a precondition for the interconnection of research and design. Euler described the connection as a win-win situation: "Science can select its research topics from problems in practice, while developing solutions for practice is enhanced by resorting to scientific theories" (Euler, 2014, p. 21).

23.3 Procedures of Design Research

The claim of design research is simultaneously its challenge: The challenge to develop trajectories which meet the goals of "refining locally valuable innovations and developing more globally usable knowledge for the field" (Design-Based Research Collective, 2003, p. 7). The latter refers to the context-sensitive theory building. For this, design research projects show a cyclical architecture in their proceedings (Euler, 2014; Gerholz, 2014; Plomp, 2013; van den Akker, 2007). Different models which describe the procedure of a design research process from several authors exist in the literature. They are similar to each other and emphasize the characteristics of design research, especially the iterative, phase-oriented and cooperative procedure.

Reeves (2006, p. 58) presents a linear model with four phases. In *phase one*, the analysis of the practical problem based on the educational context is elementary. *Phase two* represents the design phase in the sense of the development of solutions for the problem given. The solutions should be informed by existing design principles and technological innovations. *Phase three* is characterized by the iterative process of testing and refinement. And *phase four* refers to the reflection of the whole design research process with the aim of describing design principles for the problem given. Design principles are the generalized experiences and insights realized during the design research process and can be applied in similar contexts.

Reeves' phases can change back and fore during the process, however, the linearity is rarely possible in a real design research process. The iteration and cooperation especially need a more flexible framework. Bearing this in mind, McKenney and Reeves (2018, p. 67) suggest a more generic model to conduct design research processes. Their model (see Fig. 23.1) follows three phases in an iterative structure with a dual focus on theory and practice. In this model, the implementation is ongoing: (a) analysis and exploration, (b) design and construction, and (c) evaluation and reflection.

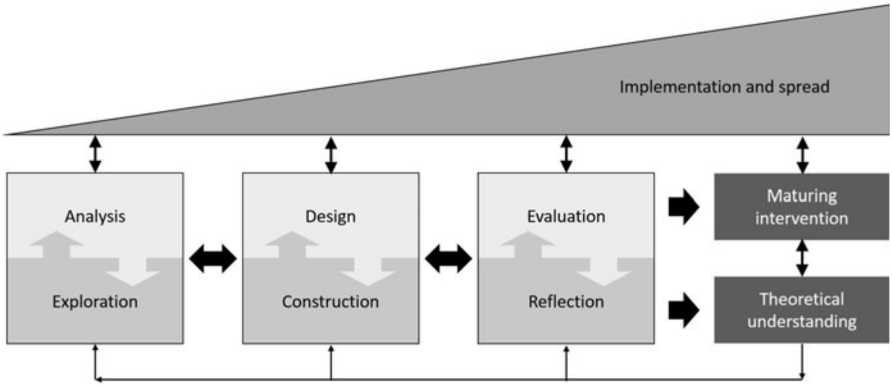

Fig. 23.1 Generic model for conducting design research. (McKenney & Reeves, 2018, p. 82)

(ad a) Analysis and Exploration Analysis refers to the problem identification and diagnosis. On the one hand, the collaboration with the practitioners to get an understanding of the problem, the target context and the needs of the practice are included in the analysis phase. On the other hand, the scientific exploration (e.g. literature review) to gain theoretical inputs for the understanding of the problem, the context or other topics is needed. Based on the analysis, the exploration is made, and similar problems and their solutions are explored.

(ad b) Design and Construction Design and construction is the beginning of the solving process to reach the aim. It is the creation of the conceptual model. McKenney and Reeves describe this as a deliberative-generative cycle in which a potential solution to the problem will be explored, mapped and considered. Construction refers to a prototyping approach, in the sense of taking the design ideas and bringing them into practice. The educational intervention is conceived from a practical perspective. From a theoretical perspective, the results of the scientific exploration are introduced. This also implies reasoning for the procedure of the intervention.

(ad c) Evaluation and Reflection Evaluation is used here in a broad understanding. It means the empirical testing and can be related to the intervention and the effectiveness of the process of implementation, the local viability or broader institutionalization. Reflection means the examination from the research and development perspective, with the things achieved so far. The aim from the research perspective is to produce a theoretical understanding based on the results available (e.g. empirical results, experiences of practitioners with the intervention) and the combined activities of practitioners and researchers. The aim of the practice perspective is to develop ideas for the redesign of the intervention.

The *Design and Construction* particularly shows that the collaborative design process is a hallmark in design research. However, the phases of design are underexposed in the design research literature. Reinmann (2014) described the

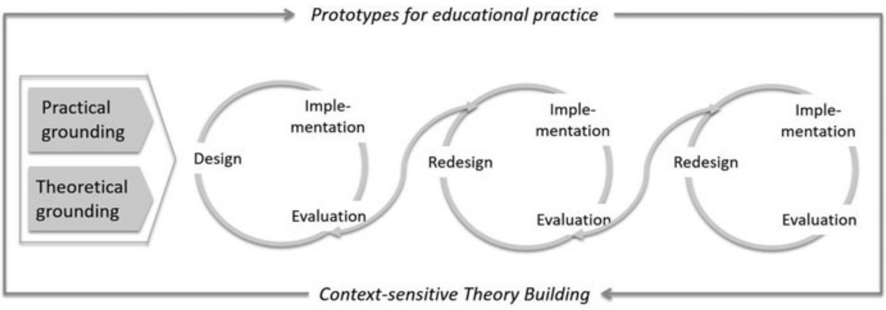

Fig. 23.2 Model of doing design research. (In Orientation Gerholz, 2014)

design process as a scientific act in design research which gives creativity a place. Easterday et al. (2017) understand the design process as an iterative matter with five different steps: (a) the *focus phase* refers to the scope of the project (e.g. identifying the general problem and initial direction of the project); (b) the *understand phase* aims at a deeper analysis of the problem (e.g. methods, such as literature review or concept mapping); (c) the *define phase* includes the operationalizing of the aims and defining criteria to assess the achievement of the aims; (d) the *conceive phase* aims at the principles and the arguments for a prototype (e.g. working on blueprints); and (e) the *build phase* represents the realization of the prototype. The model underlines that the perspectives or cooperation of research and educational practice are elementary during the design process. In addition, the iterative philosophy during the steps in the design process is shown.

Not everything can be strictly planned in design research processes. Therefore, the approaches presented are more heuristics to structure design processes. Bearing the models presented in mind, it is obvious that doing design research is a multidimensional process. Design research is a creative and systematic process at the same time. It is also carried out in a practical and scientific manner. Design research is both solution-oriented for the educational practice and theory-oriented for the science.

All in all, the cyclical architecture is a common ground in design research models (see Fig. 23.2). The design cycles have the design, implementation and evaluation of the phases. From a chronological point of view, the design cycles proceed in a linear manner. Nevertheless, feedback loops exist in the practice of design research. Every event cannot be planned during the connection between science and practice. Instead, design research processes are characterized by unexpected events. It should be inspired by systematic and creative as well as intentional and inventive activities. It is important here if a planned or unexpected event has appeared or activity has been done that the implications for the design research process are reflected from a practical and theoretical perspective.

Regarding the tabletBS.dual project, for instance, not every one of the 50 schools were in the same phase. The first implementation process in some VET schools took longer. Some schools had bigger challenges with the IT infrastructure with the consequence that the design developed could not be implemented on time and had to be postponed. However, the design research process in some schools passed off smoothly, in a similar way to Fig. 23.2 above. Nevertheless, the design research process was ongoing throughout the whole project in order to develop suitable prototypes for the tablet lessons, which contributed to solving the initial problem. Theory building based on the work in the project in the sense of context-sensitive theory building – as will be shown in Sect. 23.5 – was also done using the tablet for supporting learning and simulating future digital working situations. Of course, it is not a question of a broad generalization; it is more a question of a deepening description of the educational context and what can work effectively in this case.

23.4 Methods in Design-Based Research

The evaluation of the prototypes implemented is also a central part in design research projects. This reflects the iterative idea of design, implementation and evaluation in design cycles. Research methods during the evaluation should be used integratively. It is less a question of controlling all the influencing variables but rather one of observing, describing and reflecting the interconnection between the design or prototype and its effects in the context of application (Raatz, 2015). McKenney and Reeves (2012), for instance, differentiate between three types of evaluation in design research projects:

1. *Alpha testing* refers to the early assessments of the design ideas and prototype (e.g. coherence of the design, feasibility in the field, and consideration of theoretical and practical insights). Alpha testing is anchored in an early stage of design research projects. Research methods could be the oral questioning of experts or designers.
2. *Beta testing* aims at the evaluation of the implementation of the prototype in the educational context. The focus lies on the functionality of an intervention and the interplay of intervention and context.
3. *Gamma testing* refers to the overall question: Under which conditions is an effective use of the prototype developed given? It is a matter of the final release or, more precisely, a highly stable version of the prototype.

Regarding the three types of evaluation proposed, it is obvious that evaluation in design research projects depends on the progress of the design process. It is not a question of the reproducibility of effects; it is more a question of the discursive description of the interplay between design and implementation in a given context (Reinmann & Sesink, 2014). This underlines the importance of the cooperation between researchers and practitioners. The daily life experiences of the practitioners can offer indicators for the objectives of the evaluation.

The results of design research processes also aim to offer general results. It is a matter of context-sensitive theory building. Context-sensitive means that results of a design research process do not represent 'statements of law' or empirical results for a wider range. Instead, design research results give an orientation for a different educational context. They "provide guidance and direction, but do not give 'certainties'" (Plomp, 2007). The results can be characterized as design principles. The latter "represent the supporting pillars of the interventions to be designed that substantiate (empirically, theoretically or plausibly) the fostering of the goals" (Euler, 2017, p. 4). An example of the description of design principles is: "If you want to design intervention X (. . .), then you are best advised to give that intervention the characteristic A, B and C (. . .) and to do that via procedures K, L and M, because of arguments P, Q and R" (van den Akker, 1999, p. 9). Design principles emerge during the design research process based on theoretically and empirically guided procedures. They are more prescriptive statements for a given context to achieve specified goals (Euler, 2017, p. 2). In the following section, we present the design research process in the tabletBS.dual project that also included the design principles discovered of using tablets in VET schools.

23.5 Illustration of a Design-Based Research Project in Vocational Education

23.5.1 Context and Interest of the Study

The tabletBS.dual project is structured as a design research project incorporating more than 50 vocational schools. The main interest in the project was: How to design vocational education lessons with tablets to prepare vocational students for the requirements in digitized working processes?

This interest should be seen from two perspectives:

1. How can the digital competence of the VET students be fostered regarding their future professional field and vocational profiles?
2. How can the professional development of vocational teachers be stimulated to prepare vocational students for the requirements in digitized working processes? This includes the design of suitable vocational lesson sequences with tablets and the sensitization of teachers to the future requirements of digitization in the world of work in order to incorporate them into teaching. This refers to the professional learning of vocational teachers as their daily work practice (Harteis et al., 2014, p. 3). Thus, professional learning in the design research project was embedded in the daily work practice of vocational schools and the cooperation between science and practice in the project. In the following, we focus on the professional learning of the vocational teachers (for the fostering of digital competencies of the VET students, see Gerholz et al., 2020).

Design and evaluation are cooperatively connected in four design cycles in the design research project. One design cycle is established in a school semester. Each

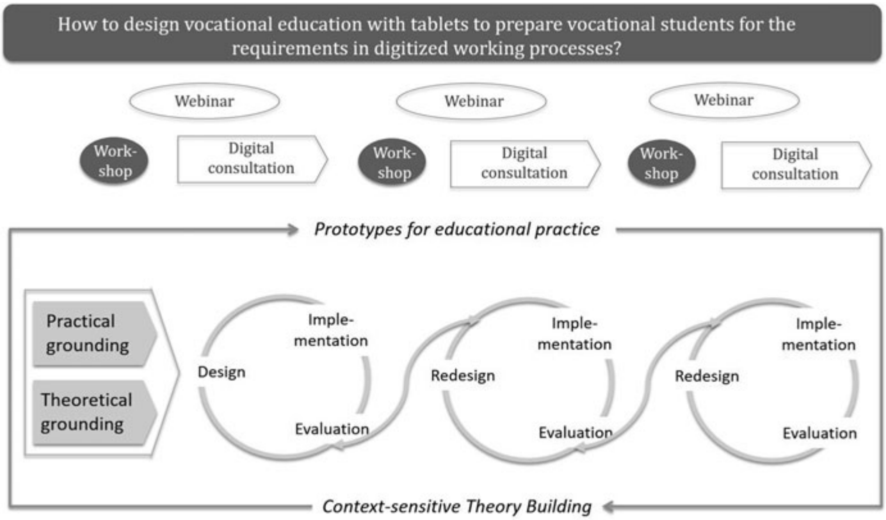

Fig. 23.3 Structure of design research in tabletBS.dual

design cycle integrates different elements of design, implementation and evaluation. Figure 23.1 gives an overview of the structure of the design research project. Various cooperation activities between science and practice were used in the design process across all design cycles (see Fig. 23.3). This involved consulting and coaching the teachers to develop prototypes for educational practice with tablets through *digital consultation sessions* and *analogue workshops*. Regarding the digital transformation, the changing working requirements can be discovered through cooperation between the learning venues – vocational school and training companies. Accordingly, *webinars* were offered to the teachers at the vocational schools and instructors in the companies to gain a mutual insight into changing working requirements in a vocational profile in times of digitalization.

New lesson sequences (3–5 vocational lessons) with tablets were developed in each design cycle in cooperation between VET teachers and researchers. These lessons were evaluated concerning experienced learning actions of VET students and the connection between the use of the tablets in the vocational lessons and their effects in the learning process. According to McKenney and Reeves (2012), this can be seen as beta testing because the effects of the lessons as an intervention can be considered more closely.

23.5.2 Theoretical Framework of the tabletBS.dual Project

The theoretical context of the design research project is based on (a) a curricular and (b) a pedagogical approach.

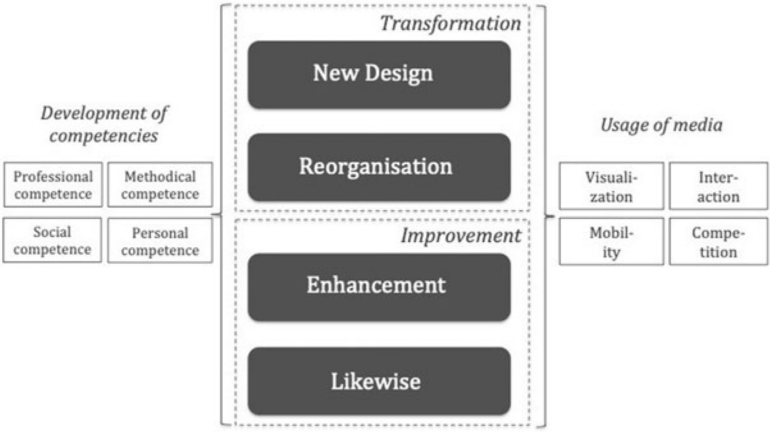

Fig. 23.4 LEaRN model

(ad a) Curricular Approach The curricula of vocational schools in the German VET system are structured in "learning areas." The latter are pedagogical reconstructions of the working fields in a given vocational profile. The requirements to fulfil a working situation are the basis for the construction of learning situations (Ertl & Sloane, 2004). Learning situations constitute a learning area and can be described as problem-based case studies. Pedagogical methods should be informed by the vocational action process. The learning process of the VET students should be a complete action. This is similar to the competence-based education approach that learning processes should be aligned with needs in a society or an economic sector (Mulder, 2012).

(ad b) Pedagogical Approach Learning situations represent, in the sense of a pedagogical adaption, the current requirements in a vocational working field. Studies have revealed that working requirements during digital transformation are changing from operative tasks to planning and controlling tasks (Fischer et al., 2018; Geisberger & Broy, 2012). Therefore, the application and use of digital devices and work techniques must be addressed and implemented at school. Design and reflection models for educational instruction are indispensable to realize this implementation of digital media into vocational educational processes (KMK, 2017).

One possible way of using digital devices such as tablets in schools is the LEaRN model (see Fig. 23.4) that combines two perspectives: On the one hand, simulating future skill requirements in a digitalized world of work with digital devices (development of competencies); and, on the other hand, using digital devices to support the learning processes (usage of media). Bearing both perspectives in mind, the intensity of digitalization in VET lessons can be described in four levels. The latter represent the changing working requirements due to digital transformation (Gerholz, 2020; Gerholz & Dormann, 2017):

1. *Likewise:* This level describes working activities previously carried out in an analogue form, which are now being replaced by digital technologies (e.g. paper-based documentation activities are now digitally implemented with digital devices).
2. *Enhancement:* At this stage, digital technologies open up possibilities of improving or enhancing working processes through digital technologies that were previously not available in an analogue form (e.g. virtual communication processes that are carried out with the help of the Internet via video telephony).
3. *Reorganization:* This includes working activities which can only be realized through digital technologies. This is the case, for instance, when previous operative activities are replaced by monitoring activities (e.g. the production process is controlled and monitored by a cyber-physical system and the employee only intervenes in the event of an error).
4. *New Design:* Completely new working scenarios are emerging at this level. Workers, located in different parts of the world, are working together in virtual teams to solve a problem in the production process, which is controlled by a cyber-physical system (e.g. virtual reality glasses enable the creation of holograms in the form of a production process).

The levels focus on the changing working requirements in a digitalized world of work. In addition, digital devices can be used to support the learning processes with (a) visualization (e.g. videos, concept maps with tablets), (b) interaction (e.g. data exchange for partial results in a problem-solving process or virtual communication in a digital role play), (c) mobility (e.g. cooperating in a group via an online system) and (d) competition (e.g. whether learning outcomes have been achieved or not can be determined, for example, via classroom response systems) (Gerholz & Dormann, 2017; Honey & Hilton, 2011).

The LEaRN model represents a heuristic which brings together the current discussion of the pedagogical potential of digital media and devices, on the one hand, and offers orientation for vocational teachers to design learning situations with digital devices in a competence-based matter, on the other. Thus, the LEaRN model is the conceptual framework in the design research project.

23.5.3 Realization of the Design-Based Research Approach

As has been shown above, the project is implemented in different design cycles focusing on design, implementation and evaluation in an iterative way. Close cooperation with teachers in all phases plays a crucial role in achieving the project aims.

23.5.3.1 Design Phases in the Cycles

The design process in tabletBS.dual is performed in cooperation with the teachers. Digital consultation sessions and analogue workshops are offered to support the teachers. According to McKenney and Reeves (2012), the workshops can be assigned, therefore, to *step one in the design process* because the teachers develop a first structure and can discuss it with other teachers and the researchers. The webinars can be attributed to *step two*. The aim is the improvement of the cooperation between the learning venues. The two learning venues – training companies and vocational schools – come together in the webinars, which are initiated and organized by the researchers, and experts of training companies clarify the influence of digitalization on the work environment and business processes. The webinars, as online in-service training, should sensitize the teachers to the vocational reality in companies. Three webinars were conducted during the design research project. The digital consultation session represents *step three*, where the researchers and practitioners, in collaboration, finalize this tablet lesson for a vocational profile. The design process focuses on the development of a vocational lesson regarding the LEaRN model. We counselled the teachers for this in digital consultation sessions in each design cycle. The discussion about the vocational lessons with tablets developed by the teachers was a central element in the design process.

23.5.3.2 Evaluation Processes in the Design Cycles

The objective of the beta testing (see Sect. 23.2) is to evaluate the interventions concerning the professional development of teachers, which are represented by (1) digital consultation sessions and (2) webinars, as well as (3) the lesson sequences implemented with tablets.

(ad 1) Digital Consultation Session The training evaluation inventory (Ritzmann et al., 2014) was adopted for the evaluation of the digital consultation sessions. After the first two design cycles, the teachers were asked about the subjective enjoyment, perceived usefulness, subjective knowledge gain and attitude towards training in order to capture the outcome dimensions of the digital consultation sessions as intervention.

(ad 2) Webinars The aim of the evaluation was to find out whether the participants were satisfied with the content of the webinars and if they could derive a benefit for their daily teaching practice with the tablets. Consequently, in addition to general questions about the organization of the webinars, questions were included about the relevance of the content and the speakers.

(ad 3) Lesson Sequences with Tablets The evaluation concept of the lesson sequences focuses on the learning situations while considering the LEaRN model and the curricular approach of problem-based learning situations. During the lesson sequences, students must answer questions about their emotional state (Schallberger,

Table 23.1 Categories for analyses of the lesson sequences with tablets

Category 1: Learning situations					
Problem-based learning and complexity of the situation	Vocational authentic situation	Action process (complete action)	Learning result oriented to the action process and vocational reality	LEaRN model – development of competencies	LEaRN model – usage of media
Category 2: Emotional state during the lesson					
Valence		Positive activation		Negative activation	
Category 3: Self-efficacy and learning conditions					
General self-efficacy	School concerning self-efficacy		Basic needs (autonomy, competence, relatedness)		Learning motivation
Category 4: Instructional design					
Subjective enjoyment	Subjective knowledge gain	Problem-based learning		Activation	Integration
Category 5: Interviews with students and teachers					
Vocational relevance	Motivation	Emotional state		Autonomy, competence, relatedness	Learning success

2005). Using a continuous state sampling method, students answer a short questionnaire on the tablet every 10 min. The questionnaire consisted of ten bipolar items with the construct's valence, and positive and negative activation. Valence reflects the students' general state of mind. Positive activation is a predictor of high motivation. Students are more likely to complete learning tasks in class successfully with high positive activation. Negative activation is more likely to be a predictor that the learning tasks they have started will cause difficulties and they are at risk of stopping learning. In addition, students were asked before and after the lesson sequences with tablets about their self-efficacy (Beierlein et al., 2012; Jerusalem et al., 2009), learning motivation and the fulfilling of the basic needs: autonomy, competence and relatedness (Deci & Ryan, 1991, 1999). These are all predictors of whether professional action competence develops (e.g. Spinath, 2011) and the learning process is successful. Furthermore, the students were asked about their perception of the lessons implemented with the tablets afterwards. The training evaluation inventory (Ritzmann et al., 2014) was adopted for this regarding subjective enjoyment and subjective knowledge gain as outcome dimensions and problem-based learning, activation and integration as design dimensions of instruction. In addition, the researchers interviewed VET students and teachers following the lesson regarding the pedagogical approach of the lesson sequences. This should give deeper insights into the perception of the lessons with tablets. Based on the various instruments, the lessons conducted with tablets were analyzed. The following analysis categories in Table 23.1 were used in a descriptive manner for this purpose.

23.6 Results of the Project

23.6.1 Results of the Digital Consultation Sessions

We had two digital consultation sessions each with almost 20 vocational schools in the school year 2018/2019. Video telephony and conferences are established digital technologies, thus, there were mostly no difficulties, and the impression was given that most of the teachers were already familiar with using the respective software. The online evaluation based on the training evaluation inventory by Ritzmann et al. (2014) of the digital consultation sessions was conducted at the end of design cycle 2. The response rate was over 90% (n = 20). The scales ranged from 1 = "does not apply at all" to 7 = "applies completely." Table 23.2 shows the descriptive analysis.

The analysis reveals that the vocational teachers perceive the digital consultation sessions as useful interventions for professional development and note them as a benefit for the design of lesson sequences with tablets.

23.6.2 Results of the Webinars

Between 34 and 54 people registered for the three webinars offered. An average of around 60% of them participated regularly. More than half of the teachers participating regularly took part in the evaluations. The evaluation instrument consisted of closed and open questions. Respondents were asked to rate the webinar on a five-level Likert scale. The scale ranged from 1 = "disagree completely" to 5 = "fully agree." Table 23.3 shows selected results concerning the aims of the webinar – the improvement of the cooperation of the learning venues.

It could be stated that the webinar presented the changes in working processes due to digitalization. The objective of improving the cooperation between the learning

Table 23.2 Evaluation of the digital consultation sessions (n = 20) (higher values on the Likert scale represent positive agreements)

Scale	Cronbachs α	M	SD	Example
Subjective enjoyment	.92	5.16	1.44	Overall, I liked the digital consultation sessions.
Perceived usefulness	.97	4.46	1.58	I find the digital consultation sessions useful for the preparation of school lessons.
Subjective knowledge gain	.85	4.43	1.77	I have the impression that my knowledge has expanded on a long-term basis through the digital consultation sessions.
Attitude towards training	.95	4.11	1.59	I will apply what I learned in the digital consultation sessions in my day-to-day work.

1 = "strongly disagree" to 7 = "strongly agree"

Table 23.3 Assessments of the relevance of the webinars (higher values on the Likert scale represent positive agreements)

Relevance for the teaching work	Mean		
	Seminar 1 (n = 24)	Seminar 2 (n = 13)	Seminar 3 (n = 9)
The content of the webinar was comprehensible.	4.08	4.54	4.11
The webinar provides an overview of changed requirements in the professional world in the course of digital transformation.	4.08	3.69	4.44
The contents of the webinar are useful for my teaching at the vocational school.	3.42	3.00	2.67
The webinar has motivated me to look deeper into the topic.	3.29	3.58	3.33
The webinar contributes to improved learning location cooperation.	3.25	3.25	2.75
The webinar made me aware of what needs to be taken more into account when educating young people.	3.38	3.31	3.38

1 = "strongly disagree" to 5 = "strongly agree"

venues cannot be regarded as fully achieved. However, it is evident that teachers still find it difficult to incorporate new working requirements into their lesson sequences. Some teachers indicated in the open questions that they would like to see concrete examples of implementations in vocational lessons, but this was not the intention of the webinars. The evaluations show that the webinars have achieved the aim to sensitize teachers to new requirements in a digitalized working world. The teachers participating were motivated to take a deeper look at digital transformation in vocational contexts. This can also be attributed to the positive assessment of the speakers, who were rated as very competent across all three webinars and were able to illustrate content well by using appropriate examples.

23.6.3 Results of the Lesson Sequences Implemented with Tablets

Based on the categories presented above in Table 23.1, the lesson sequences developed and implemented with the use of tablets were analyzed. A total of 10 different lesson sequences were analyzed. Three different profiles could be identified. The three profiles are described regarding the most relevant aspects and illustrated in more detail below.

Profile 1 – Limited Problem-Based Lessons Without Acceptance The analysis of the learning materials shows that the learning situations have almost no relation to the students' vocational activities in the training company. These are often unrealistic and inauthentic action situations and only prove to be complex and problem-based to a certain degree. They are broken up by relatively clear instructions. It appears that the students perceive these incoherencies. The post-lesson questionnaires show that the scores for integration and activation are relatively low. Students can contribute only a little of their own experience and work to a limited extent independently. This can be supported by the interview statements of the students. *"We do what the teacher says . . . So there are the tasks and we do that"* (d_SuS3). Due to the low professional relevance, the lessons are not very accepted by the pupils and have only a low motivational effect. Accordingly, the potential of the tablets could not really be exploited in lessons of Profile 1. This also goes hand in hand with a low subjective learning success. Profile 1 is evident in three of the lesson sequences analyzed.

Profile 2 – Task-Oriented Lessons with Acceptance The lessons are characterized by a stronger task orientation and less problem orientation. The learning situations represent realistic vocational action situations. In relation to the LEaRN model, the tablets are used primarily in the sense of perspective as the usage of media to support the learning process of the students and less for the development of corresponding vocational competences of a digitalized working world. The quantitative evaluations show that the students accept these lessons in terms of subjective enjoyment (e.g. independent creation of learning videos on a content area) and a subjective gain in knowledge is perceived. Self-directed learning is only promoted to a limited extent due to a task orientation with small-step, predefined sub-activities. The interviews with the students emphasize the lack of independence in the learning process due to the tasks given to achieve the learning objective. However, the high level of connection to vocational reality and the joy of working with the tablet is also made clear, as the following interview quotation shows: *"We just had to do the tasks. But the app was cool. We had to put a toilet and washbasin in there. I can show that to the customers later on"* (a_SuS3). Profile 2 can be assigned to four of the lesson sequences analyzed.

Profile 3 – Problem-Based Lessons with Acceptance The action situations of this profile have proven to be very realistic and problem-oriented in the analysis and are based on the digitally structured fields of action. Both the process of action and the action outcome are modelled with reference to the digital transformation. As the evaluations after the lessons show, the students perceive a high degree of independence in the lessons. The interviews show that the students recognize this high level of independence and problem orientation so that they are highly motivated. As the following quotation shows, the students recognize the high vocational relevance of the lessons with the tablets: *"Of course, someone from abroad can call us and ask us about the numbers. I work in an international company. Then you call . . . especially with [a tool for video telephony], you can then clarify everything"* (b_SuS6). Tablets are, therefore, used in a vocational logic of action in order to replicate the necessary competences of a digitalized working environment. Four teaching sequences can be classified under this profile.

Both administrative and technical training courses participated in the project. Furthermore, the project schools started at different times in the project and were, therefore, supported for different lengths of time through coaching and counselling in the workshops, webinars and digital consultation sessions. It is remarkable that differences between the three profiles identified can be observed in the analyses. Profile 1 is more typical for project schools in administrative training courses and with shorter support and counselling in the project. They were accompanied scientifically for only 1 year as part of the project. Profile 2 tends to be assigned to more technical training courses that were accompanied for a longer period – over 2 years. Project schools from profile 3 were also supported for 2 years in the project, but tend to be located in the administrative rather than technical area. The longer the teachers were supported and accompanied systematically within the context of the project, the better the teaching sequences developed were evaluated by the pupils. Over time, the lessons with the tablets became more problem-oriented and activating for the students, so that they were able to control their own learning process independently. This was shown clearly by the detailed analyses.

23.7 Outlook

The tabletBS.dual project enables us to illustrate how design research can work in practice and how phases of design, implementation and evaluation are meshed. Doing design research is less a linear research process and more a dynamic and often open process. This is challenging for the researchers because they have to keep both perspectives in mind: the design process and the theory-building process. Three conclusions for design research can be carved out when reflecting on the design research project presented in particular and the approach of design research in general: (a) documentation and structuring of time, (b) relevance of a design research portfolio and (c) dealing with data formats.

(ad a) Documentation and Structuring of Time Design research is characterized by a phase-oriented process – design, implementation and evaluation. Of course, the phases are clear in design research theory, but they represent more of a generic orientation. Design research projects are anchored in practice with the conditions of practice. Consequently, the coordination of the phases of the daily-life logic of the educational practice is accomplished. The conditions of the schools were heterogeneous in the tabletBS.dual project presented, which is why it could happen that the schools were in different phases (e.g. design and evaluation) at the same time and were, thus, also at different stages of development. Therefore, it is important to install a clear process documentation in a design research project in the sense of project management, especially regarding the structuring of times. The recognition of connections between the developments and evaluation data is central for the theory building. For that reason, a deeper documentation of times and events is also needed.

(ad b) Relevance of a Design Research Portfolio Design research processes present researchers with the challenge that they themselves sometimes become the object of their own research. Researchers have different roles in a design research process. They are designers, coordinators, supporters, observers, evaluators and interpreters at the same time. Bearing this in mind, it is necessary to document the traces of knowledge, which usually results in the production of documents structured in the form of texts. Sloane (2017) speaks of knowledge production. What happens in a design research project is presented in texts or documents (e.g. mails, protocols, evaluation results, design documentation). Text documents have also been produced in the tabletBS.dual project: protocols of the counselling sessions, concepts of the VET lessons with tablets, evaluation results of the VET lessons and design results during the workshops. The text documents produced in design research projects can provide a basis for generalizing the experiences, especially the experiences and insights of the researchers in their different roles. The idea of a design research portfolio can be helpful here (Gerholz, 2009). A design research portfolio is a collection of the text documents produced during a design research process and the interpretations and thoughts of the researchers in their different roles (e.g. short notes of the researchers in the portfolio). It can also be the framework for the documentation mentioned previously. Nevertheless, the question arises, how the text documents can be opened up for a (scientific) analysis especially in order to derive design principles. This refers to the next point.

(ad c) Dealing with Data Formats The design research approach can be described as a paradigm. Paradigms are constituted by scientific members of a community and what they have in common (Kuhn, 1978). They also define their rules as heuristic guidelines (e.g. hermeneutic, critical rationalism) (Euler, 2014). Bearing this in mind, the question is, how to deal with data formats in a given paradigm. When describing the impact of developments in design research processes, reference is made to methods and data formats that are also used in other research approaches (Design-Based Research Collective, 2003, p. 7). Exemplarily, quantitative (e.g. regarding self-efficacy) and qualitative data (e.g. interviews) as well as different data sources and methodical approaches (e.g. documents, statistical analyses) are used in the tabletBS.dual project. It is necessary to define how the handling of different data formats is carried out in a design research process. A point of reference is the question of what contribution a methodology used and a paradigm pursued can make to the description of the development and what potential there is for further development and theory building.

The conclusions described can be seen as food for thought for the further development of design research. The latter is a paradigm under development. As was written in the introduction, the approach of design-based research has become more widespread in the last decade. We hope that with the present chapter, the discussion and development of design research is ongoing.

References

Bannan-Ritland, B., & Baek, J. (2008). Investigating the act of design in design research: The road taken. In A. E. Kelly, R. A. Lesh, & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, mathematics and engineering* (pp. 299–319). Taylor & Francis.

Barab, S. A., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences, 13*(1), 1–14.

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). *Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen. Allgemeine Selbstwirksamkeit Kurzskala (ASKU)*. (GESIS – Leibniz-Institut für Sozialwissenschaften) Mannheim. https://www.gesis.org/fileadmin/kurzskalen/working_papers/ASKU_Workingpaper.pdf

Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation, 13*(3), 203–220. https://doi.org/10.1080/13803610701626127

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178.

Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Ed.), *Nebraska symposium on motivation: Vol. 38. Perspectives on motivation* (pp. 237–288). University of Nebraska Press.

Deci, E. L., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Easterday, M. W., Rees Lewis, D. G., & Gerber, E. M. (2017). The logic of design research. *Learning: Research and Practice, 4*(2), 131–160.

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105–121.

Ejersbo, L. R., Frolunde, L., Engelhardt, R., Hanghoj, T., Magnussen, R., & Misfeldt, M. (2008). Balancing product design and theoretical insights. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *The handbook of design research in mathematics, science and technology education, Chapter: 2: Design research and its argumentative grammar* (pp. 149–163). Routledge.

Ertl, H., & Sloane, P. F. E. (2004). A comparison of VET structures in Germany and England: Contexts of complex teaching-learning arrangements. In R. Mulder & P. F. E. Sloane (Eds.), *New approaches to vocational education in Europe* (pp. 27–42). Symposium Books.

Euler, D. (2012). Unterschiedliche Forschungszugänge in der Berufsbildung: Eine feindliche Koexistenz? In E. Severing & R. Weiss (Eds.), *Qualitätsentwicklung in der Berufsbildungsforschung* (pp. 29–46). W. Bertelsmann.

Euler, D. (2014). Design-research – A paradigm under development. In D. Euler & P. F. E. Sloane (Eds.), *Design-based research* (pp. 15–44). Franz Steiner Verlag.

Euler, D. (2017). Design principles as bridge between scientific knowledge production and practice design. *EDeR – Educational Design Research, 1*(1), 1–15.

Euler, D., & Sloane, P. F. E. (Eds.). (2014). *Design-Based Research: Beiheft Zeitschrift für berufs- und Wirtschaftspädagogik*. Franz Steiner.

Finch, D., Falkenberg, L., McLaren, P. G., Rondeau, K. V., & O'Reilly, N. (2018). The rigour–relevance gap in professional programmes: Bridging the 'unbridgeable' between higher education and practice. *Industry and Higher Education, 32*(3), 152–168.

Fischer, C., Goller, M., Brinkmann, L., & Harteis, C. (2018). Digitalisation of work: Between affordances and constraints for learning at work. In D. Ifenthaler (Ed.), *Digital workplace learning: Bridging formal and informal learning with digital technologies* (pp. 227–249). Springer.

Geisberger, E., & Broy, M. (2012). *Acatech Studie: Integrierte Forschungsagenda – Cyber-physical systems*. Online: https://www.bmbf.de/files/acatech_STUDIE_agendaCPS_Web_20120312_superfinal.pdf (02 Sept 2021).

Gerholz, K.-H. (2009). Das Forschungsportfolio – Ein Dokumentator und Reflektor im Rahmen von qualitativen Sozialforschungsprozessen. In H.-H. Kremer & P. F. E. Sloane (Eds.), *Paderborner Forschungs- und Entwicklungswerkstatt* (pp. 65–88). Band II. Paderborn: Eusl.

Gerholz, K.-H. (2014). Selbstreguliertes Lernen gestalten – Darstellung eines Design Research-Zyklus in der wirtschaftswissenschaftlichen Hochschulbildung. In D. Euler & P. F. E. Sloane (Eds.), *Design Research: Beiheft der Zeitschrift für Berufs- und Wirtschaftspädagogik* (pp. 215–229). Franz Steiner.

Gerholz, K.-H. (2020). Unterrichtsarbeit an beruflichen Schulen im Zuge der digitalen Transformation – Ein fachdidaktisches Modell für den Einsatz digitaler Medien. In U. Buchmann & M. Cleef (Eds.), *Digitalisierung über berufliche Bildung gestalten* (pp. 169–180). wbv.

Gerholz, K.-H. (2022). Berufliche Bildung. In H. Reinders, D. Bergs-Winkels, I. Post, & A. Prochnow (Eds.), *Handbuch empirische Bildungsforschung: Eine elementare Einführung*. Springer.

Gerholz, K.-H., & Brahm, T. (2014). Apprenticeship and vocational education. In C. Harteis, A. Rausch, & J. Seifried (Eds.), *Discourses on professional learning: On the boundary between learning and working* (pp. 143–158). Springer.

Gerholz, K.-H., & Dormann, M. (2017). *Ausbildung 4.0: Didaktische Gestaltung der betrieblich-beruflichen Ausbildung in Zeiten der digitalen Transformation*. bwp@ Berufs- und Wirtschaftspädagogik – Online, Ausgabe 32.

Gerholz, K.-H., Ciolek, S., & Wagner, A. (2020). Digitales Lernen an beruflichen Schulen wirksam gestalten: Erste Ergebnisse der wissenschaftlichen Begleitung von tabletBS.dual. *SchulVerwaltung BW, 4*(7–8), 217–221.

Harteis, C., Rausch, A., & Seifried, J. (2014). Discourses on professional learning: On the boundary between learning and working. In C. Harteis, A. Rausch, & J. Seifried (Eds.), *Discourses on professional learning: On the boundary between learning and working* (pp. 1–7). Springer.

Honey, M., & Hilton, M. (Eds.). (2011). *Learning science through computer games and simulations*. The National Academies Press.

Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W., & Röder, B. (2009). *Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht*. Skalen zur Erfassung von Lehrer- und Schülermerkmalen. https://www.erziehungswissenschaften.hu-berlin.de/de/paedpsych/forschung/Skalenbuch_FoSS.pdf

Kieser, A., & Leiner, L. (2009). Why the rigour-relevance gap in management research is unbridgeable. *Journal of Management Studies, 46*(3), 516–533.

KMK: Kultusministerkonferenz. (2017). *Bildung in der digitalen Welt: Strategie der Kultusministerkonferenz.*

Kuhn, T. S. (1978). *Die Entstehung des Neuen: Studien zur Struktur der Wissenschaftsgeschichte*. Suhrkamp.

McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research: What it is, how we do it, and why*. Routledge.

McKenney, S. E., & Reeves, T.C. (2018). *Conducting educational design research: What it is, how we do it, and why*. 2nd edition. Routledge.

Mulder, M. (2012). Competence-based education and training. *The Journal of Agricultural Education and Extension, 18*(3), 305–314.

Perna, L. W. (2016). Throwing down the gauntlet: Ten ways to ensure that higher education research continues to matter. *The Review of Higher Education, 39*(3), 319–338.

Plomp, T. (2007). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 9–36). Netherlands Institute for Curriculum Development (SLO).

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research – Part A: An introduction* (pp. 10–51). Netherlands Institute for Curriculum Development (SLO).

Raatz, S. (2015). *Die Entwicklung von Einstellungen gegenüber verantwortungsvoller Führung: eine Design-based Research Studie in der Executive Education*. Springer VS.

Reeves, T. (2006). Design research from a technology perspective. In J. V. D. Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52–66). Routledge.

Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtswissenschaft, 1*, 52–69.

Reinmann, G. (2014). Welchen Stellenwert hat die Entwicklung im Kontext von Design Research? Wie wird Entwicklung zu einem wissenschaftlichen Akt? In D. Euler & P. F. E. Sloane (Eds.), *Design-based Research: Zeitschrift für Berufs- und Wirtschaftspädagogik (Beiheft)* (pp. 63–78). Franz Steiner.

Reinmann, G., & Sesink, W. (2014). Begründungslinien für eine entwicklungsorientierte Bildungsforschung: Methodologie und Methoden medienpädagogischer Forschung. In *Jahrbuch Medienpädagogik 10* (pp. 75–92). VS-Verlag.

Ritzmann, S., Hagemann, V., & Kluge, A. (2014). The Training Evaluation Inventory (TEI) – Evaluation of training design and measurement of training outcomes for predicting training success. *Vocations and Learning, 7*(1), 41–73.

Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS): Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische Eigenschaften bei der Beschreibung intra- und interindividueller Unterschiede*. Forschungsbericht aus dem Projekt: Qualität des Erlebens in Arbeit und Freizeit. Online: http://www.psychologie.uzh.ch/institut/angehoerige/emeriti/schallberger/schallberger-pub/PANAVA_05.pdf (04 Mar 2020).

Schwartz, D. L., Chang, J., & Martin, L. (2005). *Instrumentation and innovation in design experiments: Taking the turn towards efficiency*. Stanford University. Online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.4720&rep=rep1&type=pdf (20 Oct 2021)

Sloane, P. F. E. (2017). 'Where no man has gone before!' – Exploring new knowledge in design-based research projects: A treatise on phenomenology in design studies. *EDeR – Educational Design Research, 1*(1), 1–29. https://doi.org/10.15460/eder.1.1.1026

Spinath, B. (2011). Lernmotivation. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Eds.), *Empirische Bildungsforschung: Gegenstandsbereiche* (pp. 45–55). VS Verlag für Sozialwissenschaften.

Van den Akker, J. (1999). Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1–14). Kluwer Academic.

Van den Akker, J. (2007). *Curriculum design research: An introduction to educational design research SLO* (pp. 37–52). Netherlands Institute for Curriculum Development.

Karl-Heinz Gerholz (PhD) is a Professor of Business and Human Resource Education at the University of Bamberg (Germany). His research interests are related to issues of digitalization in vocational education, democracy education, educational management and Design based Research. He acts as a reviewer for different national and international journals in the educational area.

Anne Wagner (MSc) is research assistant in field of Business and Human Resource Education at the University of Bamberg (Germany). Her research interests are related to issues of school development, digitalization in vocational education and Service Learning in Higher Education. She acts as a reviewer for different journals in the educational area.

Chapter 24
Change Laboratory Method for Facilitating Transformative Agency and Collective Professional Learning – Case from a Finnish Elementary School

Anu Kajamaa and Sakari Hyrkkö

Abstract The aim of our study is to enrich the current understanding of teachers' collective professional learning and development by introducing a method called the Change Laboratory (CL), which is a participatory work development method based on theory of expansive learning. We present and empirically illustrate how the method was employed among a group of 12 teachers and a principal for facilitating transformative agency, and how it contributed to their collective professional learning and the development of new practices in the school. Our interactional video data consists of six CL meetings and a follow-up meeting held in an elementary school which operates as one of the pioneering university-level teacher training, research and development units in Finland. The CL provided the participants with research-assisted space and tools to collectively analyse and redesign their work activities. In a dialectical learning process, the CL first triggered the surfacing of problems and tensions in the school community, then supporting the conceptualisation of problems as stemming from systemic organisational contradictions, and eventually leading to the creation of novel solutions to resolve the contradictions and transform the school's practices. Our analysis illustrates the significance of transformative agency for turning organisational contradictions as drivers for collective professional learning and development, in a bottom-up process of organisational renewal. Our findings also demonstrate the demands that such a collective process of professional learning poses for teachers, requiring continuous efforts and cultural change.

A. Kajamaa (✉)
Faculty of Education, University of Oulu, Oulu, Finland
e-mail: anu.kajamaa@oulu.fi

S. Hyrkkö
Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland
e-mail: sakari.hyrkko@helsinki.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_24

Keywords Change Laboratory · Transformative agency · Teacher agency · Professional learning and development · Expansive learning · Curriculum · Cultural-historical activity theory (CHAT)

24.1 Introduction

In Western societies, professional learning and development have traditionally been viewed as linear processes of problem-solving in which individuals gain increasing mastery of the culturally available professional knowledge and skills (e.g., Wilson & Berne, 1999; Borko, 2004; Opfer & Pedder, 2011). In recent years, emphasis has increasingly been placed on the collective endeavours of teams and organisations to develop their practices (Cordingley, 2015) and the relationship between individual and collective learning, such as in the form of professional learning communities (Van Meeuwen et al., 2020; Admiraal et al., 2021). However, many professional learning and development efforts for teachers still focus on acquiring tried and tested skills and competencies seen as relevant for adapting to changes and reforms, based on an overly simplified conception of the teaching profession (see Winch et al., 2015). Unfortunately, this adaptive approach does not pay sufficient attention to historical and contextual factors (Miettinen, 2013), specifically, the cultural-historical development of the schooling system and the wide variety of schools, each constituting a distinct context. It also tends to disregard the mediated, transformative, participatory and situated dimensions of collective professional learning (Lattuca, 2002). To enable practice transformation, it is pivotal for school-based learning environments to promote collective analysis of interaction, contradictions and construction of the context (Engeström, 2009). Therefore, attention should be directed to participatory methods of work development for facilitating expansive transformations in which the community of professionals learns to widen its possibilities for action by redesigning its own activity.

In this chapter, our aim is to enrich the current understanding of teachers' collective professional learning and development by introducing a method called the Change Laboratory (CL) (Engeström et al., 1996; Virkkunen & Newnham, 2013), which aims to support collaborative learning in and transformation of work activities and organisations. Our chapter empirically illustrates this method, documenting a case from an elementary school which engaged in a participatory workplace intervention to transform their local practices. The method is based on the dialectical tradition of cultural historical activity theory (CHAT) and its application, the theory of expansive learning (Engeström, 2015). Dialectical studies typically analyse organisational phenomena, such as systematic relationships, uncover tensions and contradictions, and detect opportunities and mechanisms for change at the multiple levels on which participants arrange their activities (Benson, 1977). In this tradition, the identity of any activity is primarily determined by its object, which includes a collective motive for the activity and emerges when human needs and the material-cognitive formations of the world meet (Leont'ev, 1978). The CL method treats historically evolved organisational contradictions within and between activity systems as drivers for expansive learning, that is a collective, non-linear learning process without predetermined end results (Engeström, 2015). The CL method provides a set of instruments for resolving the contradictions and reconceptualising the object of the collective activity via dialogue, analysis and innovating new forms of activity. The method follows the idea of equal participation of the employees in the analysis of their own work activity and aims at system-level transformations. (See Virkkunen & Newnham, 2013).

The CL case presented in this chapter was carried out at an elementary school, which is one of the pioneering university-level teacher training, research and development units in Finland with nearly 1000 students, 250 trainee teachers and 110 staff. At the time of the CL in 2015, a nationwide curriculum reform (which is instigated every 10 years) was in its early implementation phase, and the teacher training schools were expected to act as forerunners in this process, to guide and encourage local implementation efforts in schools across Finland. The transformational 2016 curriculum reform (Finnish National Agency for Education, 2014) was piloted in 2015 at the school in which the CL was conducted. The new curriculum introduced interdisciplinary learning outcomes called "transversal competence", a Finnish interpretation of the OECD's twenty-first century skills framework, along with a requirement for multidisciplinary and collaborative teaching and studying. The reform, a top-down regulatory process, created an acute need for enhanced collaboration among the teachers, necessitating the crossing of the traditional boundaries of the teacher community and individual teachers' work. For the teachers with a historically developed substantial autonomy over their own instruction and pedagogical choices, this called for locally transforming both their collective and individual practices.

Responding to the demanding educational change requirements and taking initiatives and actions to tackle such challenges calls for teacher agency (Kumpulainen et al., 2018), which is also a particularly useful concept for studying and explaining teachers' professional learning and development during such transformations (e.g., Goller & Paloniemi, 2017; Brodie, 2021). We therefore draw from a CHAT-based theory and method developed by Haapasaari et al. (2016) and Sannino (2015) to analyse expressions of transformative agency in the discourse of the CL participants (12 teachers and a principal) within the sociocultural context of an elementary school. We view transformative agency as breaking away from a historically formed frame of action (Virkkunen, 2006) and collectively taking agentive actions to address and overcome its inherent contradictions (Haapasaari et al., 2016). We examined the participants' transformative agency as a discursive, collective, continuous, non-linear and tension-laden process, closely related to its sociomaterial context and practical actions (see also Kajamaa & Kumpulainen, 2019). More precisely, using the framework developed by Haapasaari and colleagues, we traced the types of transformative agency in the participants' discourse, namely *Resisting*, *Criticising*, *Explicating*, *Envisioning*, *Committing to actions* and *Taking actions*. Further, we paid special attention to the physical interaction and the usage and collective constructing of artefacts and representations mediating the qualitative transformation of the participants' work activity (also Engeström et al., 2015; Sannino et al., 2016).

Our analysis was guided by the following research questions: What types of expressions of transformative agency can be found in the CL discussions? and How does transformative agency connect to teachers' collective professional learning and development in the CL?

Our analysis makes the case that the CL enabled the participants to articulate and engage with the organisational contradictions, aggravated by the curriculum reform and challenging their current practices, and to analyse and redesign their work activities collectively. Our findings illustrate the explanatory power of transformative agency in the analysis of expansive transformations, opening a new perspective on the little studied interplay between individual and collective learning. Moreover, emphasising collective actions and systemic change, emerging and evolving over time, it takes us beyond the examination of individual experts and their situational actions (Haapasaari et al., 2016), expanding the current understanding of professional learning and development. However, this expansive learning process represents a unique form of collective professional learning, and it was not without tensions.

24.2 The Change Laboratory Method

The Change Laboratory (CL) is a research-assisted workplace intervention method developed in the 1990s at the University of Helsinki (Engeström et al., 1996; Virkkunen & Newnham, 2013), and has thereafter been used and developed further by academics and practitioners globally. CL projects have been conducted in various fields, including health care, social welfare, media, industry, retail, banking and insurance (Virkkunen & Newnham, 2013), and in school contexts (e.g., Engeström et al., 2002; Sannino, 2008; Virkkunen & Tenhunen, 2010; Morselli & Sannino, 2021; Hyrkkö & Kajamaa, 2021; Rainio & Hofmann, 2021).

The CL method draws from cultural-historical activity theory (Vygotsky, 1978; Leont'ev, 1978) and especially from the theory of expansive learning (Engeström, 2015; Engeström & Sannino, 2010). It is a formative intervention method, referring to an emergent, developing and open-ended approach to theoretically guided research (Engeström, 2011). The CL is a research-supported method designed to respond to local needs and to support expansive learning and redesign and development of collective activity. The CL provides a set of instruments for innovating through expansive learning (Engeström, 2015). In this process, the end results of learning are not predetermined by the interventionists but instead the outcomes are designed by the participants as they work out expansive solutions to the developmental contradictions in their activity systems (Virkkunen & Newnham, 2013).

It is also a participatory method highlighting that each participant's voice, knowledge and expertise counts in the collective analysis and development of activity. Ideally, the participants will be drawn from all organisational levels and professional groups, and it is recommended that there not be more than 20 participants. A series of weekly meetings (usually a maximum of 10 sessions) takes place

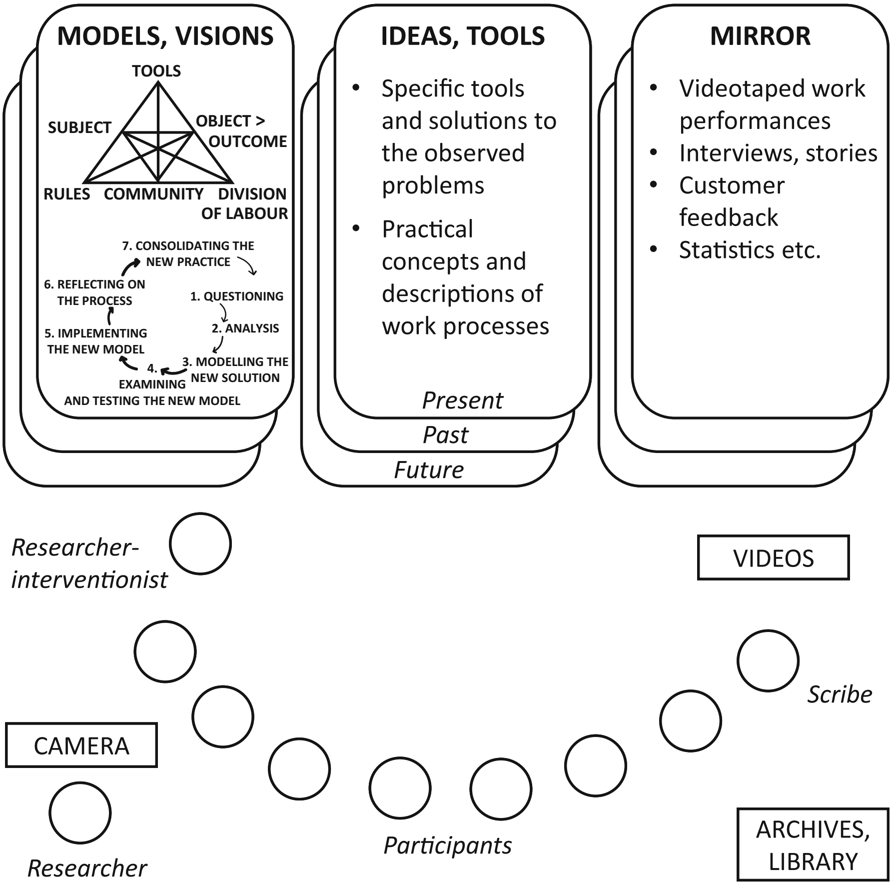

Fig. 24.1 Prototypical layout of Change Laboratory. (Modified from Engeström et al., 1996, 11; Kerosuo et al., 2010, 118; see also Virkkunen & Newnham, 2013, 16)

in a room situated at the workplace and equipped with a set of instruments to promote collective learning. At each meeting, participants are appointed to write down ideas on a set of surfaces (i.e., flipcharts) (see Fig. 24.1), and to take notes for a memo. The CL meetings generate rich discursive data (including video data) which the research team can analyse between the meetings, to prepare for subsequent meetings. Typically, one or two follow-up meetings are held some months after the main meetings to gain insights on the implementation of the new ideas and practices. Then, after all the meetings have been held, the data are typically transcribed and further analysed by applying activity-theoretical concepts, such as the notion of transformative agency.

A key activity-theoretical concept behind the CL method is the dialectical principle of learning and development of *ascending from abstract to concrete* (Davydov, 1990). 'Abstract', in this sense, means something (experiences, ideas, disturbances, problems, worries or innovative solutions) that are separated from their historical and systemic context, whereas 'concrete' refers to an understanding of these phenomena as materially contextualised: in their historical and systemic context as part of the activity system (Engeström & Sannino, 2010). This distinction between dialectical abstractness and concreteness is visible in the CL method in the three surfaces, in which the rightmost MIRROR board is used for representing systemically uncontextualised ('abstract') excerpts from the activity, such as video clips of the actual work situations of the CL participants, as well as statistical information (e.g., organisational key figures), stakeholder interviews and case examples (recorded by the researcher/facilitators) that reveal problems in the organisation and enable analysis of the activities in the CL meetings (see Virkkunen & Newnham, 2013).

In a CL, the process of *double stimulation* (Vygotsky, 1978; see also Sannino, 2015) is another grounding concept, employed to promote transformative agency among the participants (see e.g., Morselli & Sannino, 2021) with the help of "mirror data" presented on the MIRROR board. Typically, the mirror data are at the centre of the participants' attention during the first CL meeting, to stimulate reflection and discussion, and to reveal conflicts of motives (first stimulus: a consciousness of a problem in need to be solved) in the work community. This includes both the joint explication of an ambiguous situation, and the emergence of a collective will (Sannino, 2015) to influence or change the pressing situation, for example teachers' need for change in a situation in which they wish to provide high quality teaching but feeling that the prevailing school practices and management are making this impossible. As the CL progresses, the researcher can introduce theoretical concepts to the participants on the MODELS, VISIONS board, to support their analysis of the mirror data, widening their understanding of the first stimulus, and the systemic nature of their activity (representing the dialectically 'concrete'). The participants may then create auxiliary motives by means of talk or action (second stimulus), such as envisioning alternative ways of doing their work, that provide tentative solutions to cope with the critical situations (see e.g., Sannino, 2015).

A theoretical model typically used to this end in the CL is the dynamic model of the *activity system* (Engeström, 2015). It helps participants conceptualise the structure of their collective activity – consisting of complementary elements that mutually define and require each other: subject, object and community. Originating partly from Vygotsky's ideas (1978), the interactions between these elements are viewed as being possible only through mediation by culturally and historically formulated signs and tools (subject-object mediation), rules (subject-community-mediation) and a division of labour (community-object-mediation) (Engeström, 2015). Development and learning can thus be described as systemic transformation of the activity, which can take place through analysing and redefining these elements and their relationships, in other words, a process of remediation. Changes in any one element, or in the relationship between elements, also impacts other elements and their relationships. Furthermore, in today's networked practices and organisations, a minimum of two interacting activity systems is usually considered the unit of an analysis. (Engeström, 2015; Virkkunen & Newnham, 2013).

The MODELS, VISIONS board can also be used for other theoretical and conceptual tools that the researcher-facilitators see necessary and useful, such as *the model of the phases of expansive transformation of an activity* (introduced by Engeström, 2015). The purpose of the theoretical tools, especially the *activity system* model, is to help the participants to understand the systemic nature of their activity, and to provide a historical and systemic contextualisation to the contradictions behind the disturbances and everyday problems of their work, displayed on the MIRROR board.

During the CL, the participants can gradually begin to grasp and analyse a specific *contradiction* (Il'enkov, 1977), or contradictions, in their historically evolved activity system, thus 'ascending from abstract to concrete'. Although the systemic contradictions are impossible to address "directly", they are manifested in the participants' discourse in several ways (Engeström & Sannino, 2011) and can thus be analysed and turned into drivers for change and learning during the CL process. The systemic analysis and understanding of the work activity and its systemic contradictions is pivotal for dialectic learning in a CL. Without this, organisations usually end up creating targeted solutions to only tackle the immediately observable disturbances, in effect sub-optimising individual parts of the activity system. These solutions, which do not address the underlying systemic contradiction(s) behind the disturbances, often fail in the long term (e.g., Engeström, 2011).

During the CL process, collective learning is achieved through talk (e.g., Haapasaari et al., 2016) and experimentation with newly created tools and models for renewed work practices (Engeström et al., 2007). The CL process ideal-typically follows the phases of expansive transformation of an activity (i.e. the cycle of expansive learning) and it thus supports the researcher/facilitator in planning and executing the CL process to follow a series of epistemic learning actions, namely: *questioning* the current work practices, *analysing* tensions and contradictions in the work activity, *modelling* a new solution in an observable and transmittable form, *examining the model* to grasp its potentials and limitations, *implementing the model* through practical applications and enrichments, *reflecting* on the learning process and, finally, *consolidating* the outcomes into a new form of practice. Iterative transitions back and forth between individual actions (micro-cyclicity) are typical for an expansive learning process (Engeström et al., 2007; Kajamaa, 2011). The expansive learning process fundamentally aims at resolving the underlying systemic contradiction(s) and developing better work practices. Expansive learning is a process of construction and resolution of successive contradictions, which eventually may lead into the qualitative transformation of all the elements of the activity system, eliciting system-level changes in the organisation and qualitative transformation in the object of the activity (Engeström, 2015).

The third surface in the middle is reserved for IDEAS and TOOLS, such as new concepts and forms of activity, collectively created by the practitioners when analysing problematic situations during the CL process. The ideas and tools also function as a second stimulus, providing the practitioners with shared instruments to overcome the contradictions and to transform their circumstances and the work activity. The innovation of new ideas and practices requires moving between the experimental MIRROR and the theoretical MODELS and VISIONS (see Virkkunen & Newnham, 2013).

Expansive learning and transformative agency are theoretically and conceptually closely connected, even inseparable: they both require and enable each other in the sense that "expansive learning within the CL is indeed a process of formation of transformative agency" (Haapasaari et al., 2016, 243). Furthermore, both expansive learning and transformative agency are always connected to a specific contradiction in a specific historically evolved activity system. They enable overcoming the contradiction by introducing new, more adequate mediator(s) (e.g., new cultural tool or practice), to transform the collective activity and its local sociomaterial context. The newly created solutions have, for example, enhanced knowledge sharing, collaboration and understanding of the shared object of the activity system(s) in a novel way (Haapasaari et al., 2016). This process requires time and continuous efforts, as transformative agency "develops and is maintained in collective interaction over time when agentive actions gain their meaning, their consequences and their continuity in the interplay between individuals and their collective" (Engeström, 2007).

24.3 Conducting a Change Laboratory and Analysing CL Data

In this section, we describe an empirical case in which the CL method was applied in an elementary school to help a teacher community remediate their activity, riddled with tensions and conflicts, in the wake of a transformational curriculum reform. We also illustrate the analysis of CL data by describing how we executed our analysis in this case.

In early 2015, our team of researchers wanted to study how the school community was experiencing the new curriculum reform. After meeting the principal, we began to interview those interested in sharing their thoughts. The interviewees (a principal, three teachers and a pre-service teacher) described a multiplicity of tensions in the school community, interpreted by us as disturbances of the activity, such as lack of communication, commitment and sense of community as well as constant haste, aggravated by the upcoming curriculum reform. In the interviews, a need for a collective development effort to tackle these issues also came up. We therefore agreed with the principal and the teachers we interviewed that a series of CL meetings would be carried out at the school, facilitated by our research team. Participation was voluntary, and the weekly meetings would be held on the school's premises.

The videotaped interviews were edited by the research team to be used as mirror data to stimulate discussion and prompt collective analysis of the activity. Mirror

Table 24.1 Overview of the Change Laboratory meetings, discussion topics and mirror data

	Speaking turns	Attendees	Main discussion topics	Mirror data and stimuli provided by researchers
Meeting 1	575	2 new teachers 6 senior teachers	Introduction to the CL method and theory of expansive learning Voicing tensions and challenges in the current activity	Overview of the CL process
Meeting 2	564	1 new teacher 7 senior teachers 1 principal	Instructional leadership Need for a shared direction and collaboration Historical development of the organisation and its leadership	Video clips and summarising slides from meeting 1 Overview of the CL process
Meeting 3	590	1 new teacher 6 senior teachers 1 principal	Possibility of a shared vision Ideas for organised collaboration	Video clips from meeting 2
Meeting 4	592	2 new teachers 4 senior teachers	Historical analysis of the current organisation and activity Teams as a tool for collaboration Steering group as a potential new leadership structure	Video clips from meeting 3 Historical timeline of the school's development
Meeting 5	782	1 new teacher 7 senior teachers 1 principal	Roles and responsibilities of organisational entities Reflection of earlier practices Need for a concrete model	Video clips from meeting 4
Meeting 6	1491	1 new teacher 5 senior teachers	Incorporating earlier ideas in models Teamwork and leadership structures, roles and responsibilities	Summarising slide of earlier discussion points Prompt to draw models
Follow-up meeting	745	1 new teacher 6 senior teachers 1 principal	Reflection of collaboration and autonomy in daily work Need to refine the model	Overview of the six meetings Phases of expansive transformation of activity

data were also gathered by videotaping classroom situations and pre-service supervision meetings for sequences in which the disturbances were visible in everyday work (see Table 24.1 above for an overview of the mirror data and stimuli used in each meeting).

24.3.1 The Change Laboratory Meetings

Our data comprised six videotaped CL meetings, recorded with two video cameras on opposite sides of the seminar room at the elementary school. The meetings took place once a week in April and May 2015. Each meeting lasted between 93 and 103 min, and between six to nine of the 13 participants attended each meeting. The video recordings were transcribed verbatim, resulting in 276 pages of transcription consisting of 4594 speaking turns. In addition to these meetings, a seventh follow-up meeting was held in February 2016, consisting of 745 speaking turns when transcribed verbatim. Table 24.1 indicates the number of speaking turns and attendees of each meeting, along with a summary of the most central discussion topics and the stimuli used by the research team to facilitate the process. As the table shows, towards the end of the process, especially in the final session, there were significantly more speaking turns than in earlier sessions, implying shorter turns, more overlapping speech and less silence.

Each excerpt presented in the findings section of this chapter features a number indicating the speaking turn in our transcribed data and one of the following codes for the participant making the remark: senior teacher (ST), new teacher, employed at the school for less than a year (NT), principal (PR), researcher (RR) and undefined person (U), replacing their names and accompanied by a running number.

24.3.2 Data Analysis

We analysed the CL participants' discourse and actions on two levels. First, we conducted a detailed analysis of the types of expressions of transformative agency in the teachers' discourse using the analytical framework put forward by Haapasaari et al. (2016). Secondly, we analysed the dialectical progress of the participants' collective professional learning and development during the transformation efforts of their collective activity in the CL, relating our observations to the analysis of transformative agency and the theoretical and conceptual framework on which the CL method is based (see Sect. 24.2).

For the first level of analysis, we identified expressions of the six types of transformative agency (Haapasaari et al., 2016), namely *Resisting*, *Criticising*, *Explicating*, *Envisioning*, *Committing to actions* and *Taking actions* in the video data and the transcription of the meetings. The number and proportion of occurrences of each type of transformative agency along with definitions and illustrative examples from our data are listed in Table 24.2 (N.B., the expression of *Taking actions* did not appear in our CL case). The sequences of interaction containing an expression of transformative agency were then analysed. This was not always straightforward as people tend to speak over each other, get interrupted and then continue the thought, forget what they were about to say, hesitate, and so forth. Sometimes the collective creative process obscures the intentions of earlier speakers

Table 24.2 Types of expressions of transformative agency found in the data

Type of expression	Definition (Haapasaari et al., 2016)	Examples from data
Resisting 24 expressions during the CL (6.8% of all expressions)	Resisting the change, new suggestions or initiatives. Directed at management, co-workers or the interventionists.	*We'll soon have more enemies than friends in this work community [---] you are some miracle-makers if you can solve our problems.* (Meeting 1)
Criticising 93 expressions (26.4%)	Criticising the current activity and organisation. Change-oriented and aiming at identifying problems in current ways of working.	*I think many of us have become. . . saturated with the constant haste [---] teachers are supposed to be in two places at the same time, so the whole school is based on this systemic fault, the haste is in-built.* (Meeting 1)
Explicating 113 expressions (32.1%)	Explicating new possibilities or potentials in the activity. Relating to past positive experiences or former well-tried practices.	*You can develop structures in so many ways. In my previous school [---] we had class-based teams [---] they were scheduled [---] and the principal attended all meetings.* (Meeting 3)
Envisioning 104 expressions (29.6%)	Envisioning new patterns or models in the activity. Future-oriented suggestions or presentations of a new way of working.	*I think we should have a steering group with teacher members [---] and the steering group member then goes to the teams and initiates the development there.* (Meeting 5)
Committing to actions 18 expressions (5.1%)	Committing to taking concrete, new actions to change the activity.	*People have expressed the need to get more involved, so why don't we try this!* (Meeting 5)

(Hyrkkö & Kajamaa, 2021), retrospectively altering the meaning of an individual utterance in the overall context. Therefore, assigning a type of transformative agency required constant interpretation of individual utterances in relation to the discussion in general.

In the analysis we relied heavily on the video data, using the transcription mostly as a coding aid. Most often, an expression of transformative agency could be assigned for a single speaking turn. If the same speaker repeated the same idea or thought after another speaking turn (for example, interruption), this repetition would not be regarded as another expression of transformative agency, because the speaker's intention remained the same. Sometimes, a single speaking turn contained several consecutive, differing expressions of transformative agency. Sometimes, a fast-paced, like-minded dialogue with overlapping speech from several participants was interpreted to contain only one expression – as the intention and meaning were shared. Alongside discourse, we paid special attention to the participants' physical interaction and their collective construction and use of novel mediating artefacts and representations.

24.4 Findings: The Change Laboratory Facilitating Transformative Agency and Collective Professional Learning

In this section, we describe the CL process and present data examples of the expressions of transformative agency, illustrating how transformative agency contributed to the dialectical progress of collective professional learning and development.

24.4.1 Meeting 1 – Resisting and Criticising the Current Activity

At the start of the first meeting, the researchers introduced the Change Laboratory method and the theory of expansive learning to the participants, describing them as tools for aiding the work community in developing its practices. This prompted two types of responses from the participants: firstly, doubting that the CL could have any effect on the work community, and secondly, bringing up tensions and disturbances in the work activity. N.B., the principal was not present in the first meeting due to a double booking.

To the researchers' surprise and as a deviation from their planned "script" for the meeting, the teachers quickly began *Criticising* their current activity, articulating the disturbances hampering their work in an analytical manner. Consequently, most of the video excerpts of interviews and classroom situations were not shown as mirror data, and the researchers allowed the conversation to flow freely. During the meeting, the appointed board scribes used the MIRROR board to write three full flip charts of issues under the headline *"Challenges"* and on the IDEAS, TOOLS board only one sentence: *"Could structures be changed?"*

To summarise, the *"Challenges"* listed were (1) Individual interests and differing developmental needs; (2) Lack of leadership in development work; (3) Lack of organisation of work; (4) Lack of instructional leadership; (5) Lack of structures of teamwork; and (6) The "responsibility paradox" (feeling responsible for one's own instruction but being unable to extend the sense of responsibility to the level of community). These aptly worded disturbances and conflicts formed the basis of the analysis of the work activity for the entire CL.

In analysing the problems in the school's practices, the participants interpreted them as stemming from the historically developed strong autonomy of individual teachers (and the organisational practices evolved from it and to support it), and the constantly evolving object of activity in schoolwork as witnessed in the changing curricular demands. The organisation of work based on teacher autonomy was visible in the ambition of individuals to do their work as well as possible but, on the other hand, in the lack of willingness, opportunity or ability to share tasks, resulting in short-lived, individual instructional solutions, metaphorically expressed as *"pedagogical butterflies" (174, ST6, Criticising).*

The participants also saw the pressing challenges as stemming from the diverse duties of the university-level teacher training school, contributing to the multiplicity of tasks and uneven workloads, combined with the lack of effective organisational structures for distributing tasks. Behind many of these challenges was a yearning for an educational vision for the school, as first uttered by a senior teacher: (*"We are missing a uniting vision, educational vision, or actually envisioning on any level. This is probably the one major issue causing the incoherence here." 218, ST4, Criticising*). At this point, the establishment of an educational vision was seen as the key task of school leadership. Leadership issues thus became a central topic in the second meeting, with the principal also present.

Dominated by two forms of transformative agency, *Resisting* and *Criticising*, this phase of the CL initiated a shift from individually experienced difficulties toward collective professional learning and development. Already at this early phase of organisational renewal, the teachers began to view the organisational issues as systemic, leading to the realisation of a collective need for change.

24.4.2 Meeting 2 – Elaborating and Structuring the Challenges: Criticising and Explicating

To employ the conflicts and tensions voiced by the participants in the first meeting as a 'first stimulus', the researchers started this meeting by showing video clips from the first meeting, accompanied by a presentation slide with a list of topics, highlighting the disturbances that the participants had identified in their work activity. The presentation slide, acting as mirror data, was visible for the entire meeting, to mediate the conversation.

The first discussion topic on the slide, namely *"Lack and weakness of instructional leadership"*, sparked a conversation, initiated by the principal, problematising the notion of centralised leadership in the school community of autonomous experts. This was a significant step towards conceiving an organisational tension, experienced on an 'abstract' level as general lack of direction and structure for work, as a 'conflict of motives' stemming from an underlying systemic issue, a historically evolved contradiction in the activity system. It was realised by the participants that only improving a single aspect of the work community (such as introducing stronger leadership) was inadequate for resolving the conflict. Consequently, in the first expression of *Envisioning* at the meeting, the idea of distributing leadership tasks to pairs or teams was brought up as a potential solution by a senior teacher (*"...I guess the school life would develop quicker if [---] instructional leadership would be developed by pairs or teams" 668, ST5, Envisioning*).

Towards the end of the meeting, an idea for a new practice was suggested. This took place by drawing from an example of successful collaboration by the pre-service teachers in the school, whose example of improvised collaboration was seen by the principal as a potential model for professional collaboration between subject teachers and class teachers: (*"I'm calling for these small streams of ideas that we have already done [---] they could grow into larger currents, practices for the whole school." 888, PR1, Explicating*). However, this proposal, along with other similar utterances recognising new potential in the activity, still viewed the change as a responsibility of individual teachers and their choices, not something that would require collective efforts and a system-level structural transformation.

In this phase of the CL process, *Criticising* of the current activities continued. However, this phase also captures how the participants further learned to distance themselves from the individually experienced difficulties and jointly began to identify the systemic causes underlying them. Their professional learning thus progressed toward developing a collective ability to *Explicate* the developmental needs of their community.

24.4.3 Meeting 3 – Seeking a Shared Object Through Explicating and Envisioning

In the third meeting, video clips of the previous meeting were shown to stimulate a discussion that took off on the school's mission statement and the need for a shared educational vision that would permeate all school activity from local curriculum work to teaching and pre-teacher supervision. The discussion was thus moving toward, or circling around, the shared 'object of activity'. A first tentative reconceptualisation of the object was reached by binding together the societal, curricular and educational dimensions of schoolwork, and for the first time in the CL bringing "children" into the focus of discussion: (*"ST6 mentioned acknowledging the surrounding society there, so aren't all the transversal competence goals related to how the child will cope, what skills they get for functioning in the future society?" 1313, ST3, Envisioning*).

This was a turning point in the CL: after naming children (or their learning) as the shared object of activity, the organisational problems expressed earlier gained a new developmental focus, functioning as a "motivating force that gives shape and direction to activity" (Engeström, 2008, 89), driving the envisioning of new practices. The teachers expressed a feeling that even such a rudimentary idea of a unifying vision could act as a starting point for a new developmental direction. In other words, instead of relying on individual scripts, the teachers found common ground for an agreement about the school's vision, or "purpose to exist". Consequently, the rest of the third meeting was marked by developing practical solutions for enhanced collaboration.

In this phase, the conflicts related to developing work practices identified earlier, were further analysed by the participants and the researchers in light of the shared object. Two types of transformative agency, namely *Explicating* and *Envisioning*, importantly shifted the discussion from mere naming of the school's various development needs, towards explicating an expanded and shared conceptualisation of the object of the joint activity. We can also witness a redefinition of the 'subject' based on the idea of teamwork as the participants bridged their ideas and work actions to the planning of new forms of joint activity.

24.4.4 Meeting 4 – History Timeline Mediating the Development Efforts

The fourth meeting focused on the history of the school and how the school's development over time might help explain the current structures, practices and problems. A picture of a chronological timeline of the school's key events and transformations, drawn by members of the research team based on teacher interviews conducted between meetings, was used as a stimulus to initiate discussion. While historical analysis had already been carried out even before the introduction of the timeline, making the school's historical development visible sparked a markedly analytical discussion involving several active participants and characterised by consecutively alternating expressions of *Criticising* and *Explicating*.

Many current organisational structures and practices, previously taken for granted, were identified as incompatible with the needs expressed in the CL. Especially the two new teachers participating mentioned experiences of different kinds of practices from schools they had previously worked at, naming them as potential alternatives to existing ways of working: (*"…we had these instructional teams for [---] curricular matters [---] and we also had [---] groups based on school subjects" 1908, NT1, Explicating*). These experiences from different contexts, especially concerning teamwork, proved to be a significant source of ideas for the group in developing new structures of teamwork and leadership, such as an instructional steering group as a new administrative structure for distributing leadership duties.

Towards the end of the fourth meeting, elements of a new organisation structure to support collaboration at all levels of the school community – principals, a steering group, teacher teams, and the general teachers' meeting, and their roles and responsibilities – were envisioned by the participants using the IDEAS, TOOLS board (see Fig. 24.1).

In this phase, establishing a shared understanding of the school's history enabled the participants to connect new ideas and development plans to this historically evolved "bigger picture" and the political and the societal factors underlying the historical developments. The history timeline was thus an important learning device for developing the participants' shared understanding of organisational change as a complex, time consuming and continuous process. In this phase, their professional learning and development were characterised by back-and-forth movement between *Criticising* and *Explicating* expressions and between the old and the new forms of activity.

24.4.5 Meeting 5 – Sharing Responsibility and Envisioning Through Practical Suggestions

As a continuation of the previous meeting, the discussion took off about the tasks, roles and responsibilities of the newly designed teacher teams. As a stimulus, the principal had brought along a list of workgroups currently operating in the school. The list of dozens of workgroups dedicated to various schoolwork and development duties, many of which were non-operational, helped the participants grasp the ineffectiveness of the school's current teamwork structures, mediating a discussion about the changing professional demands and the inability of the current structures to match them. Our analysis of transformative agency in this meeting shows a gradual shift towards more *Explicating* and *Envisioning*, illustrating the group's increased ability to view schoolwork as a systemic activity, oriented towards student learning, and to develop specific solutions through recognising the contradictions inherent in the activity. An important mandate for transformation was provided by the principal, who assumed a facilitative role in this meeting, sharing the responsibility for envisioning new organisational practices with the teachers (*"When we meet on August 10th [---] in the planning meeting, how are we going to launch this?" 2697, PR1, Committing to actions*).

Towards the end of the meeting, we identified numerous *Envisioning* expressions with active participation from the teachers. In a brief passage of enthusiastic, overlapping speech by multiple teachers, a distribution of duties and responsibilities between teacher teams, teachers' general meeting and the principal was jointly articulated. At this point, the participating teachers had not only repeatedly expressed transformative agency and empowerment to change their local context but had also assumed a role as teacher leaders who were able to negotiate organisational structures and roles, recorded on the IDEAS, TOOLS board. This phase was particularly important for the emergence and development of the participants' joint responsibility over organisational change. From the viewpoint of collective professional learning and development, the dominant types of transformative agency, *Explicating* and *Envisioning*, not only contributed to the responsibility-taking but also promoted the efforts to materialise the organisational change as the teachers jointly made decisions and envisioned practical solutions.

24.4.6 Meeting 6 – Constructing a Model to Facilitate Educational Change

Unlike the earlier meetings, for the sixth meeting the researchers did not bring videotaped clips to view and analyse with the participants, but instead used a presentation slide as a stimulus, with topics listed from earlier discussions related to teamwork and shared instructional leadership, visible throughout the meeting. The researchers also encouraged the group to start drawing a model to materialise and summarise ideas articulated during the previous meeting, directing them towards the expansive learning action of *modelling* the new solution.

In innovating new practices and organisational structures during the meeting, the participants relied heavily on external stimuli to mediate the process. These included the presentation slide provided by the researchers as well as the list of work groups provided by the principal in the fifth meeting (through reminiscing), as the newly suggested teamwork practices were mirrored against the long list of work groups and their tasks.

During this meeting, expressions of *Envisioning* were prevalent and were especially identified as the participants engaged in the modelling process. The group drew four preliminary versions of the teamwork and leadership structure on a flipchart before reaching an agreement, ready to draw the fifth and final model. During the process, the preliminary versions of the model acted as 'second stimuli' and mediating artefacts, helping the group progressively to create potential new forms of collaboration, enriching the model. The group also tested the model by critically discussing it and by simulating their work activities within the new model through talk.

The model drawn in the CL supported the teachers' professional learning and development as it materialised the collaborative forms of work by establishing teacher teams with clearly defined roles, tangible tasks and designated responsibilities. Instructional leadership and decision-making over local curriculum work and the school's duties as a university unit were effectively distributed across the organisation.

24.4.7 Follow-Up Meeting – Reflection of the Implementation

The new teamwork and leadership model was implemented in the autumn semester of 2015 as the school adopted the new curriculum. In a follow-up meeting of the Change Laboratory process, held in February 2016, seven teachers and the principal reflected on their experiences of the implementation of the new model into their daily work. The model had provided a good basis and structure for developing new collaborative practices in the work community, but it had only been taken up by some teacher teams while other teachers chose to hold on to their autonomy and individual work practices. New ideas for cultivating the model to fit the reality of the daily work better were also initiated and discussed. In the conversation, it became evident that the central contradiction defined in the CL process – teacher autonomy vs. the elevated need for collaboration – aggravated by the new national core curriculum, was still causing disturbances and conflicts in the work community. This shifting between the old and new models of work is typical for an expansive learning process in its implementation phase (Virkkunen & Newnham, 2013; Engeström, 2015). Furthermore, the new curriculum's incompatibility with the school's status as a university research and pre-teacher supervision unit was discussed, marking a shift towards conceptualising the issue as a contradiction between the school's activity system and its neighbouring activity systems, namely those of the university and the governmental education authorities. In conclusion, the transformation of work activity was still an ongoing process, and no final solutions had yet been reached.

24.5 Discussion and Conclusion

Our chapter contributes to activity-theoretical development efforts which have yielded productive and sustainable solutions to complex challenges in schools, contributing to teachers' professional learning and development (e.g., Zhang et al., 2011; Brevik et al., 2019; Lund & Vestøl, 2020; Junor Clarke & Fournillier, 2012). Our findings also add to the studies which have widened the understanding of the significance of tensions, contradictions (e.g., Engeström et al., 2002; Virkkunen & Tenhunen, 2010; Rainio & Hofmann, 2021) and transformative agency (e.g., Sannino, 2008; Brevik et al., 2019; Kajamaa & Kumpulainen, 2019; Lund & Vestøl, 2020; Grant, 2020; Morselli & Sannino, 2021) in the collective creation of educational innovations and practice changes within schools. Furthermore, our study contributes to the much-needed knowledge and reflection on how the CL method is currently being applied as a tool for professional learning and development in educational settings and what kinds of consequences and results it can yield (Virkkunen & Newnham, 2013).

In this chapter, we aimed to enrich the current understanding of teachers' collective professional learning and development by introducing a participatory work development method called the Change Laboratory (CL). Our findings illustrate how the CL offers a method for facilitating collective professional learning and development through the articulation and resolution of organisational contradictions. This process involves engaging the participants in the articulation of uncontextualised ('abstract') disturbances in their current work practices and gradually moving towards analysing their work as a systemic entity with historically evolved contradictions behind the disturbances – and, eventually, taking actions to create novel practices to resolve the contradictions. This process requires both expansive learning and transformative agency among the participants.

As the CL process unfolded, an especially pressing contradiction identified by the teachers was related to the inadequate tools for reconceptualising the changing object of the joint activity, specifically, the pupils and their diverse and evolving learning needs, especially in the light of the new core curriculum. The existing instructional and leadership practices had become inadequate in responding to the changed needs. This contradiction was aggravated by the upcoming curriculum reform that laid bare the individually motivated and partly contradictory developmental motives in the school organisation (stemming from the historically formed autonomy of the teachers), which gradually became conceptualised as contradictions related to the subjects, the division of labour and the rules of the teacher community. The transformation, achieved through the establishment of new teamwork and leadership practices, can be interpreted not only as a reconceptualisation of the object but also as a redefinition of the subject, expanding the conception of autonomous teachers with a notion of collaborating teams. The new collaborative process involving the school's local curriculum work, teaching and instructional leadership, oriented towards a new, expanded object of activity, provided an instrument for shifting from individually driven work practices toward more collaborative and shared forms of work and leadership.

Our analysis of the types of expressions of transformative agency in the participants' discourse shows the explanatory power of this framework for facilitating collective professional learning and development in the CL process, leading to practice changes at the level of an entire organisation. Our detailed analysis provides new insights into the mechanisms of how such a transformative development of work practices gradually unfolds in a process of collective professional learning. Although the conversation in the meetings in our case was at times tense and even heated, the critical voices, manifesting as *Resisting* and *Criticising* expressions up until the end of the CL, did not hinder the creation of new solutions and resolution of the contradictions. In other words, the complexity and tensions involved were not harmful but also triggered and maintained collective professional learning (also Kajamaa & Kumpulainen, 2019). On this basis, we wish to highlight the potential of the CL method for offering a safe space for the participants to turn the contradictions into drivers for collective learning and development (Engeström, 2011; Virkkunen & Newnham, 2013; see also Engeström & Sannino, 2011).

Our study also generates new knowledge about the ways the material mediators support participation and create opportunities for shared decision making and responsibility taking, which in our case were also pivotal in enhancing collective professional learning and development. Our case analysis not only describes discursive change efforts, but also elucidates how these are put to use by the creation and enactment of mediators by the actors involved. Furthermore, our findings echo with previous activity-theoretical studies showing that the materialisation of the results of the development projects into shared tools, such as leadership models designed by CL participants, endorses the diffusion and sustainability of the outcomes of organisational change efforts (Engeström et al., 2007; Kajamaa, 2011).

Furthermore, in a follow-up meeting held 9 months after the CL process, the group had a chance to reflect on both the CL and the implementation process, revealing that not all the problems in the work community had been resolved. Despite some significant improvements at the school, the newly implemented collaborative practices had also aggravated yet further contradictions, especially regarding some teachers' professional identities and how they related to the new practices, as well as the school's diverse tasks as a university research and pre-service teacher training unit. To tackle these newly arisen challenges and further facilitate the expansive learning process, another follow-up meeting might prove useful.

To conclude, the Change Laboratory method can provide a powerful tool for collective professional learning and development in educational settings, by facilitating the participants' transformative agency and expansive learning. The method requires volition from the participating organisational actors to actively engage in the critical analysis and development of their daily work activity (Virkkunen & Newnham, 2013). Along with the previous literature, we wish to highlight the role of the facilitator as pivotal, not in directing or predefining, but acting as a chair of the discussion, asking reflective questions and offering tools to aid the participants to reflect and analyse their work activity, and to transform their circumstances in a way meaningful to them, their students/clients and the society. For the facilitator, conducting such a process requires research knowledge and deep understanding of the principles of the method. In addition, the method calls for creativity, flexibility (e.g., accepting deviations from original plans to adjust them to the developmental stage of the given context) and sensitivity toward the participants' multiple voices, opinions, ideas and emotions. Thus, each CL process is unique, and as it happens as intertwined to the actual work practices, it potentially produces sustainable new activity models and other consequences. In our experience, it is crucial that the facilitator maintains trust and resilience throughout the process. Echoing Engeström et al. (2007), we emphasise the importance of acknowledging that the collective identification and analysis of contradictions, to generate collective sense, meaning and expansive learning, is a time consuming and demanding collaborative journey which may take months and even years.

References

Admiraal, W., Schenke, W., De Jong, L., Emmelot, Y., & Sligte, H. (2021). Schools as professional learning communities: What can schools do to support professional development of their teachers? *Professional Development in Education, 47*(4), 684–698.

Benson, J. K. (1977). Organizations: A dialectical view. *Administrative Science Quarterly, 22*(1), 1–21.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15.

Brevik, L. M., Gudmundsdottir, G. B., Lund, A., & Strømme, T. A. (2019). Transformative agency in teacher education: Fostering professional digital competence. *Teaching and Teacher Education, 86*, 102875.

Brodie, K. (2021). Teacher agency in professional learning communities. *Professional Development in Education, 47*(4), 560–573.

Cordingley, P. (2015). The contribution of research to teachers' professional learning and development. *Oxford Review of Education, 41*(2), 234–252.

Davydov, V. V. (1990). *Types of generalization in instruction: Logical and psychological problems in the structuring of school curricula*. National Council of Teachers of Mathematics.

Engeström, Y. (2007). From stabilization knowledge to possibility knowledge in organizational learning. *Management Learning, 38*(3), 271–275.

Engeström, Y. (2008). *From teams to knots: Activity-theoretical studies of collaboration and learning at work*. Cambridge University Press.

Engeström, Y. (2009). From learning environments and implementation to activity systems and expansive learning. *Actio: An international journal of human activity theory, 2*, 17–33.

Engeström, Y. (2011). From design experiments to formative interventions. *Theory & Psychology, 21*(5), 598–628.

Engeström, Y. (2015). *Learning by expanding: An activity-theoretical approach to developmental research*. Cambridge University Press.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review, 5*(1), 1–24.

Engeström, Y., & Sannino, A. (2011). Discursive manifestations of contradictions in organizational change efforts: A methodological framework. *Journal of Organizational Change Management, 24*(3), 368–387.

Engeström, Y., Virkkunen, J., Helle, M., Pihlaja, J., & Poikela, R. (1996). The Change Laboratory as a tool for transforming work. *Lifelong Learning in Europe, 1*(2), 10–17.

Engeström, Y., Engeström, R., & Suntio, A. (2002). Can a school community learn to master its own future? An activity-theoretical study of expansive learning among middle school teachers. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century*. Blackwell.

Engeström, Y., Kerosuo, H., & Kajamaa, A. (2007). Beyond discontinuity: Expansive organizational learning remembered. *Management Learning, 38*(3), 319–336.

Engeström, Y., Kajamaa, A., Lahtinen, P., & Sannino, A. (2015). Toward a grammar of collaboration. *Mind, Culture, and Activity, 22*(2), 92–111.

Finnish National Agency for Education. (2014). *National core curriculum for basic education*.

Goller, M., & Paloniemi, S. (2017). Agency at work and its relationship to learning and professional development: An introduction. In M. Göller & S. Paloniemi (Eds.), *Agency at work: An agentic perspective on professional learning and development* (pp. 1–14). Springer.

Grant, C. (2020). Double stimulation and transformative agency for leadership development of school learners in Southern Africa. *Pedagogy, Culture and Society, 30*(1), 53–70.

Haapasaari, A., Engeström, Y., & Kerosuo, H. (2016). The emergence of learners' transformative agency in a change laboratory intervention. *Journal of Education and Work, 29*(2), 232–262.

Hyrkkö, S., & Kajamaa, A. (2021). Distributed creativity and expansive learning in a teacher training school's Change Laboratory. In S. Lemmetty, V. Glaveanu, K. Collin, & P. Forsman (Eds.), *Creativity and learning: Context, processes and impact*. Springer.

Il'enkov, E. V. (1977). *Dialectical logic: Essays in its history and theory*. Progress.

Junor Clarke, P. A., & Fournillier, J. B. (2012). Action research, pedagogy, and activity theory: Tools facilitating two instructors' interpretations of the professional development of four preservice teachers. *Teaching and Teacher Education, 28*(5), 649–660.

Kajamaa, A. (2011). *Unraveling the helix of change: An activity-theoretical study of health care change efforts and their consequences*. University of Helsinki.

Kajamaa, A., & Kumpulainen, K. (2019). Agency in the Making: Analyzing students' transformative agency in a school-based makerspace. *Mind, Culture, and Activity, 26*(3), 266–281.

Kerosuo, H., Kajamaa, A., & Engeström, Y. (2010). Promoting innovation and learning through Change Laboratory: An example from Finnish health care. *Central European Journal of Public Policy, 4*(1), 110–130.

Kumpulainen, K., Kajamaa, A., & Rajala, A. (2018). Understanding educational change: Agency-structure dynamics in a novel design and making environment. *Digital Education Review, 33*, 26–38.

Lattuca, L. R. (2002). Learning interdisciplinarity. *The Journal of Higher Education, 73*, 711–739.

Leont'ev, A. N. (1978). *Activity, consciousness and personality*. Prentice-Hall.

Lund, A., & Vestøl, J. M. (2020). An analytical unit of transformative agency: Dynamics and dialectics. *Learning, Culture and Social Interaction, 25*, 100390.

Miettinen, R. (2013). *Innovation, human capabilities, and democracy: Towards an enabling welfare state*. Oxford University Press.

Morselli, D., & Sannino, A. (2021). Testing the model of double stimulation in a Change Laboratory. *Teaching and Teacher Education, 97*, 103224.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407.

Rainio, A., & Hofmann, R. (2021). Teacher professional dialogues during a school intervention: From stabilization to possibility discourse through reflexive noticing. *Journal of the Learning Sciences, 30*(4–5), 707–746.

Sannino, A. (2008). From talk to action: Experiencing interlocution in developmental interventions. *Mind, Culture, and Activity, 15*, 234–257.

Sannino, A. (2015). The principle of double stimulation: A path to volitional action. *Learning, Culture and Social Interaction, 6*, 1–15.

Sannino, A., Engeström, Y., & Lemos, M. (2016). Formative interventions for expansive learning and transformative agency. *Journal of the Learning Sciences, 25*(4), 599–633.

Van Meeuwen, P., Huijboom, F., Rusman, E., Vermeulen, M., & Imants, J. (2020). Towards a comprehensive and dynamic conceptual framework to research and enact professional learning communities in the context of secondary education. *European Journal of Teacher Education, 43*(3), 405–427.

Virkkunen, J. (2006). Dilemmas in building shared transformative agency. *Activités, 3*(1), 43–66.

Virkkunen, J., & Newnham, D. S. (2013). *The Change Laboratory: A tool for collaborative development of work and education*. Sense Publishers.

Virkkunen, J., & Tenhunen, E. (2010). Finding a concept that integrates specialists' know-how – The case of a special school for handicapped and neurologically ill children. *Actio: An International Journal of Human Activity Theory, 3*, 1–24.

Vygotsky, L. S. (1978). *Mind in society: The psychology of higher mental functions*. Harvard University Press.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education, 24*, 173–209.

Winch, C., Oancea, A., & Orchard, J. (2015). The contribution of educational research to teachers' professional learning: Philosophical understandings. *Oxford Review of Education, 41*(2), 202–216.

Zhang, M., Lundeberg, M., & Eberhardt, J. (2011). Strategic facilitation of problem-based discussion for teacher professional development. *The Journal of the Learning Sciences, 20*(3), 342–394.

Anu Kajamaa (PhD) is a Professor of Education (Continuous Learning), in the Faculty of Education, University of Oulu, Finland. She draws from Cultural-Historical Activity Theory and sociocultural approaches, and has conducted Change Laboratory projects in schools, teacher education, prison services, health care and entrepreneurship contexts. She has had nearly 100 publications accepted and has received several international awards for her research and teaching. She is a certified Change Laboratory facilitator, trained by professors at the University of Helsinki. She is also an Honorary Professor at the Centre for Research in Professional Learning (CRPL), at the University of Exeter, UK.

Sakari Hyrkkö is a Doctoral Researcher at the Faculty of Educational Sciences, University of Helsinki, Finland. His research interests lie in collective and interpersonal structures of organisational activity, creativity, learning, agency and leadership, drawing from Cultural-Historical Activity Theory and its applications. He is also a member of the faculty's Learning, Culture and Interventions (LECI) expert research group.

Chapter 25
Professional Learning Analytics: Understanding Complex Learning Processes Through Measurement, Collection, Analysis, and Reporting of MOOC Data

Allison Littlejohn, Eileen Kennedy, and Diana Laurillard

Abstract Global, organisational and technological changes are transforming the world of work. To respond to these changes, professionals need to be able to adapt and upskill flexibly, elevating the need for lifelong professional learning. Technological systems are being used to provide professional learning at scale, supported by learning analytics to scaffold the learner. Learning Analytics has been defined as 'the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs' (Siemens G, Long P, EDUCAUSE Rev 46(5), 2011, p. 30). It comprises a range of methods, from 'predictive analytics' that identify 'at risk' learners, to 'multimodal' methods that bring together diverse data to help inform learners. Alongside potential benefits of Learning Analytics are a number of problems of datafication, such as increased workplace surveillance, automation of decisions and making use of available data, rather than identifying specific data needed to answer research questions. This chapter considers these and other methodological challenges, through presentation of case examples of forms of Learning Analytics from professional development Massive Open Online Courses (MOOCs). Each example considers the complex processes within professional learning and how these might be analysed in future using new methods for Professional Learning Analytics.

Keywords Professional learning · Learning analytics · Online learning · Digital education · Digital data · Value creation

---

A. Littlejohn (✉) · E. Kennedy · D. Laurillard
UCL Knowledge Lab, Institute of Education, University College London, London, UK
e-mail: allison.littlejohn@ucl.ac.uk

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_25

25.1 What is Learning Analytics?

Global, organisational and technological changes are transforming the world of work. More and more professionals are working together via networked, digital environments, leading to an escalation in digitalisation and datafication, as digital data becomes a central component of almost every form of work (EU, 2021). Datafication is the trend towards turning many aspects of work into data which is transferred into information which offers a new form of value. As professionals work and collaborate using digital tools and environments, they leave various forms of digital traces and data (Littlejohn & Margaryan, 2013, p. 1). These data can be exploited using data analytics to make learning processes visible in ways that enable scientific research and/or support professionals with their learning. Professional Learning Analytics systems are being developed to monitor professionals as they learn and to automate and scale learning support. These systems are being designed for a diverse range of learning contexts, from formal educational programmes or vocational training, to situations where professionals learn through everyday work.

In this chapter, we define Learning Analytics (LA) as a form of data analytics aimed at 'the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimising learning and the environments in which it occurs' (Siemens & Long, 2011, p. 30). This definition was conceived to describe Learning Analytics in formal, educational setting, such as universities or schools and has to be extended as Professional Learning Analytics (PLA) applied in workplace contexts.

Analytics techniques are already being used in a variety of ways to understand and support work processes. For example, PLA is being used by workers to improve formal and structured learning: for example by trainers to help scaffold learning processes or by organisations to examine competency development in the workplace. PLA methods can be used alongside classic (analogue) educational research methods, for example digital data can be compared with how each learner self-reports how they have learned in interviews.

In this chapter we provide an overview of Professional Learning Analytic methods. We argue that some current methods do not provide accurate insights into the complex learning processes that professionals engage in for three reasons. First, we argue that learning analytics tends to be dominated by data science and computational science approaches that are based on formal education and do not take into consideration the complexity of learning and the social science research that has investigated it. We question whether and how a broader range of disciplines might come together in ways that offer new insights. Second, analytics techniques tend to measure and quantify behavioural and/or affective data available for analysis, rather than using learning science to identify the broader range of qualitative data needed to understand and make decisions about learning. While these data could be useful in understanding and interpreting learning, they should not direct the analysis, but should instead be used only if they help answer the research questions. This leads to a third major problem, that data science techniques analyse and interpret these quantitative data in a mechanistic and reductive way, without due consideration of the influence of the complex interactions that govern each learner's work and learning environment. We argue the need for a broader, more interdisciplinary approach to researching professional learning in digital environments that focuses on the value of the learning process and outcome to the professional.

Our first step is to examine the origins of Learning Analytics as a form of automated support to scale up learning opportunities, which we discuss in the next section.

25.2 Learning Analytics as a Form of Automated Support for Learning

The idea of automated, digitised support for learning has its origins in formal education and is not new. Almost 40 years ago Benjamin Bloom published his paper on 'The 2 sigma problem' (Bloom, 1984). Based on empirical data gathered by Anania (1982) and Burke (1984), Bloom compared the outcomes of learners taught using conventional classroom teaching with students taught under conditions of one-to-one tutoring. Bloom found that under the tutoring conditions the average learner's assessment outcome was about two standard deviations above the average of the control group in the classroom (Bloom, 1984). While it can be argued that outcomes cannot easily be measured through conventional tests, these findings point to the importance of tutoring, which scaffolds each learner's progress in ways that are not possible using group instruction. Bloom questioned whether educators could formulate conditions for teaching and learning that enable the majority of learners under group instruction to attain the levels of achievement that can be reached under tutoring conditions (Bloom, 1984). A proposed solution was to replicate one-to-one tuition in an affordable way by using technology-based systems to provide personalised support to learners (Anderson et al., 1995).

Since then, a range of intelligent tutoring systems (ITS) have been developed (see for example Ahuja & Sille, 2013; Mousavinasab et al., 2021) using data analytics to provide automated feedback to learners, without the intervention of a human teacher. ITS comprise four interacting components: (a) the student model which represents the learner's cognitive state, (b) the pedagogical model which includes appropriate instructional interventions, (c) the knowledge model which contains the content, and (d) the user interface model which supports dialogue. An assumption underpinning the design of ITS is that cognitive skill acquisition involves the formulation of thousands of rules relating task goals and task states to actions and consequences. These rules are developed using production-rule formulas that represent goal-oriented knowledge (Anderson, 1983). As more data is gathered and processed the ITS improves its ability to support the learner using Artificial Intelligence. For example, in disciplines where problems have right and wrong answers, a digital tutor or 'bot' can provide the learner with immediate feedback, consisting of short, directed messages, conditional on the learner's answer (Anderson et al., 1995). The more data the bot is exposed to, the better it can perform. ITS tend to emphasise the expert knowledge as the central component of the instructional process. The learner model identifies common errors in the learner's knowledge and selects the appropriate feedback and next task on the basis of predefined rules. The system tends to provide the learner with hints or supplementary content to help them acquire knowledge, rather than asking questions of learners and giving feedback in ways that might help them learn how to learn, as an experienced teacher would (Kluger & DeNisi, 1998). This means that the ITS performs poorly as a teacher.

To address this problem, analytics techniques have been developed to support learners by replicating or enhancing what the teacher does by processing large amounts of student-generated performance data. Critics of these systems are concerned that the data that is gathered is not the form needed which leads to over-simplification of the interpretation of these data (Selwyn, 2019). In conventional classroom settings, teachers use multisensory cues – such as facial expressions, body language and vocal responses – to assess how well students are learning, and adapt their teaching strategies accordingly (Titsworth, 2001). This ability to scaffold learning by observing and adapting to each learner's needs is a fundamental part of good teaching practice in small group teaching, where experienced teachers instinctively process complex and sometimes tacit information. However, in even moderate class sizes and with limited time, teachers are not always able to do this adequately. Automated tutoring in theory could provide more personalised feedback than is possible in a classroom. Learning Analytics systems can measure and analyse a range of data, such as keystrokes, text, time on task, geolocation, discourse, and even, with the right additional technology, facial expression, temperature and skin conductivity. These different forms of data provide a basis on which analytics systems can interpret learning situations and (potentially) provide support.

An advantage of analytics systems is that they can process huge quantities of data from large numbers of learners. The greater the number of people who provide data, the greater the potential to analyse data in meaningful ways that provide research insights and inform the design scaffolds and supports for learners. Data at scale provide insights, such as how many students are likely to drop out of a course or highlight sections of a course learners find difficult. Teachers working in smaller classes can predict the scale of dropout from a course as they get to know learners. Teachers usually take action to support and scaffold individual learners to help them progress. However, these predictions can be more challenging as the number of learners scales to hundreds or thousands and as the time the learner is in direct contact with a teacher is reduced, as in the case of large classes or Massive Open Online Courses (MOOCs) – online courses that are often (though not always) open access and freely available.

In this sense, Learning Analytics is most helpful where large numbers of people are learning, because these situations make it more difficult for a teacher to observe the performance of a single learner and also large amounts of data can be gathered and analysed. However, not all professional learning situations have these large numbers and the data gathered may not be those data that are needed but those that are simply available automatically.

While there are clear benefits afforded by Learning Analytics, there are also a number of problems in drawing firm conclusions from the interpretation of these data. One problem is that analytics systems are not as skilled or nuanced as teachers in their interpretation of learners' responses. System coding is based on a number of assumptions about how a teacher would respond to learners, how they modify their teaching depending on learner behaviour, body language, facial expressions and so on. These codes can reduce complex social settings to a small number of factors. For example, learners' self-regulation might be measured through time management or planning behaviours because of a correlation between planning behaviours and course outcomes. However, time management and planning are influenced by a wide range of socio-environmental factors. Teachers intuitively understand that learning is bound to the experiences, beliefs and motivations of individual learners as well as their broader socio-cultural context, relationships and environment and it is important to consider all of these factors. Learning is inseparable from the personal histories and experiences of learners and it is difficult to know how someone is progressing in their learning unless all of these factors are taken into account. However, teachers in post-compulsory educational settings, such as workplaces, colleges and universities, do not have an appreciation of each learner's whole-life experience. Learning Analytics systems offer a solution, but currently they only focus on the narrow range of variables they are programmed to measure and analyse. These narrow range of data are helpful for large classes where teachers or trainers do not have the time or ability to analyse the progress of individual learners, and automated, individualised feedback gives the learner advice to act upon.

A second problem relates to the volume of data needed to make informed conclusions about each learner's progress. Qualitative, multimodal accounts of learning provide the sorts of data needed to understand learning, but these data are difficult to measure, analyse and draw conclusions from (Littlejohn & Hood, 2019). Gathering these narrative accounts requires the collection and analysis of large volumes of qualitative data needed to understand whether a learner is learning. To get around this problem and simplify analyses, Learning Analytics systems tend to measure and use a relatively narrow range of (usually quantitative) data rather than those data that could provide a more accurate analysis and interpretation. Data scientists understand this problem and are experimenting with new analytics methods that gather, combine and analyse multimodal data, such as environmental data, location data, facial expressions and body gestures that, together, provide a more comprehensive understanding of each learner's situation. This is known as Multimodal Learning Analytics (MMLA) (Di Mitri et al., 2018). However, combining and analysing large amounts of different types of data requires a great deal of computational power as well as an interdisciplinary approach to create the algorithms for data analysis.

A further issue relates to the diverse epistemologies of the different disciplines that need to be integrated to gather and interpret data needed to explore and measure value for professionals –including learning science, social science and business studies, computational science and data science,. This makes it difficult to bring theories, methods and approaches together (Littlejohn & Pammer-Schindler, 2021). The next section reviews Learning Analytics methods and considers their relevance for understanding complex professional learning and development.

25.3 Learning Analytics Methods

Researchers and developers of Learning Analytics have adapted a number of methodologies from disciplines including data science and computational science, some of which are widely used in business. Methods include (a) *visualisation* using dashboards that illustrate each learner's progress in relation to a pre-prescribed curriculum pathway or in relation to the learner's progress relative to other people (Clow, 2013), (b) *predictive analytics* using modelling techniques that forecast the likelihood of a learner falling behind or dropping out of a course (Shum & Ferguson, 2012), and (c) *recommender systems* and *adaptive systems* that aim to personalise learning by providing options in terms of learning resources or pathways, depending on the learner's profile, behaviour or emotions (Nistor & Hernández-Garcíac, 2018).

Visualisation systems present data analyses in ways that are intended to inform learners (or teachers) about their progress in a course. For example, the outcomes of a Social Network Analysis can be visualised to indicate the learner's position within a learning network to illustrate their 'connectedness'. This technique is based on the assumption that the connectedness is important for learning and that the learner's position within this network may be strengthened through interactions with peers and tutors using social media tools such as blogging and microblogging tools or by linking with others through online discussion threads. Other visualisation methods illustrate dispositional data to allow learners to make informed choices about how they learn. For example, Shum and Crick (2012) link data on learners' ability to self-regulate their learning with assessment data to feed back to learners how they might amend their learning in ways that allow them to achieve their goals. The authors use a survey instrument that asks each learner questions about how they learn and assesses their 'Learning Power' – each learner's ability to self-regulate their own learning. These data are compared with the data of other learners held in a global Learning Warehouse (Shum & Crick, 2016). The system generates a 'spider diagram' that provides the learner with a visualisation of their ability to, for example, set goals, seek help, self-evaluate while learning. There is evidence that this visualisation methods helps build learners' self-awareness to supports learner's development of self-regulation skills (Shum & Crick, 2016). The system also provides cohort summary statistics which can be used to inform teacher's pedagogical interventions which, combined with predictions about learners' progress, can be helpful for teachers in pinpointing ways to scaffold learners in large cohorts that would otherwise be difficult to analyse and support.

Predictive analytics were developed for formal education settings (for example university online programmes) to assess each learner's likelihood of achieving 'success'. Success measures range from completing an assessment to reaching a specific stage in the programme. Predictions may be made by comparing the learner's data to those of other learners. For example, Jiang et al. (2014) found early alert factors related to a learner's behaviour in the first week of a MOOC that can signal whether or not a participant would complete the course. Predicting whether a participant will drop out is complex. The predictive capability of a system depends on having tracked prior data and identified the indicators that produce a trajectory that ends in dropout. Different cohorts of learners can behave differently, so forecasting predictions based on previous data (even for the same course) assumes cohorts have similar behaviours (Littlejohn & Hood, 2018). Another assumption underpinning predictive analytics is that learners want to complete the course and attain a qualification and will therefore respond to advice that completing the assessment will make it more likely that they complete a course. However, there is ample evidence that not all learners in MOOCs are motivated to complete assessments and that, even if they do not complete the course, they may still learn (Littlejohn et al., 2016). This is likely to be the case for professionals looking for insights to solve professional problems rather than pursuing a qualification or certificate.

Other factors that have been used as a proxy for success in MOOCs include time management (Balakrishnan & Cooetzee, 2013), lighter workload, higher autonomy and more flexible assessments (Skrypnyk et al., 2015). This methodology relies on proxy indicators that have a number of important underlying assumptions about learner behaviour and their motivations to learn. There are a number of problems with time management as a proxy for self-regulation, since time management is one of many inter-related factors that contribute to self-regulation. Correlating self-regulation with time management reduces an intricate, qualitative process to a measurement of one part of the process and does not take account the whole system. The analysis uses data that are not representative of how each learner is learning, reducing a complex process to one or two behavioural indicators. These proxies can be helpful in specific learning settings, for example in large-scale courses or training where learners aim to complete the course to gain a qualification, but the course numbers are so large that the teacher finds it difficult to give individualised feedback. Here, automated feedback is helpful. However, open and flexible learning introduces a range of learner motivations and goals with highly variable patterns of engagement (Conole, 2013). There are many instances where professionals do not intend to complete a course or gain a qualification. For example, professionals 'drop in' to a course to learn a specific concept or skill to help them with their job. Milligan and Littlejohn (2017) analysed the motivations of professionals who registered for a MOOC on 'Introduction to Data Science', offered by the University of Washington on the Coursera platform. The course was designed for people with a moderate level of programming experience. Over 8 weeks, 50,000 learners, from 197 countries participated in the course. Survey data from 788 learners provided evidence of a wide range of motivations, goals, learning behaviours and perspectives. Some participants aimed to learn a specific programming skill but did not intend to engage in the whole course. They could achieve their personal goals without completing the course. Learning Analytics has to take account of each learner's goals, motivation and agency in order to provide an accurate analysis. It is difficult for teachers or trainers to know the individual goals and motivations of learners and the value that engaging in a course brings to them, especially where there are large numbers of learners, so this is an area where Learning Analytics can be of benefit.

Recommender systems offer learners guidance, including advice about learning resources, people they can connect with or learning pathways they might choose. These recommendations are based on different kinds of data (demographic data, behavioural data and sometimes emotional data) gathered from the learner and analysed against previous data from a large number of learners, usually gathered from the same or previous cohorts within a course. Skrypnyk et al. (2015) analysed learners' demographics and groupings to provide learners with personalised recommendations about how they might improve their learning. *Adaptive analytics systems* are based on similar methodologies for data tracking, and analysis through these systems automatically changes a course design or pathway in response to the learner's data.

Emerging analytics methods are gathering and expanding the types of data that are analysed to support learning, expanding beyond demographic and behavioural data to include dispositional and affective data that indicate how learners feel about their learning as well as data on learner motivations (Gasevic et al., 2017). These data allow for more influential recommendations and adaptations of learning resources. Data are gathered in multimodal formats (Nistor & Hernández-Garcíac, 2018), through digital traces that indicate the choices learners make: eye tracking that analyses where the learner's attention is placed, facial recognition and skin conductivity systems that analyse facial expressions as a proxy for affective states, temperature and skin conductivity detectors measuring moisture in skin as a proxy for level of anxiety, and location data (Martinez-Maldonado et al., 2018). However, sentiment analysis is complex and depends on a wide range of factors, including culture and setting.

Although computational systems are able to analyse large volumes of data, they are not able to respond in the same ways as teachers or trainers. Current Learning Analytics methods tend to use the data that are available and easy to measure and analyse. While these analyses offer insights into learner choices and behaviours, new methods are being developed in ways that offer learners and teachers more support than oversimplified analyses of happenstance data. These methods are experimenting with the types of data that can be captured to provide more meaningful, individualised support.

Some of these new methods combine complex, multimodal data to allow more in-depth analysis. For example, Gillani and Eynon (2014) examined the strategies of hundreds of learners as they engaged in online discussions, using semiotic analysis and network analysis techniques, they identified 'significant interaction networks' embedded within discussion forums. These networks serve as a proxy for forms of active learning within digital environments. Other new methods of MMLA combine brain imaging (MRI) data with behavioural (keystroke) data examining writing and pausing and linguistic analysis to understand how use of language influences learning when learners are engaging with text to inform future learning processes. Future analytics methods will combine a wider range of multimodal data to inform more meaningful conclusions. These methods could examine not only the outcomes of learning but the value created through participation in learning. The next section outlines why this is important for professional learning.

25.4 Learning Analytics for Professional Learning

Learning Analytics methods largely have been developed for use in formal educational settings, such as courses or programmes, where the main motivation of the learner is to attain a qualification. The motivations and settings for professionals as they learn for work often are different and, therefore, require novel analytics methods.

Professional learning includes the range of activities professionals engage in to stimulate their professional knowledge, to improve work performance and to ensure their practice is up-to-date (Littlejohn & Margaryan, 2013, p. 2). These activities range from formal courses including educational programmes, training and workshops, to non-formal, on-the-job learning such as learning through work, observing or talking with a more experienced colleague or being mentored. There is a growing body of evidence that professional learning is most effective when integrated with work (See for example Tynjala, 2008; Fuller & Unwin, 2003). Professional learning is often deeply embedded within and difficult to distinguish from work (Eraut & Hirsh, 2010; Engestrom, 2016). Eraut (2004) emphasises the importance of on-the-job learning, where learning goals are aligned with work goals, making work and learning difficult to distinguish. For example, a manager in the finance sector might learn new inter-cultural competencies through working with colleagues from around the world (Littlejohn et al., 2016). Her learning goals are not structured and her environment is not a bounded ‘classroom’ environment, therefore it is challenging to measure data that indicates how she is learning.

In summary, as professional learning is scaled, there is a need to find automated ways to scaffold and support learners. Learning Analytics provides useful analyses, support and feedback to large numbers of learners in ways that would be difficult for teachers or trainers who are working with large cohorts. The domain of Learning Analytics has its origins in formal education, such as higher education or school learning. However, since professional learning is a very different context compared with formal education, we need to consider what counts as being of value to professionals and what do they hope to gain from their professional learning. The motivations for professionals to learn are often instant, practical and aligned with work tasks. Therefore, it is unlikely that the approach developed for formal education will be easily transferable to the professional learning context. For example, some of the indicators of learning used in analytics systems are adopted from formal education (e.g. course completion, assessment grades). These indicators are less helpful for professional learning than for formal education, because professionals have different motivations for learning – for example they may want to learn a new concept rather than a complete a whole course. There is, therefore, a need to rethink indicators of professional learning and what data is needed. The following section presents case studies that consider what is of value to professionals and how to generate the data points needed to demonstrate professional learning.

25.5 Case Studies from Professional Development MOOCs

The previous section highlighted that the types of data measured and used in learning analytics are only partially relevant to understanding the impact of learning for professionals. In this section we propose other measures that indicate learning. We begin by considering professional learning in Massive Open Online Courses (MOOCs).

MOOCs in their current form became established around 2012 (Laura Pappano, 2012). Universities began to open up their courses to online recruitment for free, and the novelty of these experiments led to tens of thousands of enrolments. Although there were hopes that such free online courses could be a viable vehicle for widening participation in undergraduate study, it was quickly discovered that MOOCs were dominated by professionals who already had degrees or postgraduate degrees (Hollands & Tirthali, 2014).

MOOCs are asynchronous online courses that enable learners to navigate through course content at their own pace. They are free to enrol, but often learners are required to pay for certificates or prolonged access (upgrade). Early MOOCs were little more than a collection of videos and quizzes, but as the platforms developed, other learning features were added such as discussion forums and peer review which offered more potential for professional peer-to-peer learning. Many MOOCs remain instruction-led, but research has shown that social and collaborative learning designs are possible on the large scale and can be effective for professionals (Laurillard & Kennedy, 2020).

As open, online courses with large numbers of participants, MOOCs provide an ideal opportunity to collect digital traces of learner engagement on the large scale. Since MOOCs are free to enrol on, there is little chance for universities that run them to recoup any return on investment. Where there is a cost for a certificate of attendance, this is also generally kept low, and in any case these have little interest for professionals. Being essentially large-scale and free, with all the attendant benefits this brings for openness and access to all, it is not financially feasible to provide the kind of tutor-based feedback as would be typical for an undergraduate course. This is why MOOCs are unsuitable for undergraduates, unless they are blended into a tutor-supported course. The model is very suitable for continuing professional development, however, where participants are knowledgeable enough to help each other, and do not need one-to-one tuition or tutor assessment, hence the value of using automated assessment and peer review in MOOCs.

As online platforms, MOOCs collect data as learners enrol and engage with the content. Although enrolments rarely count as massive, with active learners in the low tens of thousands at best, they still generate a lot of data automatically collected by the platform. These data have the potential to help educators understand what is happening in their courses, even if they cannot engage individually with more than a handful of participants.

25.5.1 What Data Is Available in MOOCs?

Table 25.1 shows the kinds of data automatically collected by MOOC platforms, and their use for understanding learning.

Summaries of these data are often presented to educators via a dashboard of some kind, or the datasets themselves can be downloaded for later analysis. The kinds of data collected are decided by the platforms, led by what data are most easily available.

Measures of engagement in MOOCs that show that participants have enrolled and started learning are clearly valuable. It is also useful to know where learners are from (e.g. by country) and how far they have progressed, and whether and to what extent they have engaged with activities such as discussion and peer review. Video views, and the points during the video that participants stopped watching, can indicate engagement and disengagement with learning content, as can progression or cessation of progression, through the content steps (the FutureLearn platform refers to each unit of learning, such as a video or an article, as a 'step'). However, data alone will not explain why participants stopped watching a video or stopped moving through the course. This is important because sometimes these measures are used to indicate findings that they cannot support. For example, it is tempting to use the

Table 25.1 Typical datasets available through MOOC platforms and their uses

Dataset	Source of activity	Useful insights
Enrolments	Participants enrol and share demographic information	Numbers and reach of course e.g. to target group
Step activity	Participants land on and click steps to complete	Level of engagement in the course
Video views	Participants watch videos (part or whole)	Level of interest in videos
Quiz responses	Participants select multiple choice quiz answers	Attention to learning content
Peer review assignments & reviews	Participants submit assignments and review assignments automatically allocated to them	Peer learning activity; content requires qualitative analysis to determine quality
Comments	Participants post a comment	Level of social presence or social learning activity (requires qualitative analysis)

cessation of step activity to indicate that a course design is flawed. However, professionals may not be motivated to complete the entire course, a participant who stops moving forward at a particular point may have achieved their own learning aims, or may have been drawn away to competing priorities. Platforms have begun to use profiles or archetypes based on pre-course surveys to group learners by motivation and other characteristics, for example learners who use MOOCs to improve their career prospects (Walker, 2018). Despite this, data that are automatically collected are never going to tell course designers or researchers everything they want to know. For this reason, it is up to the course designer to embed in-course surveys, discussions and activities (possibly through external tools) that collect the sorts of data that can help answer questions that are important to them. Data from the platforms may answer many of the "what" questions, but the method has to be followed up by post-course surveys and interviews in order to answer the "why?" questions with any confidence.

The limitations in pre-designed platform learning analytics for highlighting the impact of learning in a MOOC indicate that we need to go beyond what is provided by the platform if we want to use MOOCs to improve the quality and value of professional learning. Each educator has to decide what counts as a measure of the value of their MOOC for the professional participants, and then organise the additional data collection they need. The next section considers how we might do that.

25.5.2 A Value Creation Framework for Multimodal Analytics

A method to measure the value of learning for each professional can be found in Wenger et al.'s (2011) framework for tracing value creation in social networks or communities of practice. It is a good starting point for constructing such a methodological approach for professional learning courses, because it supports the "triangulation of multiple sources and types of data" (p. 8) which can be used to attribute outcomes to specific activities. The framework therefore aims to go beyond mere correlation between different sets of data to establish causal links. Moreover, Wenger et al. (2011) observed that while the course participants themselves are the primary recipients of the value created, so too are other stakeholders, including their organisations and their clients or students. The authors argued that what counts as value creation needs to be negotiated by the members of a community or network, and cannot always be neatly defined from the outset. As a result, value creation should be considered in the context of narrative, i.e. personal and collective value creation stories should be constructed to bring meaning to the data.

Wenger et al. (2011) proposed that there are five cycles of value creation. The first cycle is the *immediate value* of the activities and interactions undertaken. This includes social support that participants can derive from membership of a professional community or network – for example, getting tips from colleagues and emotional or practical support with a difficult work problem. The second cycle is *potential value,* also described as "knowledge capital" (Wenger et al., 2011, p. 19) since it involves learning things whose value is to be realised later, such as skills or information. Potential value can also involve developing relationships or gaining access to resources or an increased capacity for learning. The third cycle, *applied value,* involves putting knowledge capital into practice, for example, trying out a suggestion. However, it is not certain that such applications of knowledge gained will be beneficial, so the fourth cycle is *realised value* and is the evidence of improved performance. The final cycle, *reframing value*, is the value that is created when participants use the evidence of impact to reconsider their goals and strategies and what counts as success. This can happen at an individual or organisational level.

MOOCs designed according to a social learning paradigm can have some similarities with the communities and networks that Wenger et al. (2011) use as their focus, particularly MOOCs that target groups of professionals and provide a platform for participants to interact and learn from each other. The five cycles of value creation offer a complex framework to consider the kinds of data that indicate the impact of MOOCs. For example, the platform and other data could provide evidence of the following cycles of value creation:

1. Immediate value: enrolments, evidence of engagement and social interaction in the comments and other contributions, pre- and post-course surveys, post-course interviews;
2. Potential value: quiz responses, self-reports in surveys, comments; peer assignments and reviews;
3. Applied value: contributions in collaborative activities, e.g. Padlet, Googledocs etc.; comments, peer review assignments;
4. Realised value: follow up survey responses for self-reports of impact on others;
5. Reframed value: self-reports of changes in approach e.g. participants becoming a MOOC mentor, skills learnt or online/blended learning within their organisation, or achieving institution-wide policy changes.

While evidence of applied, realised and reframed value that can be accessed via platform datasets can be extremely limited, the list above shows there are many ways we can supplement these data, e.g. with follow up surveys or selected participant interviews that provide the kind of value creation stories that Wenger et al. (2011) propose. Such value creation stories could elicit participants' journeys through the five cycles – that is, their account of engagement in the MOOC, identification of specific resources or ideas deemed of value, how they were applied and with what effect, and their reflections of shifting understandings of their practice. Here, therefore, we propose a way of using platform data within a larger framework of value creation as a type of MMLA, defined by Zhou et al. (2021, p. 29) as a "method that integrates multiple data sources to analyze learners' interactions and examine complex learning processes". The next section presents a selection of case studies from MOOCs that demonstrate what can be gained from this approach.

25.5.3 Tracing the Cycles of Value Creation with MMLA

The case studies presented below both derive from MOOCs that have been created through a Design Based Research approach (see Gerholz & Wagner, 2022, Chap. 23), that seeks to refine a design through progressive iterations informed by evaluation. Both MOOCs were aimed at education professionals (usually teachers). The design of each MOOC was informed by the Conversational Framework (CF) developed by Laurillard to embed a social and collaborative learning design that promoted peer communication to mitigate the lack of educator feedback. Such a design is effective for professionals who are capable of learning independently and from peers.

The first case study addresses the different motivations of professional MOOC learners. Rather than course completion, the immediate value that will keep participants learning is becoming part of a professional online community. MMLA focused on immediate value can help us measure whether MOOC designs are creating social presence. The case study takes an unusual route of using analytics to compare platforms to check a social learning design has been successfully implemented. The second case study looks at how learning analytics can help to design courses that promote potential and applied learning. Learning analytics focused on applied value can provide insights into how we can design environments more effectively to promote impact on practice.

The case studies illuminate the need to combine data to address the two challenges for learning analytics for professional learning. The first of these is to adapt learning analytics to the specifics of professional learning. As we have discussed, Learning Analytics tends to be developed in relation to formal undergraduate or graduate education rather than professional education. In formal education, the assumption is that learners will complete a high stakes assessment at the end of a course, and to not complete is seen as a failure. Students on these courses tend therefore to be motivated by completing and gaining certification. Professionals are likely to be less motivated by completing a course, however, since their focus may be to solve their professional challenges. Secondly, the evaluation of formal education tends to focus on pass rates and attaining certification. The most appropriate evaluation of professional learning is its impact on future practice. Value creation is therefore a more meaningful way of measuring the impact of professional learning.

25.5.4 Case Study 1: Analytics for Immediate Value: Designing Social Presence into a MOOC Platform

This first case study is from a MOOC on Community Based Research created as part of the RELIEF Centre (https://www.relief-centre.org) research which co-designed MOOCs with communities in Lebanon to address challenges in the context of mass displacement. The overarching goal of this research was to explore the ways that digital technologies could support education to achieve more inclusive prosperity in the contexts of mass displacement. Our preliminary research (Laurillard et al., 2018) had identified that MOOCs had the potential to support professionals to share local solutions with others on a large scale, and we wanted to explore which particular combinations of technology, pedagogy and community could make this happen. From initial workshops to engage stakeholders and partners in the co-design process, feedback identified the need to create MOOCs in both English and Arabic. Our ambition was to create as far as possible the same learning experience on two different MOOC platforms – one Arabic and one English.

The challenge was to create a social learning environment with thousands of participants and a set of often limited technical features for discussion with MOOC platforms. Within the Value Creation Framework, this relates most closely to immediate cycle of value, which focuses on creating social presence. The first case study explores how data from the platform was able to show that this had been achieved in a particularly difficult context.

We used the FutureLearn MOOC platform (https://www.futurelearn.com) for the English version of the MOOC since it has a social media style design for discussion that promotes social learning beneath every content step.

For the Arabic version of the MOOC, we chose the Jordan-based Edraak platform. However, the platform affordances were less developed for social learning than FutureLearn, and most Edraak MOOCs featured only video steps and quizzes. Discussion forums were a feature of the platform, but were positioned separately from the content, rather than embedded with the content steps.

To create an equivalent social learning design on Edraak, we embedded a discussion component beneath each step, so that learners could participate in discussions without having to leave the step and go to the separate discussion forum space. Nevertheless, posting a comment required more steps on Edraak (e.g. creating a comment, a title and using mark-up tools for formatting) than on FutureLearn, and receiving notifications of replies was less immediate. For these reasons, and the lack of familiarity with social learning among Edraak participants, we wanted to find evidence that the experience was equivalent across both platforms.

We evaluated the effectiveness of our design intervention by comparing the data on discussion activity on both platforms. Course measures provided by FutureLearn are available through the platform through summary statistics. Measures include: joiners (enrolments); learners (the percentage of joiners who visit at least one step); and social learners (the percentage of learners who post at least one comment). A data transformation process was required to obtain comparable measures from the Edraak datasets. The results are shown in Table 25.2:

Table 25.2 A comparison between enrolments and social learning activity across Edraak and FutureLearn

	FutureLearn		Edraak	
	Count	%	Count	%
Joiners	815		2862	
Learners	520	63.80%	1605	56%
Social learners	82	15.80%	242	14.90%

The data shows that the construction of a social learning design within the Edraak platform (the combination of a discussion prompt and discussion element embedded beneath a content step) is comparable to FutureLearn. Moreover, because of the higher number of joiners (learner enrolments) on the Edraak platform, there was a greater total number of participants in the discussion on the this platform.

This means that we achieved a comparable level of social learning across both platforms, despite the technical and cultural/pedagogical differences. However, the platform data alone do not show the value participants attributed to social learning. The post-course survey showed that 93.7% of Edraak learners ($n = 578$) rated discussions with peers somewhat, very or extremely important for their learning, compared to 90% of FutureLearn learners ($n = 33$), which does indicate immediate value gained from discussion. To understand the value of peer discussion in depth, however, it would be preferable to obtain more detailed and qualitative data. The low number of responses to the post-course survey from FutureLearn participants reduces the meaningfulness of these figures, which are further complicated by similar ratings given to other pedagogical features of the platform. In subsequent MOOCs designed by the RELIEF Centre, evaluation surveys asking specific questions about discussion and collaboration elements of the design were embedded earlier in the course, generating a much more substantial response rate. In summary, therefore, the platform data helped demonstrate the comparative social learning experience, but needed to be supplemented by survey data to enable a better understanding of their value for participants.

25.5.5 Case Study 2: Analytics for Applied Value: Promoting Peer Review

While there is potential for the data traces left by participants to be used to measure immediate and potential value, it is more difficult to measure applied value in MOOCs, and very difficult indeed to measure realised and reframing value. However, for applied value, activities can be designed to encourage participants to share the ways that they are applying ideas to their practice. For example, one of the clearest indicators is the percentage of participants completing a peer reviewed assignment that invites participants to apply their learning to practice.

The second case study focuses on the peer review activity in Blended Learning Essentials, a MOOC designed to support professionals working in the Vocational Education Sector to integrate blended learning in their teaching. The research aim was to evaluate the extent to which peer sharing of practice could be supported in a MOOC, and the research findings informed the approach taken in the RELIEF MOOCs described above. We invited participants to submit a learning design relevant to their practice for peer review, and then to review others' learning designs. The demanding nature of these MOOC steps (where participants submitted a learning design and then reviewed 3 others), in comparison with previous steps

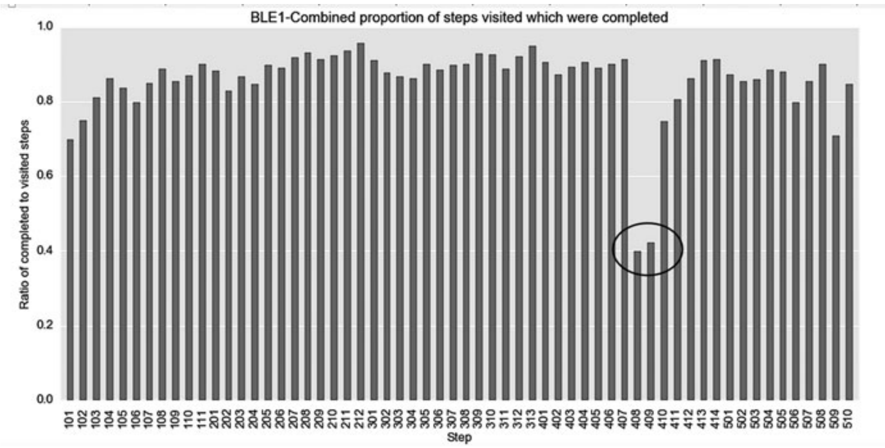

Fig. 25.1 Ratio of completed to visited steps in Blended Learning Essentials: Getting Started Phase 1 (runs 1–3)

(which might feature a video or an article to read), meant that the completion rates of the MOOC steps featuring the peer review were much lower, at 45%, than for the other steps, at close to 90% on average. When participants visit a step, the platform records that action. When participants mark the step complete, that action is also recorded. For most steps, marking a step complete is achieved simply by clicking a button. However, for a peer review step to be marked complete, the participant has to submit a response to the assignment and a specified number of reviews of others' submissions. Figure 25.1 shows the ratio of steps visited to completed from the combined data from the first three runs of Blended Learning Essentials: Getting Started, showing the drop in completion for the peer review steps 409 and 410.

Despite the dip in participation, there were over 900 submissions across the three runs of the MOOC, which also accumulated a resource of vocational teacher's design ideas for blended learning. Moreover, the submissions were of high quality showing how ideas were being applied to practice (applied value), with only 28 assignments in total not containing a detailed learning design.

Participants' comments in the discussion indicated they attached a high value to the reviews received, with participants saying the process gave them reassurance (immediate value), and offered positive suggestions for changes to their designs (potential value). The participants also said they found value in performing the reviews, because, for example, reviewing someone else's submission enabled them to reflect on their own use of technology in planning their teaching (potential value).

So, we could see from MMLA that peer review was a high value activity, and indicated that changes to the course design that encouraged wider engagement would benefit would participants benefit greatly. As a result, we redesigned the peer review activity in a later run to scaffold engagement with these challenging steps. We spread out the work required for the assignment over the previous steps, so

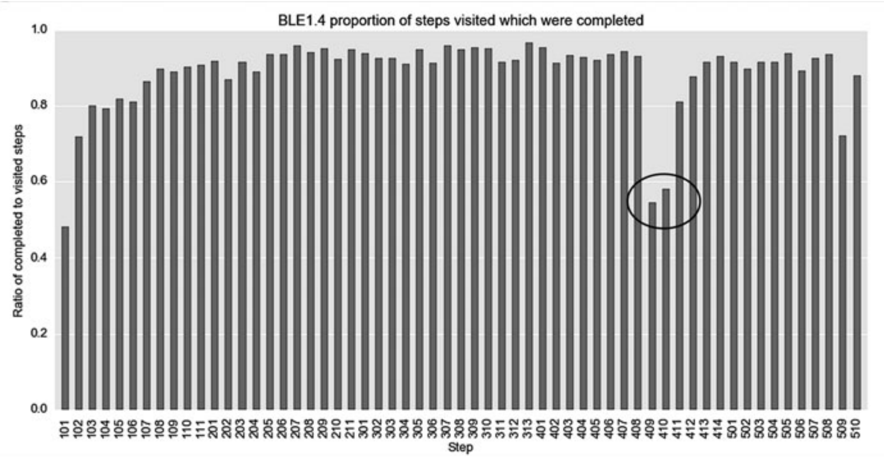

Fig. 25.2 Ratio of completed to visited steps in Blended Learning Essentials: Getting Started Phase 2 (run 4)

that it was easier for participants to complete the peer review steps. Following these changes, which included renumbering the peer review steps 408 and 409, it is possible to see an improvement in the ratio of steps completed to steps visited for the peer review steps, increasing from 40% in phase 1 to 55% in phase 2 (Fig. 25.2).

The Value Creation Framework is able to guide the questions we ask of platform data to show the many ways that MOOCs can create value for participants. However, MMLA will not be able to tell the whole story. Since applied, realised and reframing value are largely generated after the MOOC, platform engagement data cannot evidence this. Here, interviews were able to show impact on participants' learners. For example, a teacher trainer who had embedded the course within a Post Graduate Certificate of Education (PGCE) in Post-compulsory education reported widespread change of practice among his trainees, saying that in 10 years of observation, he would only see trainees use "Powerpoint and occasionally interactive whiteboards and not much else", however, since the MOOC he witnessed trainees becoming more innovative and "experimenting with some of the packages they seen used in the MOOC" (applied value). Other interviewees reported observing teachers' use of quiz tools and the positive effect on learner engagement (realised value). Interviews also revealed that the course had changed institutional approaches to blended learning "I use the course to deliver CPD [Continuing Professional Development] at the college" (reframing value).

This case study shows that, by embedding MMLA in a Value Creation Approach they can show us where value is being created for participants, and where we can make changes to increase this value. It also shows the limitations of MMLA, and where off-platform data can complete the picture.

25.6 Conclusions

In this chapter we have provided examples illustrating how LA provide insights about professional learning. In the race to scale up professional learning globally, new ways of supporting learners at scale are needed to help teachers to understand how to support learners in large cohorts. The idea of automated systems to support learners at scale has been around for at least four decades and has led to the development of a range of analytics methods.

Learning Analytics provide helpful insights made available as data visualisations, recommendations, adaptations and predictions. However, these analytics methods largely have been developed for formal education and are not always relevant for professional learning contexts.

There are two main reasons for this, first, professional learning is more complex than formal education and requires reconsideration about what are the indicators of and influences within professional learning and, therefore, what data is needed. This requires consideration of how professionals learn, both in formal education and also through everyday work tasks. LA tends to use data available on a digital platform. While these data provide some insights into learning performance, the data that are useful to analyse learning are likely to be multimodal (e.g. digital trace data indicating learners' emotional states using facial recognition, heartrate, body temperature and skin conductivity – a method used in 'lie detector' tests to measure stress levels). There are already examples of these sorts of data being gathered using computer cameras as well as wearable, smart devices, such as smart watches and wristbands. However, the interpretation of these data are complex, controversial and raise questions around their use and human rights issues (Berendt et al., 2020). With the likelihood of misinterpretation, these sorts of data should be used with caution.

Second, analyses are based on assumptions underlying data analysis around how professionals learn and the interpretation of these data. Some assumptions that are helpful in formal education (for example, the assumption that all learners intend to complete a course) may not hold true in professional learning – especially in settings beyond the classroom when professionals are learning on-the-job or applying knowledge learned in the classroom to work tasks.

To resolve these problems, we propose a widening of the types of data measures and analysed. MultiModal Learning Analytics (MMLA) are already being developed for professional learning. However, these forms of analyses still tend to adopt indicators from formal learning settings and theories. We propose a methodology that considers more broadly the different kinds of value participants gain from professional development based on a Value Creation Framework.

Wenger et al.'s (2011) Value Creation Framework can identify data to be gathered across a variety of modalities, ranging from conventional forms of digital data (trace data) measured through learning analytics, to textual data gathered as written or spoken data, to face recognition, gesture data and so on. Together these different data types provide detailed indication of the value created for professionals as they learn. Different professions will benefit from distinct multimodalities that are related to their professional practice. For example, teachers tend to use facial recognition, body language and spoken feedback to assess a learner's progress while health professionals use gestures, postures and so on (Weldon et al., 2013).

To be effective in understanding complex learning processes, Professional Learning Analytics has to better reflect the demands of diverse professions and has to be better integrated within work. An indication we have reached this goal is when learning is indistinguishable from work.

References

Ahuja, N. J., & Sille, R. (2013). A critical review of development of intelligent tutoring systems: Retrospect, present and prospect. *International Journal of Computer Science Issues (IJCSI), 10*(4), 39.

Anania, J. (1982). The effects of quality of instruction on the cognitive and affective learning of students (Doctoral dissertation, University of Chicago, 1981). *Dissertation Abstracts International, Ļ2*, 4269A.

Anderson, J. L. (1983, January). Autonomous systems intelligence. In *Proceedings of the 1983 annual conference on computers: Extending the human resource* (pp. 229–233).

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207.

Balakrishnan, G., & Cooetzee, D. (2013). *Predicting student retention in Massive Open Online Courses using Markov models* (Report No. UCB/EECS-2013-109). University of California at Berkeley. Retrieved 30 July 2021 from https://www2.eecs.berkley.edu/Pubs/TechRpts/2013/EECS-2013-109.pdf

Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: Learner choice and fundamental rights. *Learning, Media and Technology, 45*(3), 312–324. https://doi.org/10.1080/17439884.2020.1786399

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Burke, A. J. (1984). Students' potential for learning contrasted under tutorial and group approaches to instruction (Doctoral dissertation, University of Chicago, 1983). *Dissertation Abstracts International, 44*, 2025A.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695.

Conole, G. G. (2013). MOOCs as disruptive technologies: Strategies for enhancing the learner experience and quality of MOOCs. *Revista de Educación a Distancia (RED)*, (39).

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning, 34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Engeström, Y. (2016). *Studies in expansive learning: Learning what is not yet there*. Cambridge University Press.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273.

Eraut, M., & Hirsh, W. (2010). *The significance of workplace learning for individuals, groups and organisations*. Skope.

EU. (2021). *Digital automation and the future of work*. Retrieved 30 July 2021 from https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656311/EPRS_STU(2021)656311_EN.pdf

Fuller, A., & Unwin, L. (2003). Learning as apprentices in the contemporary UK workplace: Creating and managing expansive and restrictive participation. *Journal of Education and Work, 16*(4), 407–426.

Gasevic, D., Siemens, G., & Rosé, C. P. (2017). Guest editorial: Special section on learning analytics. *IEEE Transactions on Learning Technologies, 10*(01), 3–5.

Gerholz, K.-H., & Wagner, A. (2022). Design-based research – Grounding, understanding and empirical illustration in the context of vocational education. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 513–534). Springer.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education, 23*, 18–26.

Hollands, F., & Tirthali, D. (2014). *MOOCs: Expectations and reality*. Center for Benefit-Cost Studies of Education. https://doi.org/10.1109/EDOC.2006.60

Jiang, S., Williams, A. E., Warschauer, M., He, W., & O'Dowd, D. K. (2014). Influence of incentives on performance in a pre-college biology MOOC. *The International Review of Research in Open and Distributed Learning, 15*(5), 99–112.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current directions in psychological science, 7*(3), 67–72.

Laurillard, D., & Kennedy, E. (2020). MOOCs and professional development: The global potential of online collaboration. In C. Callender, S. Marginson, & W. Locke (Eds.), *Changing higher education for a changing world* (pp. 157–170). Bloomsbury Academic. https://doi.org/10.5040/9781350108448.0020

Laurillard, D., Kennedy, E., Wang, T., Escorcia, G., & Hooker, M. (2018). *Learning at scale for the global south*. Quezon City.

Littlejohn, A., & Hood, N. (2018). *Chapter 4: Reconceptualising learning in the digital age: The [un] democratising potential of MOOCs* (p. 64). Springer.

Littlejohn, A., & Hood, N. (2019). *Reconceptualising learning in the digital age: The [Un] democratising potential of MOOCs*. Singapore: SpringerBriefs in Education.

Littlejohn, A., & Margaryan, A. (Eds.). (2013). *Technology-enhanced professional learning: Processes, practices, and tools*. Routledge.

Littlejohn, A., & Pammer-Schindler, V. (2021). Technologies for professional learning. In D. Gijbels & C. Hartelis (Eds.), *Handbook of research approaches on workplace learning*. Springer.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education, 29*, 40–48.

Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018, March). Physical learning analytics: A multimodal perspective. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 375–379).

Milligan, C., & Littlejohn, A. (2017). Why study on a MOOC? The motives of students and professionals. *The International Review of Research in Open and Distance Learning, 18*(2), 92–102.

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163.

Nistor, N., & Hernández-Garcíac, Á. (2018). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior, 89*, 335–338.

Pappano, L. (2012, November 2). The year of the MOOC. *New York Times*. Retrieved from https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace

Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. Wiley.

Shum, S. B., & Crick, R. D. (2012, April). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 92–101).

Shum, S. B., & Crick, R. D. (2016). Learning analytics for 21st century competencies. *Journal of Learning Analytics, 3*(2), 6–21.
Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*(3), 3–26.
Siemens, G., & Long, P. (2011, September/October). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5). Retrieved 30 July 2021 from http://er.educause.edu/articles/2012/7/~/link.aspx?_id=82AE6F528BDC4EBFBC17FCD75B3E3E5C&_z=z
Skrypnyk, O., de Vries, P., & Hennis, T. (2015, May 18–20). *Reconsidering retention in MOOCs: The relevance of formal assessment and pedagogy*. Paper presented at the Third European MOOCs Stakeholders Summit, Université Catholique de Louvain, Mons, Belgium.
Titsworth, B. S. (2001). Immediate and delayed effects of interest cues and engagement cues on students' affective learning. *Communication Studies, 52*(3), 169–179.
Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154.
Walker, T. (2018). *Who are our learners? Part 1: What we did and why – FutureLearn*. Retrieved 29 July 2021 from https://www.futurelearn.com/info/press/research-insights/learners-part-1
Weldon, S. M., Korkiakangas, T., Bezemer, J., & Kneebone, R. (2013). Communication in the operating theatre. *Journal of British Surgery, 100*(13), 1677–1688.
Wenger, E., Trayner, B., & De Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework Rapport 18*. Heerlen.
Zhou, Q., Suraworachet, W., & Cukurova, M. (2021). Different modality, different design, different results: Exploring self-regulated learner clusters' engagement behaviours at individual, group and cohort activities. In *Proceedings MAIED 2021* (pp. 28–40). Online: http://ceur-ws.org/Vol-2902/

Allison Littlejohn (PhD) is a Professor and the Director of the UCL Knowledge Lab, a centre exploring the future of learning and communication with technology at University College London. She is a learning scientist, specialising in professional and digital learning. Her research focuses on the role of professional learning in addressing global challenges and her work has made contributions to the understanding of how people learn for work across the Energy, Finance, Health, Education and International Development sectors. Formally Dean (Learning & Teaching) at the University of Glasgow, UK; Academic Director for Digital Innovation at the Open University; Founding Director of the Caledonian Academy at Glasgow Caledonian University, UK; and a Professor of Education, University of Dundee, UK.

Eileen Kennedy (PhD) is a Principal Research Fellow based at UCL Knowledge Lab. She researches ways of realising the transformative potential of digital technologies for higher education with the Centre for Global Higher Education and the RELIEF Centre, where her focus is digital education in the contexts of mass displacement.

Diana Laurillard (PhD) is a Professor of Learning with Digital Technologies, UCL Knowledge Lab. She leads research on digital design tools for teachers and scaling up online learning. Formerly Head of the e-Learning Strategy Unit at DFES (2002–2005); Pro-Vice Chancellor for Learning Technologies, the Open University (1995–2002); member of the Dearing Committee.

Chapter 26
Longitudinal Case Study Research to Study Self-Regulation of Professional Learning: Combining Observations and Stimulated Recall Interviews Throughout Everyday Work

Katrien Cuyvers, Piet Van den Bossche, and Vincent Donche

Abstract Professional learning reflects critical processes of change whereby one modifies and extends prior competencies while performing one's job. Over the past two decades, the need has emerged and grown for insights on how employees take responsibility for their own learning and engage in self-regulation of professional learning. However, the process of measuring professional learning as well as self-regulation of professional learning during everyday work has raised difficult methodological problems for various reasons. The retrospective, cross-sectional, self-report measurement techniques often used, tend to de-contextualise learning from the complex environments in which professionals operate. Under such techniques, study participants are asked to make abstractions of this complexity to self-report regarding possibly implicit, multifaceted competencies and metacognitive strategy use as features of self-regulated learning. In this chapter, we offer an alternative

K. Cuyvers (✉)
Faculty of Social Sciences, Department of Training and Education Sciences and Antwerp School of Education, University of Antwerp, Antwerp, Belgium

Faculty of Behavioural, Management and Social Sciences, Department of Educational Science, University of Twente, Enschede, The Netherlands
e-mail: Katrien.cuyvers@uantwerpen.be

P. Van den Bossche
Department of Educational Research and Development, School of Business and Economics, Maastricht University, Maastricht, The Netherlands

Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium
e-mail: Piet.vandenbossche@uantwerpen.be

V. Donche
Faculty of Social Sciences, Department of Training and Education Sciences, University of Antwerp, Antwerp, Belgium
e-mail: Vincent.donche@uantwerpen.be

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_26

approach via a longitudinal multiple case study design combining long-term observations with immediate consecutive stimulated recall interviews, towards building a more dynamic and situated understanding of professional learning through which to explore participants' self-regulation. Using both 'on-line' and 'off-line' measurement techniques, the proposed interactive approach was empirically applied to investigate self-regulation of professional learning in medical practice. Without pretentiously suggesting that this is the ultimate research solution, we aim to outline the approach, its opportunities and challenges, how to tackle these challenges, and how the approach's research insights could function to advance theory-building on professional learning in general—and self-regulation of professional learning in particular—in everyday work.

Keywords Professional learning · Self-regulated learning · Workplace learning · Longitudinal multiple case study · Long-term observations · Stimulated recall interviews

26.1 Introduction

Due to rapid changes in society and working lives, employers and employees have sought out strategies to ensure a certain level of competence at the job in the past few decades (Tynjälä, 2008). Besides traditional classroom training, forms of workplace learning ranging in their degree of integration with work offer abundant opportunities in this respect. The importance of integrating learning with the job has become a widespread belief and emerging practice among researchers, practitioners, and policymakers. In this chapter, 'professional learning' is defined as learning in the workplace which is entirely integrated with work (Cuyvers et al., 2021).

Supportive conditions are required to enable professional learning in the work environment (Ellström, 2001). Individual-related factors also play a vital role in professional learning (Tynjälä, 2013). For instance, researchers have assumed a professional's ability to self-regulate one's learning to be critical for ongoing improvements in performance and in the adoption of new ways of working (Cuyvers et al., 2021; Littlejohn et al., 2016).

Together with the increasing awareness on the importance of professional learning, interest has grown over the past two decades for insights on how employees go about shaping their learning process, or 'self-regulate professional learning'—for which hereafter we use the acronym 'SRpL' (Cuyvers et al., 2020). Measuring SRpL in real-time, ongoing professional experiences has become an important goal amid expanding on the work of those who have advocated for a dynamic and situated understanding of SRpL (Cuyvers et al., 2021; Endedijk & Cuyvers, 2021; Littlejohn et al., 2016). However, the process of measuring SRpL in all its complexity, as well as of grasping its role in improving workers' skills and other outcomes, has given rise to major methodological challenges and questions in empirical research (Cuyvers et al., 2020).

In this chapter, we first elaborate upon the concept of SRpL. We then describe challenges in prevalent methodological paths for SRpL. Finally, we propose an alternative methodological approach to capture SRpL as an ongoing process in real-life work environments.

26.2 The Concept of Self-Regulation of Professional Learning

SRpL refers to professionals' ability to proactively, reactively, or implicitly engage in self-regulatory strategies to shape their learning process, elicited by the challenges in daily practice (Cuyvers et al., 2021). In SRpL in the workplace, self-regulatory strategies are behavioural, cognitive, metacognitive, and affective in nature (Cuyvers et al., 2021; Sitzmann & Ely, 2011), taking up different roles in the professionals' learning process. That is, some of the strategies engaged in are conditional for other strategies initiating, advancing, and evaluating the learning process (Cuyvers, 2019; Cuyvers et al., 2021). Interrelated self-regulatory strategies engaged in by the professional with feedback loops dynamically compose the unfolding learning process which can be evoked by a work challenge on the one hand, and interrupted by the work being performed on the other (Cuyvers et al., 2021). Given its highly implicit nature and intertwining with performance, professional learning and SRpL could be hard for observers and even learners to recognise or distinguish between. Based on empirical research on medical specialists in the clinical environment (Cuyvers, 2019; Cuyvers et al., 2021), Figure 26.1 depicts self-regulatory strategies engaged in, with arrows indicating the interrelatedness among strategies.

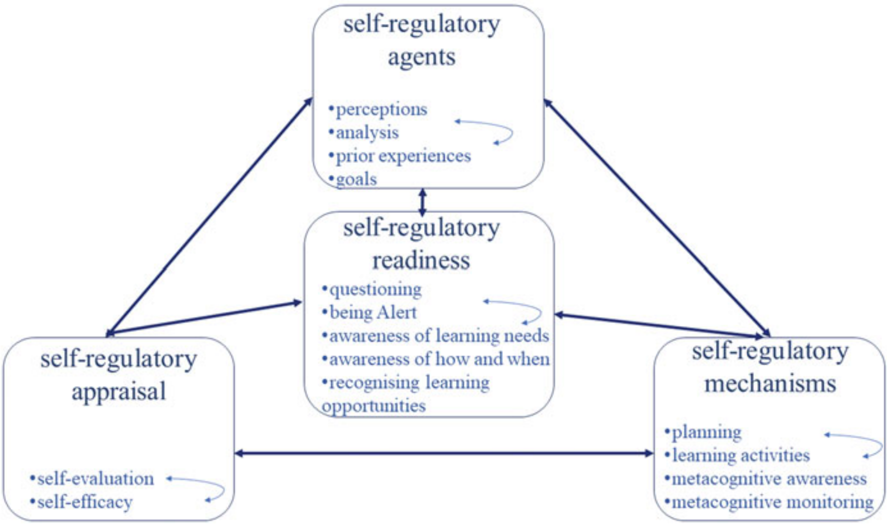

Fig. 26.1 Interrelated self-regulatory strategies in SRpL. (Cuyvers, 2019)

As Cuyvers et al. (2021) described, in medical practice, the process of SRpL starts when a medical specialist encounters a performance-related situation in which learning could take place. This calls for alertness on the part of the medical specialist (self-regulatory readiness, as depicted in the centre of Fig. 26.1) regarding the current situation: alertness both for the danger of routine in performance and for competencies being challenged. In this alertness, the medical specialist questions and reflects on their own performance and competencies required (self-regulatory readiness strategy). This questioning is engaged in, in relation to an activation of awareness (self-regulatory readiness) both of the competencies, and of the potential gaps at hand for the medical specialist. Hence, they can recognise the situation and patient as an opportunity for learning (self-regulatory readiness strategy). Amid these self-regulatory readiness strategies, the medical specialist could perceive the patient or situation in different manners (self-regulatory agents): for instance, as either tense or exciting (affective perception; self-regulatory agent) or as difficult (cognitive perception; self-regulatory agent) in relation to the questioning engaged in as self-regulatory readiness strategy.

Thus, the medical specialist could analyse the situation (self-regulatory agent) not only from the perspective of job performance, but also from that of learning in relation to identifying missing competencies, knowledge, or skills. Engagement in the self-regulatory strategies visualised in Fig. 26.1 evolves while the medical specialist pursues strategies related to other strategies. Given the complexity of the unfolding of SRpL over time as a process, its interrelated components, and its integration with work, important challenges for empirical research come to the fore, as outlined in the next section.

26.3 Measurement Challenges in Prevalent Methodological Paths

A recent systematic literature review showed that systematic efforts to measure SRpL began in the last decade (Cuyvers et al., 2020). The existing body of research revealed various challenges related to such measurement. To start with, most of the SRpL studies reviewed transferred measurement approaches to self-regulated learning (SRL) from educational to work settings (Cuyvers et al., 2020). Doing so, the de-contextualised, cross-sectional, self-report techniques used in such research have tended to isolate the phenomenon of interest from the complex environment in which professionals operate in, often measuring overall professional development activities on a more general level (Cuyvers et al., 2020).

Despite widespread evidence for the validity of instruments assessing SRL in educational contexts (Schunk & Greene, 2018), most of these instruments have not been developed to gauge less intentional, less planned professional learning in the workplace. Also, study participants are asked to make retrospective abstractions of the complexities in their work, to self-report on the behavioural, cognitive, metacognitive, and affective self-regulatory strategies used. As such, the dynamic process of SRpL, with interrelated strategies also highly intertwined with performance, has tended to be measured as a relatively static aptitude, potentially biased by memory failure or the notion of socially-desirable answers (Cuyvers et al., 2020; Moorman & Podsakoff, 1992; Rausch, 2014; Veenman, 2011).

To tackle the issue of gauging its dynamics, research on SRL in educational contexts over the last decade has focused on capturing SRL during the task itself, as a process in the actual learning environment—referred to as 'on-line'—via trace data such as eye-tracking, log files, and physiological sensors (Spliethoff & Abele, 2022, Chap. 8; Azevedo et al., 2010, 2018; Jossberger, 2022, Chap. 21; Winne, 2010). As on-line event measures, trace data have often been collected from technology-enhanced learning environments (Bernacki, 2018). In measuring SRpL, the ability to use trace data depends on the presence and role of technology in the work of the professionals under investigation. Also, although online event measures such as traces are described as interesting to measure SRL, the value of other techniques such as think aloud, surveys, and interviews, especially when augmented, is still expressed by researchers in the field (Azevedo et al., 2010; Winne, 2010).

Moreover, in research on SRL in the workplace during internships, besides 'off-line', self-report questionnaires (Bransen et al., 2020; Brydges et al., 2020; Vrieling-Teunter et al., 2021), diaries or structured learning reports have been used as off-line event measures as well (Endedijk et al., 2016; Rausch et al., 2022, Chap. 3). But although it offers valuable insights on SRL situated in the workplace as a learning environment, this method is time-consuming and labour-intensive for professionals, and it does not allow for a more process-driven assessment of SRpL.

Finally, professional learning and regulatory strategies are often largely implicit as well as complex,—with its integration with work making it difficult to distinguish between learning and work for valid measurements. It requires a complex array of competences, thereby defying simple representations of what is being learned and how (Cuyvers et al., 2016; Furner & Steadman, 2004; Eraut, 2000, 2004; Rausch, 2014).

Thus, although the importance and research field of SRpL have grown considerably in the last decade, the prevalent measurement methods and instruments have hardly grasped the complexity of the process of SRpL in real-life contexts.

26.4 Proposing an Alternative Methodological Approach for SRpL Measurement

To meet the challenges we described, we introduce here an alternative interactive approach to measure SRpL. We propose a longitudinal multiple case study to explore SRpL as it actually unfolds in real-life professional learning environments. In this design, long-term direct observations as an on-line event measurement technique are combined with immediate stimulated recall interviews as a self-report off-line event technique. While we do not mean to suggest that this is the only viable alternative, in triangulating multiple data collection techniques across time, the situated longitudinal methodological perspective does go beyond the well-trodden path of cross-sectional, off-line self-report questionnaires. It enables the reduction of the so-called 'mono-method bias' when measuring SRpL as it evolves over time, accounting for the conditions in real-life work environments.

In the next sections, we first discuss the methodology, followed by a brief illustration of a previously reported empirical application (Cuyvers et al., 2021). We then elaborate upon the method's opportunities and address some major challenges which have arisen in bridging the gap between the proposed on-line and off-line techniques used. Finally, we put forth suggestions for future research on SRpL.

26.4.1 Why a Longitudinal Multiple Case Study Design Matters in Exploring Self-Regulation of Professional Learning

Methodological literature in social sciences has often advocated for the use of case studies to research real-life phenomena (Yin, 2014). This approach allows for in-depth explorations of contextual characteristics and conditions in relation to the phenomenon under investigation (Yin, 2014). Multiple case study designs have been particularly known for facilitating the replication and pursuit of theoretical propositions, leading to the acquisition of more compelling and generalisable evidence (Yin, 1999, 2014). Since case study research relies on multiple sources of evidence, thereby triangulating data, it tends to increase the credibility of findings (Noble & Heale, 2019; Stake, 1995; Yin, 1999, 2014).

Gauging how SRpL unfolds requires an in-depth investigation of temporal and sequential features (Cuyvers et al., 2020; Endedijk & Cuyvers, 2021). The longitudinal tracing of how respondents and processes change over time could enable the achievement of this goal (Bernacki, 2018; Yin, 2014). As such, a longitudinal multiple case study design could address the demand to measure SRpL as a real-time, dynamic process situated in professionals' work environments. Using a multi-method approach in this way, with both on-line and off-line data collection, could successfully tackle challenges in measuring behavioural, cognitive, metacognitive, and affective self-regulatory strategies. In line with Yin (2014), we propose long-term direct observations as a key method in case study research, augmented with immediate, consecutive stimulated recall interviews.

Not only has theory proposed observations as an on-line measurement to assess actual ongoing behaviours (Veenman, 2007; Wolters et al., 2011), but also as a means to deliver rich evidence on what one is learning, and how, while performing one's job (Eraut et al., 1998; Eraut, 2000, 2004). Despite the value of direct observations in case study fieldwork, they do not allow for the simultaneous study of engagement in covert cognitive and metacognitive regulatory strategies in SRpL. Indeed, researchers have yet to figure out how covert regulatory strategies are externalised into observable behaviours. Nevertheless, observations provide clues for the tracking down and recalling of covert and implicit learning which might have taken place (Furner & Steadman, 2004).

To address the shortcomings in observations, we propose the integration of stimulated recall interviews into the longitudinal case study design, as these clues can be used as prompts to mediate verbalisation and elicit participants' thoughts and strategies from a real-life activity (Henderson & Tallman, 2006; Wolters & Won, 2018). Previous research on metacognition recognised stimulated recall interviews as a useful means to capture metacognitive strategy use (Henderson & Tallman, 2006; Veenman, 2005; Veenman et al., 2006; Wolters et al., 2011). To establish validity, such literature recommended the use of a protocol with open-ended questions, including all relevant dimensions of the assessed construct (Henderson & Tallman, 2006; Wolters & Won, 2018). Notwithstanding the fact that only conscious thoughts and strategies can be reported, research indicated that stimulated recall was satisfactorily reliable when participants were prompted and questioned within the limits of a 48-h period (Henderson & Tallman, 2006).

To summarise, research aiming to map SRL at the workplace, including SRpL, and its process across time can benefit from applying longitudinal case study research designs. Especially when this is combined with long-term observations of verbal and non-verbal behaviours as prompts for immediate stimulated recall interviews to elicit the verbalisation of thoughts and embedded metacognitive strategies. The interactive approach using observation and self-report measures (via on-the-spot, workplace interviews) can offer important comprehensive insights into overt actions and behaviours, as well as into covert thoughts and metacognitive strategies.

We now turn to illustrating the application of the proposed methodology via an empirical study (Cuyvers et al., 2021), before detailing the method's opportunities and challenges, as well as how to tackle the latter.

26.4.2 How Medical Specialists Self-Regulate Their Learning: Illustration of the Research Approach

We employed the longitudinal case study design introduced above to investigate the SRpL process during medical specialists' daily practice in a clinical environment (Cuyvers et al., 2021). The study aimed to unravel the dynamic SRpL process by investigating (1) which overt and covert SRpL strategies medical specialists adopted in a real-life clinical environment (RQ1), and (2) how the process of SRpL evolved dynamically through time in relation to physicians' job performance (RQ2). Thirteen physicians from diverse specialties participated in this study: an endocrinologist, a cardiothoracic surgeon, a gynaecologist, a neurologist, a neurosurgeon, two emergency physicians, three radiologists, an intensive care specialist, a pathologist, and a paediatric reconstructive urologist. All came from hospitals in Flanders (the Dutch-speaking part of Belgium).

To illustrate the research method, the upcoming section details the key measurement actions performed in the study. First, we describe which behaviours we decided to observe, and why ('behaviours under observation'). Second, we elaborate on the choices regarding the observations' time aspects. Third, we report on how we carried out the observations. Fourth, we describe how we conducted the stimulated recall interviews. Finally, we describe the process of data preparation, followed by its analysis (Cuyvers et al., 2021).

26.4.2.1 Behaviours Under Observation

As mentioned earlier, the existing body of research on SRL in various contexts has not yet determined which behaviours could be observed as externalisations of SRL strategies. Hence, for the purposes of this study, our observations focused on both the verbal and non-verbal behaviours of physicians. We followed the physicians around ("shadowing") while noting all conversations and discussions with colleagues of the same or other medical specialties or disciplines, and with patients and family (Cuyvers et al., 2016). We recorded each interaction in as much detail as possible—person, time, what was said—to glean a thorough perspective from the observations on how interactions could provide opportunities for SRpL. We also paid attention to and made notes on facial expressions, as a non-verbal behaviour, and on all behavioural indications of learning and potential SRpL strategy use. Making thick, rich descriptions which were accessible to others was an important goal in our fieldwork process, to allow for replication and/or verification and de-briefing, and help to minimise investigator bias and ensure the validity and reliability of data collection (Bakeman, 2000; Morse, 2018).

26.4.2.2 Time Aspects of the Observations

All physicians in our sample were observed during their daily medical practice. We selected the time slots based on the different professional activities of each specialist. This could mean that the researcher accompanied the physician during surgery on Monday morning, consultations on Wednesday, administrative duties on Thursday afternoon, and formal meetings on Friday. Hence, we could witness a variety of situations and co-occurring competencies for which learning could be necessary or beneficial, and access potential learning experiences.

Consultations, ward rounds, informal meetings, and surgery all provided valuable opportunities for observation. Moreover, writing requests for technical investigations, reading patient files and other reports, and consulting on radiographic images were examples of professional activities that could lead to learning, offering observable clues for ongoing SRpL. In this study, the medical specialists were 'shadowed', which meant that the researcher sat by, stood by, and followed the physicians as they went about their duties, while using equipment to make written and audiotaped records of the activities and interactions (see detailed description Sects. 26.4.2.3 and 26.4.2.4).

Thorough considerations on time-related aspects led to intermittent shadowing of the medical specialists: on average, each physician was observed four times, for 4 h, so on average for a total of 16 h. We chose time intervals according to their schedules to, on the one hand, allow space for the investigation of possible dynamics in SRpL strategy use; and on the other, to reduce the risk of memory failure and reminiscence bias, and hence of invalid data (Yin, 2018). As such, we organised a maximum interval of five working days between two observation moments. If this rule could not be kept, for example due to a planned vacation, a telephone call was made by the researcher to the physician to follow up on data collection regarding SRpL strategies engaged in between the last visit of the researcher and the last day before the physician's break.

26.4.2.3 Performance of the Observations

Ethical considerations discourage video recordings in medical practice, and the complexity of the behaviours under observation hinders the adoption of a structured scheme to note a priori determined behaviours. The collection of unstructured data in the clinical environment can be considerably less complicated, if answers to open-ended questions during everyday work can be quickly registered. We used a clipboard, paper, and pencil for this purpose. To facilitate registration and later transcriptions, we employed a protocol pre-developed by the researcher for making observation notes, as illustrated in Fig. 26.2.

26.4.2.4 Process and Structure of the Stimulated Recall Interviews

All indications for potential SRpL we observed prompted in loco stimulated recall interviews on beliefs, thoughts, and intentions associated with specific overt behaviours and potential covert SRpL strategies. In our study, this sometimes took place during lunch or a short coffee break, but mainly in between two patient consultations or ward rounds, while going from one room to another, immediately before and after surgery, during the periods of surgery which allowed the physician to answer questions, and following emergency situations. Also, each new observation moment started with asking what had happened after the researcher had left an earlier observation moment. We used a stimulated recall protocol with open-ended questions to explore developed based on the insights of SRL and professional learning elements (Cuyvers et al., 2020) to explore, in a semi-structured manner, the nature and process of SRpL (see Fig. 26.3 for sample questions).

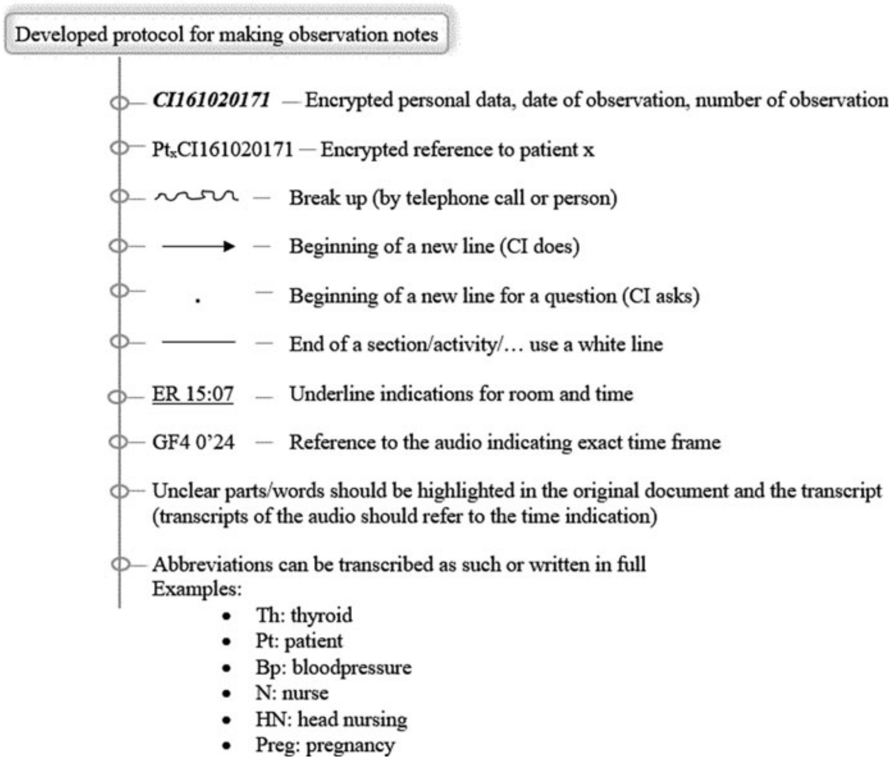

Fig. 26.2 Illustration of the pre-developed protocol during observations

26.4.2.5 Data Preparation for Analysis

All stimulated recall interviews were audio-recorded and transcribed. Field notes were also transcribed. After collecting and transcribing the data, we had to prepare it for analysis.

Because SRpL is a process, and we propose measurements to enable its capture, the final set of data analysed should reflect such a process. Therefore, we chronologically integrated all data into a longitudinal database for each case. In matching the type of data collected to the process, we distinguished between self-report data and observational data to clearly classify the description of what was overtly seen and heard and what was covert but made explicit via the stimulated recall interviews, respectively. We used different colours to indicate the different types of data in our analysis, but underlining the data from the observations while italicising that from the stimulated recall interviews is also possible, as illustrated in Fig. 26.4. This longitudinal database excerpt relates to a cardiothoracic surgery in which Henry is about to perform an anastomosis with four coronary bypass grafts.

The time indications of the stimulated recall interviews in the field notes allowed us to integrate the stimulated recall data into the observational data as closely as

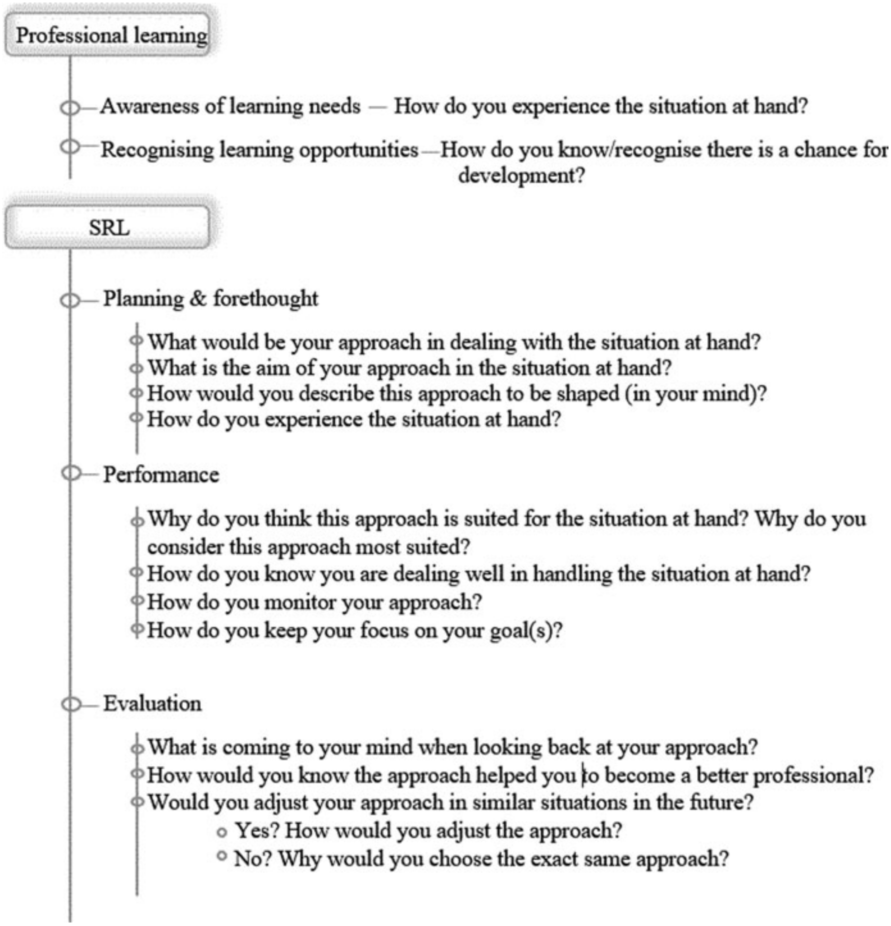

Fig. 26.3 Sample questions in the SRpL stimulated recall protocol

possible to the actual thoughts and potential covert strategies engaged in. The longitudinal databases for each participant yielded the data to be analysed.

26.4.2.6 Analysis and Main Results

In our study (Cuyvers et al., 2021), we selected three cases from the entire sample and archived the longitudinal datasets in the Nvivo 12 software, which supports qualitative data analysis. To answer the first research question, we performed qualitative content analysis via a code tree based on the mainstream SRL theories, interpreted and expanded on with first insights on SRL strategies potentially important for SRpL (Cuyvers et al., 2020, 2021; Cuyvers, 2019). We conducted a deductive, vertical, within-case analysis, assigning codes to parts of the transcripts

"H24042018

8.05: *At the ICU… …*

Audiorecording n°6: 00.00 (on the way from the ICU to the operating room)

Henry says that the challenge is to perform the procedure correct but also to not make steps unnecessary for the individual patient… Henry says this latter is difficult…He says he is alert for the danger of making mistakes. He says that people make mistakes, people doubt themselves from time to time… Henry says he is still searching for his own way, without forgetting things, skipping steps, or not performing all the necessary steps thoroughly… Henry says that he is lefthanded which means he has to do things differently…

…

Henry continues his preparation to perform the operation, looks at his watch… He disinfects his hands asking the assisting nurse to check a certain machine because it is very noisy."

Fig. 26.4 Illustration of the integration of observational and stimulated recall data in a longitudinal database, with an excerpt of the data of cardiothoracic surgeon Henry, from a 23-page document

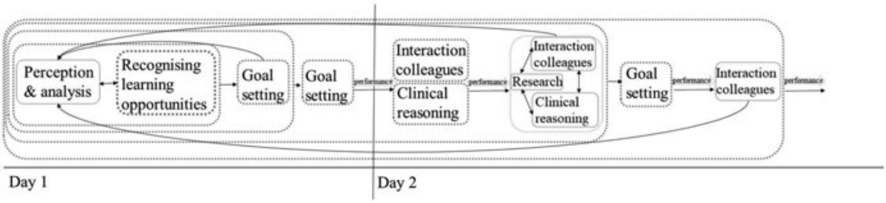

Fig. 26.5 SRpL trajectory composed by relations between SRpL strategies in which physicians engaged

which we interpreted as referring to SRpL strategies. We sought commonalities and differences among the data within each case, thereby creating internally homogeneous and externally heterogeneous codes and categories (Graneheim & Lundman, 2004).

We followed this within-case analysis with a cross-case analysis, comparing textual interpretations, their commonalities and differences across cases, to obtain more compelling evidence regarding the SRpL strategies found. Principles of the grounded theory approach were applied to identify SRpL strategies inductively and further build up the theory on SRpL. The first author analysed the longitudinal data independently. Evidently, data were coded and categorised, and constantly compared, thereby carefully scrutinising differences and similarities in the data. Tentative categories and potential differences in interpretations were critically discussed and assessed with the other authors in several peer debriefing sessions to increase the credibility of the findings (Graneheim & Lundman, 2004).

We engaged in a sequential intrapersonal investigation of each longitudinal dataset independently, as part of the within-case approach addressing the second research questions. For this purpose, we indicated with arrows all the relationships we identified between SRpL strategies used by the medical specialists. Figure 26.5 illustrates this by depicting a specific SRpL event across a two-day time span.

Several SRpL activities took place and connected with each other then. More specifically, within a selected learning event, we noted the physicians' apparent SRpL strategies, drawing arrows between them to indicate their relationship. We had arrows lead back to the SRpL strategies referred to more than once. Feedback loops originated in that way. The combinations of the earlier and newer SRpL strategies were framed, composing the ongoing process, as seen in Fig. 26.5. As such a so called SRpL trajectory across time emerged from the data, along with the dynamics of the process. We sought meaningful patterns for these trajectories within and across cases. We ensured the trustworthiness of the SRpL trajectories by further discussing the results in several peer debriefings (Graneheim & Lundman, 2004).

Based on these analyses of the data collected (Cuyvers et al., 2021), we found SRpL in the clinical environment to be shaped by a broad variety of SRpL strategies initiating, advancing, and evaluating the process of SRpL. Alongside the existing frameworks, we identified and classified as 'readiness strategies' the SRpL strategies without which SRpL could not take place. We also found SRpL to be an interrelated, dynamic process, unfolding over time with feedback loops among strategies and therefore constantly being adapted. Lastly, we concluded that physicians' work performance was a key driver of SRpL (Cuyvers et al., 2021).

26.4.3 Opportunities and Challenges of the Proposed Methodology

This case study design involves both on-line and (self-report) off-line data collection. As the next sections will show, the combination of these two techniques brought valuable opportunities for our empirical research, but also major challenges when applied to the case of SRpL of physicians (Cuyvers et al., 2021).

26.4.3.1 Opportunities and Challenges in Data Collection

One major opportunity the proposed interactive approach provided was a *comprehensive insight* on overt actions and behaviours as well as on covert thoughts and strategies. However, it was difficult to arrive at a *precise definition of the behaviours of interest*. Observations tend to be tied to judgments referring to the process of selection, filtering, discriminating, and sorting (Bratich, 2018; Gobo & Marciniak, 2011). Therefore, a clear and precise definition of the concept under investigation, informed by theoretical insights, is a prerequisite for making systematic and replicable observations (Bakeman, 2000). Since the existing literature has failed to clearly determine the aspects of observable behaviour that could point to SRpL, researchers must resort to collecting 'unstructured' data during fieldwork. Moreover, although *ecological validity* may constitute a great opportunity when investigating contextual factors along with covert strategy use and overt learning behaviours

(Wolters et al., 2011), not everything can be registered during observations in the field. Hence, using our protocol led to important choices in structuring data collection and ensuring the quality of the observations, as we chronicle below.

First, we studied both verbal and non-verbal communication. Our detailed recording of conversations and discussions via our tailored protocol allowed for the collection of trustworthy and credible data on how interactions could facilitate opportunities of SRpL. Second, our focus on a variety of accessible professional activities yielded rich descriptions of professionals' on-the-job performance and how it integrated SRpL. Third, our registration of time—to structure our field notes according to the course of the day—proved important to validly incorporating the transcribed stimulated recall interviews into the longitudinal dataset. Fourth, we found our registration of clues indicating prompts for the consecutive stimulated recall interviews, and providing references to the audio recordings, to be key to the reconstruction of the flow of reality in subsequent analysis. Lastly, although it lacked sophistication, our use of a clipboard, paper, and pencil proved to be well-suited for data collection which followed the pace of work, and for the successive transcriptions for analysis.

Despite the difficulties we faced in precisely defining the behaviours to look for, our *interactive approach* provided major opportunities for the required *elicitation of covert cognitive and metacognitive strategy use*. The verbalisation of the observed behaviours, interactions, and facial expressions helped the professional to recall and reveal thoughts and metacognitive processes that occurred at the time of the activity, as also indicated by Henderson and Tallman (2006). By re-engaging with the event, the subject provided clues enabling the acquisition of insights on non-observable strategy use. This also prompted conversations regarding events which the researcher would otherwise possibly not have witnessed. In relation to the *most appropriate questions* to elicit covert strategy use in everyday work, the stimulated recall protocol based on theoretical insights demonstrated the capture of SRpL. Thus, the reliability of the findings was enhanced (Yin, 2014).

The *interactive approach* proposed in this chapter revealed the possibility of measuring SRpL dynamically, by *taking time into account*. This offered a major advantage over cross-sectional measurements. However, prior research lacked parameters on *when and how long* to observe, *how many* observation moments there should be, and *how much time between* observations would be needed to capture SRpL as it unfolded in time. Moreover, given the concerns regarding the validity of retrospective self-report measurements (Veenman, 2007), we still must ponder the *best timing* to conduct stimulated recall interviews.

As mentioned earlier, important considerations guided our choices regarding the time-related aspects in our empirical study. We used on average four observation moments for each physicians, with an average duration of 4 h each. This duration was seen as sufficient time in the field and at the same time not too long, reducing the risk that researcher fatigue would influence the validity of the observation. We synchronised the variation in time slots with our study subjects' different professional activities and set a maximum five-day interval between two observation moments. This intermittent replication logic proved well-suited to ensuring more external validity and transferability (Yin, 2014; Morse, 2018). Regarding the timing for stimulated recall interviews, the literature has shown that retrospective reconstruction may lead to invalid interpretations of metacognitive skills, rather than correct recollections from memory (Veenman, 2007). To mitigate memory shortcomings, we used mostly non-directive prompts to elicit immediate consecutive recall, as close as possible to the actual event, which helped to ensure the reliability and validity of the data (Henderson & Tallman, 2006; Veenman, 2007).

Our research addresses an obvious plea for alternative valid measurements of SRpL through multi-method approaches (Cuyvers et al., 2020), such as in our combining on-line and off-line techniques with long-term observations and stimulated recall interviews. However, *obtrusiveness* of the researcher is a challenge when collecting data in the field (Angrosino & Rosenberg, 2011) and using verbalisation of observations as clues for stimulated recall interviews. Addressing this as an ethical issue (Angrosino & Rosenberg, 2011; Bratich, 2018; Roulet et al., 2017) requires the approval of the ethics committees and consent of all subjects concerned. Yet harder to overcome is the related validity challenge at the heart of our proposed interactive approach. As stated earlier, generating an 'ecologically valid' understanding of actual events appeared as a great opportunity. Not only did this approach discourage participants from painting an invalid 'ideal picture' when reality is being observed (as indicated by Furner & Steadman, 2004), but it also allowed for metacognitive strategies to be triggered remarkably close to the time of the actual event. However, such triggering (for instance in an interview situation) could also be viewed as an unintended intervention, which could push the degree of obtrusiveness and thus potentially influencing the validity of the findings (Roulet et al., 2017). Indeed, although inquiring into metacognitive strategy use during interviews is required in this approach, external regulation of the learning process could take place (via the interviewer asking questions), leading to intervention bias and perhaps even confirmation bias (Roulet et al., 2017). On the one hand, inviting professionals to think and speak about metacognitive strategy use during their practice (e.g. monitoring and reflection) could influence and perhaps lead to certain answers; on the other hand, long-term field observations enable researchers to ascertain whether a systematic tendency towards this externally-regulated metacognitive strategy use is at work.

Another related challenge recognised in the literature on methodologies (Ericsson & Simon, 1993; Henderson & Tallman, 2006) was subjects' potentially limited ability to *articulate* complex metacognitive strategies; in our study, they had to have the vocabulary to reflect, for example, metacognitive awareness and monitoring. Asking the right questions is key for this purpose, and paraphrasing subjects' answers could offer further solutions in this matter, but again, the latter could increase obtrusiveness and lead to bias. Another danger (Henderson & Tallman, 2006) is that habituation may supplant conscious strategy use, with only strategies in participants' consciousness being reported. Also, metacognitive strategy use could be absent, and inquiring into unconscious or absent strategies could cause bias.

The existing literature offered us no clear-cut solutions for these challenges. However, some suggestions can be made, which will be described in the next section.

26.4.3.2 Opportunities and Challenges in Data Analysis

The longitudinal case study design based on direct observations and stimulated recall interviews yielded *a lot of rich* empirical data. Such a rich dataset can leave researchers 'in a fog' for quite some time, unsure of what to *analyse* (Yin, 2018) or of the appropriate unit of analysis to choose. Evidently, this choice depends on the research questions posed. Investigating the concept of SRpL and its constituting self-regulatory strategies requires a grain size *unit of analysis*. When investigating the dynamic nature of SRpL, the unit of analysis is brought to the relations and sequences between self-regulatory strategies engaged in by professionals. The *selection of the learning events* related to professionals' engagement in SRpL can be another challenge in this sense, which calls for the researcher to address the question of what constitutes a learning event and how it can be identified. In sum, although the unit of analysis in research on SRpL is this concept occurring in close relation to work, its distinctive key characteristics designate this concept and lead to its different characteristics as relevant for the actual research question and necessary units of analysis.

A final challenge to be noted concerns the *interpretation of the SRpL data collected*, particularly in reference to the *distinction between strategies that regulate learning and those that regulate performance*. Careful consideration and rigorous empirical thinking, with reflexivity on the part of the researcher, are thus necessary, and a conscious sensitivity in a (repeated) cyclic analytic approach will help the researcher in this matter (Cresswell & Miller, 2000; Yin, 2018). We urge the researcher to always keep in mind during the analytic process, whether the data clearly demonstrates if or how the self-regulatory strategies observed explicitly served the purposes of work, learning, or both. Our engagement in peer-debriefing sessions proved critical to data interpretation convergence and to ascertaining the validity of findings (Cresswell & Miller, 2000; Lincoln & Guba, 1985).

26.4.4 Suggestions for Future Research

In relation to the opportunities and challenges mentioned above, we make further suggestions which may inform the decision-making process for future research on SRpL and/or professional learning.

One notable innovation in our design and method was the time aspect. The choices we made led to reliable findings on the components of SRpL in medical practice and on its dynamic evolution through time (Cuyvers et al., 2021). We reiterate here some important reflections and offer suggestions. That is, spending enough time in the field, with long-term observations is needed to access sufficient potential learning experiences that ensure more possible variation of behaviour and regulatory strategy-use. Moreover, intermittent observation moments are needed to allow for measurements of the dynamic aspects of SRpL. To define the appropriate time interval, insights on how and when change presumably reveals itself are necessary (Yin, 2014). Prior studies had been inconclusive regarding the ideal time frame for this: research on learning in the workplace relied upon visits of 1 or 2 days (Eraut, 2004; Furner & Steadman, 2004), while research on student teachers' SRL used six weekly learning reports (Endedijk et al., 2016). For future measurements of SRpL, we propose four observation moments, with an average duration of 4 h each, which proved to yield rich and credible findings in our empirical study (Cuyvers et al., 2021).

To address the challenge of researcher obtrusiveness, we first urge continuous critical scrutiny and reflection on the part of the researcher, as indicated by Guillemin and Gillam (2004), to assure the rigour and validity of the research. Second, we find it necessary for researchers to reflect on their own role in the observation context, making an explicit description of how this position could impact the data collection process (Bratich, 2018; Gobo & Marciniak, 2011). Researchers must take seriously the notion of a reflexive process in which their dynamics with subjects and the related contexts are an integral part of the research (Ezzy, 2013; Guillemin & Gillam, 2004; Gobo & Marciniak, 2011), to be critically reported on. Close and continual monitoring of the researcher's own interactions, reactions, roles, and biases, and related discussions with co-researchers, will support the objectivity of the research (Ezzy, 2013; Lincoln & Guba, 1985). Third, awareness for verbal and non-verbal signals in the participants' answers which might indicate bias, and alertness for signals regarding false interpretations, are necessary to ensure the validity and reliability of the data. Registration of, and reflection on, such signals should inform the analytic process.

Besides difficulties in interpreting SRpL data, we have noted challenges in identifying learning events during which professionals engage in SRpL. Indeed, the question, what counts for a learning event and how can they be identified, needs to be addressed by the researcher. Evidently, we first suggest gleaning professionals' indications of learning experiences from the existing literature (a.o. Cuyvers et al., 2016; Eraut, 2007; Eraut et al., 1998; Van Eekelen et al., 2005). Further, a clear definition of the situations that can account for learning events in which SRpL takes place is needed. To get a clearer picture of what constitutes SRpL during daily work, we recommend looking beyond the critical incidents professionals characterise as learning experiences. We suggest to define a 'learning event' as each on-the-job situation which a professional (1) explicitly relates to learning and development, (2) characterises as putting their competencies to the test, or (3) describes as an experience leading to the desire or need to perform better. We also suggest including in the definition situations in which others point out gaps in the professional's competencies, thereby offering a chance for SRpL. These clear parameters can play a key role in facilitating stimulated recall interviews; if researchers can identify situations as potential learning events based on subjects' verbal and non-verbal communication and cues, they can pose stimulated recall questions on the spot to elicit thoughts and potential SRpL strategy use.

26.5 Conclusion

It seems obvious that the de-contextualised, cross-sectional, self-report measurements often used to investigate professional learning and SRpL abate the complexity of on-the-job learning. We thus proposed a longitudinal multiple case study design as an alternative, using direct observations in everyday practice, along with stimulated recall interviews to acquire insights into metacognitive strategy use. Intermittent field observations offered us valuable cues to aurally prompt metacognitive strategy use during the stimulated recall interviews on the spot.

This interactive approach offers major opportunities to better understand the situated and dynamic nature of professional learning and the process of SRpL. However, bridging the gap between on-line and off-line techniques also comes with important challenges, such as the danger of researcher obtrusiveness and elicitation of certain metacognitive strategies. We chose medical practice as our case study (Cuyvers et al., 2021) to test the proposed approach to investigating SRpL. Research addressing other professional communities and environments could benefit from our findings and recommendations. Our approach could be used to research various aspects of professional development, especially when development over time should be taken into account, in authentic contexts and investigating observable and non-observable behaviours, thoughts, and affective, cognitive, and metacognitive processes. Examples would include the research topics of professionals' practical and generic skill development, identity formation, sustainment of occupational capacities, adaptability in ever-changing contexts, and personal, social, and contextual factors influencing developmental processes.

Furthermore, adding on-line, technology-supported data collection techniques to the proposed approach could greatly benefit the further exploration of SRpL, including many of the methods developed for research on SRL in educational settings (Schunk & Greene, 2018). 'Could', because contrary to educational settings, professional work environments are not designed for learning per se, which impacts these methods' applicability to workplace contexts. In terms of medical practice, another challenge is that in contrast with other work environments, it is highly socially interactive and dynamic, making it difficult for researchers to use particular on-line measurements. While video observations could facilitate observations during surgeries, during consultations, or situations where physicians run from here to there, they could be difficult to realise due to ethical concerns (e.g., obtaining patients' consent and honouring the confidentiality of medical information). Thus, due to the context-specificity of SRpL, these techniques clearly need further customisation, which could be done via experimentation with additional methods in the proposed design, in combining on-line and off-line techniques and in triangulating data sources to gauge these methods' ability to measure SRpL.

Overall, the proposed longitudinal case study design applied in Cuyvers et al. (2021) enabled us to disentangle the complex interdependence of work and learning. It represented a profound and systematic attempt to make implicit processes explicit and investigate SRpL contextually and in depth. The illustrated research approach yielded valuable research insights for future theory development in various fields of inquiry; it may become an important avenue for future research on workplace learning and professional learning and development. Future studies in the field of SRpL (or beyond) might therefore benefit from using this research approach, and from further developing the triangulation of on-line and off-line methods (such as direct observations and immediate stimulated recall interviews) to fine-tune the approach.

References

Angrosino, M., & Rosenberg, J. (2011). Observations on observations, continuities and challenges. In N. K. Denzin & Y. S. Lincoln (Eds), *The SAGE handbook of qualitative research* (4th ed., pp. 467–478). Sage publications, Inc.

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*(4), 210–223.

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254–270). Routledge.

Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138–159). Cambridge University Press.

Bernacki, M. L. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: The promise of in-depth case study methodologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 370–387). Routledge.

Bransen, D., Govaerts, M. J. B., Sluijsmans, D. M. A., & Driessen, E. W. (2020). Beyond the self: The role of co-regulation in medical students' self-regulated learning. *Medical Education, 54*(3), 234–241.

Bratich, J. (2018). Observations in a surveilled world. In N. K. Denzin & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (5th ed., pp. 526–545). SAGE.

Brydges, R., Tran, J., Goffi, A., Lee, C., Miller, D., & Mylopoulos, M. (2020). Resident learning trajectories in the workplace: A self-regulated learning analysis. *Medical Education, 54*, 1120–1128. https://doi.org/10.1111/medu.14288

Cresswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Cuyvers, K. (2019). *Unravelling medical specialists' self-regulated learning in the clinical environment*. Academic dissertation. University of Antwerp.

Cuyvers, K., Donche, V., & Van den Bossche, P. (2016). Learning beyond graduation: Exploring newly qualified medical specialists' entrance into daily practice from a learning perspective. *Advances in Health Sciences Education, 21*(2), 439–453.

Cuyvers, K., Van den Bossche, P., & Donche, V. (2020). Self-regulation of professional learning in the workplace: A state of the art and future perspectives. *Vocations and Learning, 13*, 281–312. https://doi.org/10.1007/s12186-019-09236-x

Cuyvers, K., Donche, V., & Van den Bossche, P. (2021). Unravelling the process of self-regulated learning of medical specialists in the clinical environment. *Journal of Workplace Learning, 33*(5), 375–400. https://doi.org/10.1108/JWL-09-2020-0151

Ellström, P.-E. (2001). Integrating learning and work: Problems and prospects. *Human Resource Development Quarterly, 12*(4), 421–435. https://doi.org/10.1002/hrdq.1006

Endedijk, M. D., & Cuyvers, K. (2021). Self-regulation of professional learning: Towards a new era of research. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning: Arising perspectives and retrospective*. Springer.

Endedijk, M. D., Brekelmans, M., Sleegers, P., & Vermunt, J. D. (2016). Measuring students' self-regulated learning in professional education: Bridging the gap between event and aptitude measurements. *Quality & Quantity, 50*, 2141–2164. https://doi.org/10.1007/s11135-015-0255-4

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 70*, 113–136.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 247–273.

Eraut, M. (2007). Learning from other people in the workplace. *Oxford Review of Education, 33*(4), 403–422. https://doi.org/10.1080/03054980701425706

Eraut, M., Alderton, J., Cole, G., & Senker, P. (1998). Learning from other people at work. In F. Coffield (Ed.), *Learning at work* (pp. 37–48). Policy Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data*. Massachusetts Institute of Technology.

Ezzy, D. (2013). *Qualitative analysis: Practice and innovation*. Routledge.

Furner, J., & Steadman, S. (2004). Eliciting professional practical knowledge: Experiences and problems. *Conference paper, British Educational Research Association Annual Conference*, University of Manchester, 16–18 September. Assessed 2 Mar 2018 from http://www.leeds.ac.uk/educol/documents/00003884.htm

Gobo, G., & Marciniak, L. T. (2011). What is ethnography? In D. Silverman (Ed.), *Qualitative research 4E* (3rd ed., pp. 103–120). SAGE.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*, 105–112. https://doi.org/10.1016/j.nedt.2003.10.001

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry, 10*(2), 261–280.

Henderson, L., & Tallman, J. (2006). *Stimulated recall and mental models, tools for teaching and learning computer information literacy*. The Sacrecrow Press.

Jossberger, H. (2022). Eye tracking in professional learning and development: Uncovering expertise development among residents in radiology. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 467–489). Springer.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Littlejohn, A., Milligan, C., Fontana, P. R., & Margaryan, A. (2016). Professional learning through everyday work: How finance professionals self-regulate their learning. *Vocations and Learning, 9*, 207–226. https://doi.org/10.1007/s12186-015-9144-1

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behavior research. *Journal of Occupational and Organizational Psychology, 65*(2), 131–149.

Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (5th ed., pp. 796–817). SAGE.

Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence Based Nursing, 22*(3), 67–68. https://doi.org/10.1136/ebnurs-2019-103145

Rausch, A. (2014). Using diaries in research on work and learning. In C. Harteis, A. Rausch, & J. Seifried (Eds.), *Discourses on professional learning: On the boundary between learning and working* (pp. 341–366). Springer. https://doi.org/10.1007/978-94-007-7012-6_17

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.

Roulet, T. J., Gill, M. J., Stenger, S., & Gill, D. J. (2017). Reconsidering the value of covert research: The role of ambiguous consent in participant observation. *Organizational Research Methods, 20*(3), 487–517.

Schunk, D. H., & Greene, J. A. (2018). Historical, contemporary, and future perspectives on self-regulated learning and performance. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 1–17). Routledge.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421–442.

Spliethoff, L., & Abele, S. (2022). Measuring professional competence using computer-generated log data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 165–186). Springer.

Stake, R. E. (1995). *The art of case study research*. SAGE.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154.

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and Learning, 6*, 11–36. https://doi.org/10.1007/s12186-012-9091-z

Van Eekelen, I. M., Boshuizen, H. P. A., & Vermunt, J. D. (2005). Self-regulation in higher education teacher learning. *Higher Education, 50*, 447–471. https://doi.org/10.1007/s10734-004-6362-0

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs. In C. Artelt & B. Moschner (Eds.), *Lernstrategien und metacognition, implikationen für forschung und praxis* (pp. 77–100). Waxmann Verlag GmbH.

Veenman, M. V. J. (2007). Learning to self-monitor and self-regulate. In P. Alexander & R. E. Mayer (Eds.), *Handbook of research on learning and instruction, educational psychology handbook series* (pp. 197–218). Routledge.

Veenman, M. V. J. (2011). Alternative assessment of strategy-use with self-report instruments: A discussion. *Metacognition and Learning, 6*, 205–211.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14.

Vrieling-Teunter, E., Stijnen, S., & Bastiaens, T. (2021). Promoting student teachers' self-regulated learning in the workplace. *Vocations and Learning, 14*, 223–242. https://doi.org/10.1007/s12186-021-09264-6

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Wolters, C. A., & Won, S. (2018). Validity and the use of self-report questionnaires to assess self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 307–322). Routledge.

Wolters, C. A., Benzon, M. B., & Arroyo-Giner, C. (2011). Assessing strategies for the self-regulation of motivation. In B. J. Zimmerman & D. A. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 298–312). Routledge.

Yin, R. K. (1999). Enhancing the quality of case studies in health services research. *Health Services Research, 34*(5), 1209–1224.

Yin, R. K. (2014). *Case study research: Design and methods*. SAGE.

Yin, R. K. (2018). *Case study research and applications, design and methods* (6th ed.). SAGE.

Katrien Cuyvers (PhD) is guest professor in Health Science Education and Learning in Organisations at the University of Antwerp (Faculty of Social Sciences, Department of Training and Education Sciences and Antwerp School of Education). She is also a post-doctoral researcher ('Learn@work') at the Spaarne Gasthuis for the University of Twente, the Netherlands. Her research activities relate to self-regulated workplace learning and co-regulation of learning in the workplace, in particular in healthcare environments.

Piet Van den Bossche (PhD) is Professor of Learning in Organizations at the University of Antwerp (Faculty of Social Sciences) and Professor of Team Learning at Maastricht University (School of Business & Economics). He is co-director of the Antwerp Social Lab. His research activities are centered on learning, cognition and performance in teams and collaborative work.

Vincent Donche (PhD) is a full professor at the Department of Training and Education Sciences of the Faculty of Social Sciences at the University of Antwerp, Belgium. Within the research group EduBROn, he conducts research in the domain of learning and instruction, higher education, and educational measurement.

Part IV
Discussion & Commentary

Chapter 27
Researching Professional Learning in Complex Environments: Opportunities and Challenges from a Qualitative Research Perspective

Monika Nerland

Abstract This chapter discusses the contributions in the edited volume *Methods for Researching Professional Learning and Development: Challenges, Applications and Empirical Illustrations* from a primarily qualitative research perspective. The discussion takes the different methodological challenges or shortcomings that the chapters aim at addressing as a starting point, and moves between the suggested methods to discuss opportunities and limitations. A recurrent challenge is to find adequate ways to approach and delimit work practices within which learning unfolds, especially when work becomes multi-sited, interlinked in network constellations of actors and practices, and possibly black-boxed through the use of various information systems. In these efforts, a variety of theoretical and methodological approaches is needed. This edited volume is a valuable contribution to methodological discussion and awareness, which, through its various chapters, provides condensed introductions to various approaches as well as examples of how the given approach has been used in studies of professional learning or development. A suggested way forward is to establish more strategic connections to other research fields that take an interest in work-related learning and development, such as organisation studies, information system research and infrastructure studies.

Keywords Professional learning · Professional practice · Professional expertise · Digitalisation · Transformation of work · Qualitative research

M. Nerland (✉)
Department of Education, University of Oslo, Oslo, Norway
e-mail: m.b.nerland@iped.uio.no

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_27

27.1 Introduction

This book emerges from the challenges faced today by research on professional learning and development (PLD), and more specifically from an envisioned need to broaden the methodological repertoire and keep up with recent methodological opportunities that are yet to be adopted in a more widespread way in this field of research. The aim is, therefore, to provide an overview of less-established and emerging methods that are relevant to research on professional learning. The value of such a collaborative initiative is at least threefold: First, any research field should aspire to keep up with methodological innovations in its wider research context, in this case in the social sciences and in educational research more broadly. Second, when the focus is on professional learning and development, there is a need to adapt research topics and methodological approaches to the ongoing changes in the organisation of professional work. It is generally acknowledged that important professional learning happens through participation in work practices. How work environments offer opportunities for learning at work is, therefore, of great importance, and research needs to keep up with important changes in these environments. Third, any research field would need to monitor the development of the field and facilitate internal knowledge-sharing for further advancement. Initiatives such as the volume presented here are a welcome opportunity to see the field of PLD research from a methodological point of view and to consider how it develops from within and relative to its neighbouring research areas.

Methodological innovations have different origins. One driver for change is the new opportunities for data gathering and analysis that come with digitalisation. This includes opportunities to make productive use of larger datasets and automatically generated data flows, as well as data generated from the use of advanced tools and instruments in work practices. Further, methodological advancements can be driven by changes in the phenomenon under examination—in our case, PLD—that call for a rethinking of the unit of analysis and approaches to collecting and analysing data. For instance, new ways of organising work and employment may call for approaches that better grasp phenomena such as multi-sited work practices, or the emergence of self-employed careers in the platform economy. Third, methodological innovations may develop cumulatively through the refinement and perfection of long-existing approaches. Here, advancements can be driven by a need to strengthen the transparency of the approach to enhance the credibility, or by efforts to improve its ecological validity. A fourth driver of change relates to the shifting notions of research and its contribution to society. For instance, the current emphasis on user involvement in research and the growing orientation towards analytics in organisational life may both, yet in different ways, encourage interventionist methodologies and soften the boundaries between research activities and other spheres of professional life.

The chapters in this volume relate to these driving forces in various ways, and take different challenges or unexploited opportunities as their starting point.

Although the current chapter primarily discusses the contributions from a qualitative methodological point of view, it becomes clear through several chapters that methodological advancements in part seek to overcome this division. The organisation of the volume in its three parts highlights methods for data collection, data analysis, and broader research approaches, which in itself is an interesting manifestation of differences in methodological concerns. Part I addresses methods for data collection, comprising chapters on both qualitative and quantitative approaches that seek to explore data sources that can complement or expand our understanding of individuals' learning or experiences. Part II, addressing methods for data analysis, is largely focused on quantitative approaches. Part III presents several qualitative approaches that typically include a broader set of organisational processes and relations in their units of analysis. Some chapters in this section also address interventionist and participatory research designs.

The positioning of contributions in the different sections of the book may reflect dominant concerns in the two main approaches. Qualitative research is often characterised by taking social processes and experiences, and how these are created and given meaning, as their objects of enquiry. Denzin and Lincoln (2018, p. 10) described qualitative research as "a set of interpretative, material practices that make the world visible", highlighting how researchers interpret the social world by enacting a set of material practices to represent, explore, and make sense of social phenomena. Qualitative research can be directed towards processes as they unfold, towards personal experiences, and towards characteristics of specific social environments. However, the focus of the analysis will often include relations between dynamic processes and entities as they unfold in situ or are experienced by practitioners. Hence, the ecological validity of the research and whether the research design offers an adequate reduction of the social world are recurrent concerns. The predominance of quantitative approaches in Part II of the book may reflect the relatively stronger emphasis on techniques for achieving reliable measures and enabling replicability through these procedures. This may also reflect the fact that qualitative approaches will less easily separate between data collection and analysis as distinct processes, as these often become intertwined in ongoing explorative and interpretative work.

In the next section, I will discuss the contributions in this book from a primarily qualitative research perspective. However, recognising that research on PLD often makes use of several data types and procedures, I will also include some comments on initiatives that combine qualitative and quantitative approaches. I will take the different challenges or shortcomings that the chapters aim at addressing as a starting point and move between the suggested methods to discuss opportunities and limitations. In Sect. 27.3, I relate the discussion to changes in professional work that in different ways pose new challenges to research on PLD, before the chapter concludes with some reflections about future developments in this field of research.

27.2 The Contributions: Problems Addressed, Insights Generated

A common interest in several of the chapters is to better grasp manifestations of PLD in the situation where they occur, and to account for contextual features that shape learning and experiences in situ. The problem of relying on self-reported data in retrospect when making claims about learning situations and processes has been explicitly raised by several authors (Cuyvers et al., 2022, Chap. 26; David et al., 2022, Chap. 9; Lemmetty et al., 2022, Chap. 18; Rausch et al., 2022, Chap. 3; Paloniemi et al., 2022, Chap. 5). Finding ways to generate rich data about PLD processes as they unfold is thus an important methodological challenge. Qualitative approaches are, in principle, well suited to meet these demands. The methodological toolkit offers opportunities for recording actions and interactions, logging experiences, or narrating accounts as they emerge, as well as approaches to analysing processes of work and learning in their sociomaterial context. The problem is that such forms of data generation are often time consuming and may disturb the work practices themselves by asking participants to generate accounts as they go, or by having one or more researchers present in a participatory observational mode. The chapters examine and discuss several alternatives for addressing these challenges.

27.2.1 Studying PLD Through the Practices and Experiences of Individual Professionals

Two of the chapters present different ways of using diaries as an approach to collect and analyse professionals' experiences related to their work environments. The chapter by Rausch et al. (2022, Chap. 3) suggests that diaries can be useful for uncovering the informal and often unconscious aspects of learning at work, illustrated by studies that prompt workers to report during work by way of paper-based forms or voice recorders. The Rausch et al. show how diaries can take a range of forms and purposes and advise the use of short scales and single items to enhance efficient registration. One question in this regard is the extent to which the format of reporting can or should be predefined by the researchers. Events and experiences can, as this chapter illustrates, be logged to gain information about their occurrence and frequency, and how this varies across groups or situation types. However, if the research interest is how informal learning processes emerge, it may be difficult for the researcher to predefine situations and categories for reporting such information—and especially so in complex work environments, which I will discuss later. Leaving participants to decide what events are significant and worth reporting may be an alternative, however at the cost of comparative opportunities. The chapter by Vähäsantanen and Arvaja (2022, Chap. 17) also discusses the use of diaries; however, in this case, it is combined with interviews to perform narrative analyses of professionals' identity construction over time. The use of diaries (and interviews) to gather narratives has the potential to generate in-depth analyses of how professionals understand themselves and how this might shift over time and with different work environments. As Vähäsantanen and Arvaja argue, such self-understanding matters for learning and agency. Especially in transition phases that involve shifts in work positions and identities, such a narrative approach is suitable. An in-depth account of learning situations and processes as they occur at work would need a rich description of the relationships between tasks, work processes, and work environments, which may or may not be sufficiently captured in the participant-generated narratives.

Another related interest in several chapters is to expand the analytical focus in research on learning experiences to include emotional and embodied dimensions, in addition to the often foregrounded cognitive dimension. Two chapters explore the opportunities for combining more traditional qualitative data sources with the use of new technologies for measuring brain activity and its neural basis (Paloniemi et al., 2022, Chap. 5; Silvennoinen et al., 2022, Chap. 7). These chapters clearly present emerging opportunities for researching aspects of the human learning experience by integrating methodological tools and technologies from neurosciences. However, the methods have limitations regarding researching PLD as an emergent, processual phenomenon, especially concerning learning as intrinsic to work practice. For instance, although measures from electromodal activity may help identify emotions that are in play in specific experiential or stimulated situations, it is not yet clear how these emotions matter for learning processes that unfold over time.

Cuyvers et al. (2022, Chap. 26) take interest in PLD as an emergent and processual phenomenon that is "entirely integrated with work" (p. XX). They specifically focus on how professionals' self-regulated learning in work situations can be studied. Cuyvers et al. (2022, Chap. 26) propose what they term a longitudinal, multiple case study approach, in which the individual professionals serve as cases, and data are collected through direct observations of professionals in work situations combined with immediate stimulated recall interviews. By using a 'shadowing' technique to follow individual workers in and across various work situations, researchers can observe how PLD unfolds in different events as a work-integrated activity, while the professionals' intentions and strategies for self-regulation are further explored in the stimulated recall interviews. This approach is productive in accounting for the situated actions as they occur in everyday work situations and at the same time generating data on metacognitive processes. It can be used for different research interests; for instance, quite a few studies in the field of organisation studies have combined the shadowing of persons with forms of immediate interviews or stimulated accounts to examine the work practices of specific organisational actors. We could imagine other research topics in PLD research as well, for instance, how professionals (learn to) use new technologies in their work environments, or how they develop strategies for working with clients or user groups. One question, however, is what qualifies research designs to be longitudinal. In the study presented by Cuyvers et al. (2022, Chap. 26), the practitioners ($N = 13$) were observed 4×4 h on average, distributed on different working days. Although the approach certainly generated a rich and voluminous data set, it seems to provide a better basis for investigating variations in the self-regulation of professional learning across work settings than for following developments in their strategy use over time.

27.2.2 *Accounting for Interactional Processes and Organisational Relations*

While the above-discussed chapters take individual professional learners as their focal point, some contributions take interest in learning as a collective and interactional accomplishment, hence focusing on the dynamic interaction in teams or groups as their unit of analysis. David et al. (2022, Chap. 9) focus on the temporal dimension of team learning by investigating the interaction dynamics of teams understood as changing patterns of micro-behaviours emerging in teamwork. The authors discuss how micro-behaviours can be understood and coded on different levels of temporal resolution, subjected to different (also quantitative) analysis techniques. An interesting question is how micro-behaviours and their role in the dynamics of interaction become re-represented through coding schemes. The proposed techniques will offer different opportunities in this regard. From a qualitative research perspective, a common distinction is made between ‘categorisation strategies’ and ‘connecting strategies’ in the analysis of social phenomena (Maxwell & Chmiel, 2014).

The first relates to strategies for examining similarity-based relations by comparing and contrasting, whereas the second examines contiguity-based relations by analysing the connections between elements and how they work on each other. In other words, what is split up through the coding processes and what is kept together as meaningful units of data is a critical issue. Here, it is interesting to consider the chapter by Filliettaz et al. (2022, Chap. 19) as a different approach to analysing social interaction. These authors propose a video-based interaction analysis that builds on principles from micro-sociological approaches to studying communication in work situations. These methods have a long history in workplace studies, where the ongoing coordination, negotiation, and accomplishment of joint work processes have been a main interest. In the chapter by Filliettaz et al. these methods are discussed in the context of adults’ learning at work. As a first phase, the researchers conduct a video-ethnographic enquiry through which selected and significant work situations are recorded. Next, these recordings are used within the frames of an interventionist design, where video recordings of participants’ work situations serve as a means for analysing the interaction together, which in turn stimulates the professionals’ awareness and learning of interactional competencies. In this way, recorded work situations are used in formal training sessions with PLD as a stated aim, and the methods serve both research- and practice-related purposes.

One critical question to this approach, which is also discussed in the chapter, is how and to what extent the participating professionals can adopt the (relatively advanced) principles of performing interaction analysis together with the researchers. Such analyses often require extensive training and theoretical understanding. Therefore, one may consider the extent to which the participants and the researchers can conduct a joint analysis, and the extent to which the granularity of the analysis should be distinguished for the purposes of supporting participants' PLD and developing the researchers' contributions to the field of research. However, the collaboration with researchers and the procedural steps such an analysis implies may provide the necessary distance for the participants to consider the recorded situations anew, and the relevance of the analysis can be secured on a continuous basis.

Whereas the chapter by Filliettaz et al. (2022, Chap. 19) refers to micro-sociology and ethnography of communication, the chapter by Lemmetty et al. (2022, Chap. 18) discusses a broader ethnographic approach in which the analytical interest is expanded to account for wider sets of activities and processes in the workplace. The authors build on organisational ethnography as a research approach and discuss how its principles and methods can be employed to study PLD as a culturally embedded and practice-based phenomenon. The overall aim is to gain a comprehensive understanding of PLD, as it is contextually framed and situated. Recognising that PLD manifests in different organisational layers and contexts, and that individuals' learning and development is still a key interest, the authors also discuss more recent developments of the use of ethnographic approaches in virtual environments as well as the emergence of what is called 'subjective evidence-based ethnography' (SEBE). Although the latter is an interesting take on examining what situations and social positioning literally look like from the perspective of participants by placing cameras at the eye level of participants' heads, it may break with some of the long-established principles of ethnography regarding including the wider practice settings and organisational environments where these recorded situations are embedded. A critical question is what counts as context when PLD is seen as contextually framed, and the context is a means to establish analytical accounts between individual actions, communities, and organisations. Nevertheless, the SEBE approach may be well suited to examine work practices on the move, as an alternative strategy to the previously mentioned shadowing technique, and it may provide close views of working tasks that are not easily observed through participant observation, such as screen-mediated work on computers or hand-held devices.

27.2.3 Fostering and Responding to Change in Education and Work Environments

Some chapters discuss approaches for interventionist studies in which researchers and professional practitioners collaborate to generate new insights and change the status quo of professional practices. In addition to the above-described chapter by Filliettaz et al. (2022, Chap. 19), this concerns the chapters on design-based research (DBR; Gerholz & Wagner, 2022, Chap. 23) and the change laboratory

(CL) approach (Kajamaa & Hyrkkö, 2022, Chap. 24). Both approaches are participatory in their character and aim to foster change in the environment where they are employed. Both have a long history in the learning sciences and within studies of work and organisational learning; however, they have possibly not yet been widely adopted in the community of PLD researchers addressed in this edited volume. DBR focuses on the design of learning environments to support specific learning processes or overcome challenges experienced in learning. Following Gerholz and Wagner (2022, Chap. 23), the approach aims to bridge the gap between rigour and relevance in educational research by developing practical designs while studying the processes and implications of the design to generate theoretical insights. The research process is iterative, moving between design, implementation, evaluation, and redesign as recurrent processes. Gerholz and Wagner (2022, Chap. 23) present a study that takes digitalisation and new competence needs in the field of VET as a point of departure, and the design object is lesson designs with tablets as tools to be incorporated in vocational education. Both students and teachers in the educational setting become participants. However, as the authors comment, the teachers' professional development was dependent on support from researchers in the project over a longer period of time. Kajamaa and Hyrkkö (2022, Chap. 24) present and discuss the CL approach as a means to facilitate teachers' collective transformative agency. This approach is grounded in cultural-historical activity theory and the theory of expansive learning, and the intervention takes place through a series of laboratory sessions in which researchers support practitioners in articulating challenges and revealing contradictions in their current work environments as a means to redesign their work practices by way of creating new ideas and tools. The CL approach is cyclic in character; however, in this case, the overall aim is to foster organisational change and expansive learning by developing the participants' transformative agency. Both the DBR and CL approaches are well suited to support collective professional learning and development in work communities through interventionist methods in which the researchers take the role of facilitators for change. These approaches might gain an increased foothold at a time when researchers are expected to engage more profoundly with users and stakeholders. A potential challenge with these approaches is generating theoretical and empirical contributions beyond the case, especially in projects where the main envisioned product of the research is located in the practices and understandings of the participants.

In addition to interventionist approaches, the edited volume comprises chapters that take other problems of change as their points of departure. Harteis (2022, Chap. 16) presents and discusses the use of the Delphi technique to investigate future visions and scenarios in a given area of work as well as how these visions might matter for professionals' orientations and development. Designated experts were invited to provide their prospective accounts in a written manner, which in turn were narrowed down, analysed, and validated through the use of different methods, including qualitative approaches. The Delphi technique is increasingly used in educational research and has its strengths in its potential to engage with emerging ideas and scenarios that have not yet materialised. Critical questions concerning this approach include what kind of qualitative analysis can be used for what purposes and related to this is the role taken by the interpreter. An important principle in many approaches to qualitative analysis is to distinguish between the participants' own terminology and categorisation of their experiences or accounts, and the (often theory-based) language of the researcher. Even in explorative studies that aim at generating visions of novel phenomena, the analysts will have to work through language; therefore, it seems important to clarify their theoretical assumptions. Another question might be the extent to which it is important to reach agreement and consensus between the participants' accounts, or whether identifying variation could be considered asset in explorative and future-oriented studies.

Lastly, some chapters aim to address more comprehensive problems related to PLD by taking advantage of larger datasets and (automatically generated) data flows. Kyndt and Aerts (2022, Chap. 15) explore the possibilities for bringing big datasets into human interpretation by way of visual analysis as a way to better grasp the complexity of professional learning as a "wicked problem". Littlejohn et al. (2022, Chap. 25) discuss new methodological opportunities for learning analytics, which are still in their early phase of development and have hitherto been employed in formal educational settings rather than in work settings. Although they differ in their methodological challenges and approaches, both chapters show how the unpredictability and complexity of professional learning at work generate challenges for the research procedures—in terms of generating data for learning analytics (Littlejohn et al.) and keeping the 'wickedness' of problems while adhering to established principles for reliable research (Kyndt and Aerts). For instance, although Kyndt and Aerts depart from a system-theoretical understanding of professional learning as wicked problems, one may ask whether the suggested methodological attention towards hypothesis testing and emphasis on replicability would imply that the problems are again 'tamed' by the researchers. Where the use of learning analytics is concerned, an interesting avenue for further research might be to explore whether and how PLD researchers can make use of 'work analytics'—that is, data generated through the use of work support systems and digital platforms for work coordination—in their analysis of professional learning as intrinsic to work.

27.2.4 Summary

These chapters have discussed a wide range of methodological approaches and opportunities for researching PLD. One insight generated from the various chapters discussed above is that researchers interested in PLD need to carefully consider their unit of analysis, and how the phenomenon under investigation is framed in space and time. This may be even more important in qualitative research, as researchers often aim to examine social processes or relations in their relevant contexts. Reducing the social phenomenon is unavoidable to make it researchable, and the complexity of relations and processes involved in PLD can obviously never be grasped in one study (Ludvigsen & Nerland, 2018; see also the Special Issue of *Learning, Culture* *and Social Interaction*, 2021, Vol. 31, Part B, devoted to ‘The unit of analysis in learning research’). Some researchers delimit their research focus to a restricted timeframe and attempt to include a broader set of factors or relations in their analysis, for instance, by analysing the different dimensions that constitute a learning experience in a given situation. Others pursue processes over a longer time span but specify the analytical focus on particular aspects of professional learning, such as self-regulation. Yet others take a specific data type or opportunity for data generation as a starting point and examine their methodological opportunities. A range of complementary approaches is needed for the research field to provide comprehensive insights into PLD and the social and organisational arrangements that can support it. For individual studies, however, defining the unit of analysis and securing its theoretical framing from a related perspective on PLD is significant to clarify how and the limitations within which the study contributes to the wider research agenda.

As pointed out in quite a few chapters, professional learning at work is closely related to the enactment of work practices. This means that we need to understand the practices of work and how these are organised to understand professional learning as intrinsic to work. Given the overall theme of the volume, most chapters in this book start out from methodological problems and experienced challenges, rather than from the work itself. I will therefore in the next section reflect on how professional work environments are growing in complexity, and what we then need to account for in research on work-based PLD.

27.3 The Complexity of Work Environments: Emerging Challenges to Researching PLD

Professional work organisations and practices are undergoing profound changes that matter for how learning demands and opportunities manifest. Several chapters in this volume argued for a need to study PLD processes in the work situations where they occur, and account for the contextual features that shape learning processes and experiences (see Sect. 27.2 above). When work processes are physically distributed and expanded in time, new interdependencies emerge between participants in various practice settings. Together with the presence of advanced tools and infrastructures, it becomes more challenging to analytically define work contexts and decide on which elements to include in the unit of analysis. In this section, I will return to the second methodological change driver mentioned in the introduction to this chapter and discuss how changes in the organisation of professional work itself bring forth new questions and elements of practice that need attention in research on PLD.

In particular, more intensified digitalisation processes play a key role in transforming work and bringing new interdependencies and demands to expertise to the fore. The impact of digitalisation processes on organisational life is widely theorised and researched in the field of organisation studies and sociology of work organisations (e.g., Anthony, 2021; Bailey et al., 2022; Pachidi et al., 2020), and increasingly so by scholars in professional and workplace learning (e.g., Harteis, 2022; Littlejohn & Pammer-Schindler, 2022). This literature highlights, first, how new and more advanced technologies are at play in professional work practices, for instance, how artificial intelligence and algorithm-based support are incorporated in tools and information systems that alter the conditions for decision-making (Anthony, 2021). This poses challenges to experienced professionals, who need to learn how to trust and interrogate automatically generated suggestions. It also challenges the way newcomers can learn and develop expertise, as important aspects of expert practice may remain black-boxed and difficult to observe for novices. How work can be organised to better support these learning processes is therefore an important question.

Second, researchers have examined how new working tasks come to the fore and have expanded the scope of the responsibilities professionals face at work. Some of these responsibilities are related to future-oriented efforts in which professionals actively engage in reconstructing work practices to accommodate wider processes of digital transformation. Other responsibilities concern the need to understand the wider set of organisational processes and interdependencies in which their work is located. One example is provided by Herzum and Simonsen (2019), who examined work in a hospital setting and identified a set of competencies that are needed in the local work environment to configure an information system for productive use. They argue that 'understanding practice' emerges as a larger competence issue than 'understanding the technology' and that professionals increasingly need to understand what others are doing in the chain of actions that make up the services.

Related to the latter issue, the transformation of professional work is also driven by a growing complexity in the diversity of actors and concerns. Grounded in the sociology of professions, Noordregraaf (2016, 2020) discusses how professionals need to relate actively to stakeholders and actors outside their professional realm to ensure that their expertise is sustained and entrusted. This implies that professionals take on responsibilities for the way work is organised to accommodate various concerns and for sorting relations and connecting with stakeholders in productive ways. In the literature on professional learning, some researchers have started to examine what it entails to engage with a wider set of actor constellations in the ongoing realisation of knowledge work (e.g., Tronsmo, 2020). A recognised yet growing research interest is the way professionals learn to collaborate across domains of expertise. Markauskaite and Goodyear (2016) argue that professionals today often work on tasks that depend on other professions and "operate in circumstances where their own professional knowledge is insufficient for success and their own professional practices have to adapt to the practices of others" (p. 25). They suggested the concept of epistemic fluency to analyse how professionals can learn to integrate different forms of knowing and adapt this to different work contexts. Moreover, to better understand how theoretical knowledge can become productive in complex practical situations, they suggest focusing on the construction of actionable conceptualisations as a unit of analysis and employing a multimodal blending perspective as an analytical approach (Markauskaite et al., 2021). This research strategy would require in-depth studies of constructive activities in situ, and approaches to data collection and analysis that account for different types of object construction as well as representations of knowledge and forms of knowing. Another contribution to researching work and learning in interprofessional collaboration is provided by Edwards (2010), who develops a perspective on relational expertise and employs analytical resources from cultural-historical activity theory to examine the practices through which professionals build relational agency and work together to overcome knowledge boundaries. Here, the established and envisioned practices of the professionals and how these are negotiated in the organisational context are foregrounded.

Together, these strands of literature underscore how the work environments of professionals are increasing in complexity in ways that call for continuous learning and development. Learning to take part as skilful practitioners is not restricted to learning how established work procedures and ways of knowing can be accomplished. Rather, professional work implies taking part in knowledge-generating activities and in service redesign, in which digital tools and infrastructures play prominent roles (Nerland & Hasu, 2021). Increased attention has, therefore, been given to how professional education can prepare students for these activities (Damşa & Nerland, 2016; Markauskaite & Goodyear, 2016). Moreover, through the way technologies and practices become linked in wider infrastructures, work processes become distributed over a wider scale and are constituted through multiple interdependencies between actors and sites (Nicolini et al., 2018). These and other transformations of work enlarge the scope of relations and actors that need to be accounted for in research on PLD. On the one hand, there is a need to expand the unit of analysis to include how the learning-through-practice that emerges in one setting is interlinked with practices and relations elsewhere in the organisation. On the other hand, researchers may need to go 'deeper' into the actual work practice and examine how its implicit epistemological underpinnings—for instance, an algorithm-based support mechanism or the view of a non-present stakeholder—inform ways of knowing in the given situation. Further, the temporal scope of the research is an issue. As more work processes are oriented towards changing and developing the ways of working, and professionals move in-between various sites, roles, and responsibilities, research designs that follow workers and work processes over a longer time span would also be of great interest from a PLD perspective.

In these efforts, a variegated set of research projects and practices that draw on different theoretical and methodological approaches are needed. In particular, a way forward could be to establish more strategic connections to other research fields that take an interest in work-related learning and development. In fields such as organisation studies, information system research, and infrastructure studies, there is a growing interest in learning and competence development as essential for work performance. Here, researchers from the PLD community would bring important insights and analytical approaches to the table. Further, the analyses of PLD could draw on expertise from these domains to better account for organising processes and the complexity of the IT systems with which professionals interact. In such collaborative efforts, it is important to expand the scope of the analysis without losing the

PLD researchers' identity as affiliated with their research field. Hence, as the complexity of work organisations moves into the analytical foci, a joint responsibility should be to keep the phenomenon of PLD discernible and secure so that the research contributions add to our knowledge about human learning and development in shifting professional, social, and institutional contexts.

27.4 Conclusion

In qualitative research on PLD, a recurrent challenge is to find adequate ways to approach and delimit work practices within which learning unfolds, especially when work becomes multi-sited, interlinked in network constellations of actors and practices, and possibly black-boxed through the use of various information systems. Efforts to combine data sources, account for interdependencies, and make implicit aspects of work practices accessible for analysis are examples that push qualitative approaches in professional learning research to include more relational and organisational dimensions. Further, other changes in the work organisation can stimulate a renewed interest in the experiences and ambitions of the learning subject. For instance, the increase in portfolio careers and the emergence of platform-mediated and self-employed work may call for approaches to better grasp learners' agency and strategies for working within dynamic and changing organisational environments.

The current book is a valuable contribution to methodological discussion and awareness, which, through its various chapters, provides condensed introductions to various approaches as well as examples of how the given method or approach has been used in studies of professional learning or development. Moreover, the volume is useful for research students in their efforts to explore and consider the opportunities that different methodological approaches provide. Today's publication regimes emphasise journal contributions, in which the limited length of the manuscripts may imply that the methodological choices and steps taken in the analysis are less transparent, leaving these important parts of the research handicraft implicit for the reader. Hence, a book like this is very valuable as the chapters present, discuss, and illustrate the use of different approaches.

With some exceptions, the methods and approaches presented in the chapters discussed here are not new in themselves. Rather, most of the qualitatively oriented chapters discuss well-established approaches to data collection and analysis in the social sciences. However, as several chapters show, these approaches are adapted to grasp emerging aspects of PLD in new and changing organisational environments. For instance, both diaries and stimulated recall interviews have a long history as data sources but their potential for generating significant insights is revitalised in organisational environments where work processes are black-boxed through technology use and important aspects of problem-solving are not observable for the researcher. The integration of work and learning in such environments thus needs to be articulated to be analysed. How professionals experience and explain these processes may be more significant in generating novel insights into PLD than the established categories and measures predefined by researchers. Nevertheless, explorative research and theorising need to go together for cumulative research insights to emerge.

In the further development of the PLD research field, a variety of methodological approaches are needed to examine the different processes and relations that such learning implies. There is a need to account for the personal dimensions of PLD, but also for collective dimensions and their relations to materiality, organisational contexts, and discourses. To take advantage of the variegated opportunities that qualitative approaches can offer, studies that take different units of analysis are needed. I envision that the different qualitative methods and approaches presented in this volume will contribute to future research as well. We will probably both see a renewed interest in phenomenological and experience-based approaches, and a growing attention to the way work and learning are accomplished and integrated in unfolding practices. We will see more studies oriented to micro-phenomena in specific professional actions, and more studies that span organisational settings and processes. I also envision a stronger presence of interventionist studies, as the processes and political frameworks regulating research, learning, and innovation become intertwined. Hopefully, we will also see more collaborative efforts between professional learning researchers and scholars in other fields targeting work organisations and practices. The complexity of PLD research as a field will then reflect and adapt to the complexity of the phenomenon of professional learning and development in a dynamic manner.

References

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly, 66*(4), 1173–1212. https://doi.org/10.1177/00018392211016755

Bailey, D., Faraj, S., Hinds, P., Leonardi, P. M., & von Krogh, G. (2022). We are all theorists of technology now: A relational perspective on emerging technology and organizing. *Organization Science, 33*(1), 1–18. https://doi.org/10.1287/orsc.2021.1562

Cuyvers, K., Van den Bossche, P., & Donche, V. (2022). Longitudinal case study research to study self-regulation of professional learning: Combining observations and stimulated recall interviews throughout everyday work. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 579–600). Springer.

Damşa, C. I., & Nerland, M. (2016). Student learning through participation in inquiry activities: Two case studies in teacher and computer engineering education. *Vocations and Learning, 9*, 275–294. https://doi.org/10.1007/s12186-016-9152-9

David, L., Endedijk, M., & Van den Bossche, P. (2022). Investigating interaction dynamics: A temporal approach to team learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 187–209). Springer.

Denzin, N. K., & Lincoln, Y. S. (2018). Introduction. The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 1–26). Sage.

Edwards, A. (2010). *Being an expert professional practitioner: The relational turn in expertise*. Springer.

Filliettaz, L., Garcia, S., & Zogmal, M. (2022). Video-based interaction analysis: A research and training method to understand workplace learning and professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 419–440). Springer.

Gerholz, K.-H., & Wagner, A. (2022). Design-based research – Grounding, understanding and empirical illustration in the context of vocational education. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 513–534). Springer.

Harteis, C. (2022). Research on workplace learning in times of digitalisation. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning* (Vol. 31). Professional and Practice-based Learning, Springer. https://doi.org/10.1007/978-3-030-89582-2_19

Harteis, C. (2022). Delphi-technique as a method for research on professional learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 351–371). Springer.

Herzum, M., & Simonsen, J. (2019). Configuring information systems and work practices for each other: What competences are needed locally? *International Journal of Human-Computer Studies, 122*, 242–255.

Kajamaa, A., & Hyrkkö, S. (2022). Change laboratory method for facilitating transformative agency and collective professional learning – Case from a Finnish elementary school. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 535–556). Springer.

Kyndt, E., & Aerts, J. (2022). Addressing 'wicked problems' using visual analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 329–348). Springer.

Lemmetty, S., Collin, K., Glăveanu, V., & Paloniemi, S. (2022). Capturing actions of communities: Towards virtual ethnography and digital tools in researching organizations and workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 397–418). Springer.

Littlejohn, A., & Pammer-Schindler, V. (2022). Technologies for professional learning. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research approaches on workplace learning. Professional and practice-based learning* (Vol. 31). Springer. https://doi.org/10.1007/978-3-030-89582-2_15

Littlejohn, A., Kennedy, E., & Laurillard, D. (2022). Professional learning analytics: Understanding complex learning processes through measurement, collection, analysis, and reporting of MOOC data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 557–578). Springer.

Ludvigsen, S., & Nerland, M. (2018). Learning at work: Social practices and units of analysis. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 147–156). Routledge.

Markauskaite, L., & Goodyear, P. (2016). *Epistemic fluency and professional education*. Springer.

Markauskaite, L., Goodyear, P., & Sutherland, L. (2021). Learning for knowledgeable action: The construction of actionable conceptualisations as a unit of analysis in researching professional learning. *Learning, Culture and Social Interaction, 31*, Part B. https://doi.org/10.1016/j.lcsi.2020.100382

Maxwell, J. A., & Chmiel, M. (2014). Notes toward a theory of qualitative data analysis. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis*. Sage.
Nerland, M., & Hasu, M. (2021). Challenging the belief in simple solutions: The need for epistemic practices in professional work. *Medical Education, 55*(1), 65–71. https://doi.org/10.1111/medu.14294
Nicolini, D., Mørk, B. E., Masovic, J., & Hanseth, O. (2018). The changing nature of expertise: Insights from TAVI. *Studies in Continuing Education, 40*(3), 306–322. https://doi.org/10.1080/0158037X.2018.1463212
Noordegraaf, M. (2016). Reconfiguring professional work: Changing forms of professionalism in public services. *Administration and Society, 48*(7), 783–810. https://doi.org/10.1177/0095399713509242
Noordegraaf, M. (2020). Protective or connective professionalism? How connected professionals can (still) act as autonomous and authoritative experts. *Journal of Professions and Organization, 7*(2), 205–223. https://doi.org/10.1093/jpo/joaa011
Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2020). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science, 32*(1), 18–41.
Paloniemi, S., Penttonen, M., Eteläpelto, A., Hökkä, P., & Vähäsantanen, K. (2022). Integrating self-reports and electrodermal activity (EDA) measurement in studying emotions in professional learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 87–109). Springer.
Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.
Silvennoinen, M., Parviainen, T., Malinen, A., Karjalainen, S., Manu, M., & Vesisenaho, M. (2022). Combining physiological and experiential measures to study adult learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 137–164). Springer.
Tronsmo, E. (2020). Changing conditions for teachers' knowledge work: New actor constellations and responsibilities. *The Curriculum Journal, 31*, 775–791. https://doi.org/10.1002/curj.60
Vähäsantanen, K., & Arvaja, M. (2022). The narrative approach to research professional identity: Relational, temporal, and dialogical perspectives. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 373–395). Springer.

Monika Nerland (PhD) is Professor in the Department of Education, University of Oslo. Her research focuses on professional learning in education and work contexts. Particular interests are how the knowledge practices of professional communities provide distinct opportunities for learning and identity formation, and how professional expertise is affected by technological and societal change. She has led several research projects that investigated these themes comparatively across professions. She is currently leading the research project CORPUS, which investigates implications of digitalisation for work and learning in the primary health sector in Norway.

Chapter 28
How to Deal with the Complexity in Research on Workplace Learning

Erno Lehtinen

Abstract Workplace learning takes place in conditions that are, in many respects, different from that occurring in on educational institution. Due to this, essential workplace learning may not always be apparent and may therefore be difficult to capture with conventional methods of learning research. The present book discusses methodological issues from the standpoint of research design, data collection and data analysis. There is a rich and inspiring collection of established and emerging approaches, all of which attempt to tackle the challenge of research on workplace learning in its complexity. Professional competences and the processes of learning that support the development of them are multi-layered. There is no possibility to collect data about them in a uniform way using a single methodology. The book presents multimodal approaches to the collection of data. The implication of this challenge call for novel approaches to data analysis and particularly well-developed theoretical frameworks for interpreting and combining the findings from different data sources. In this discussion, I conclude that the chapters in this book provide powerful approaches for future research on workplace learning, but also new challenges relating to a deeper theoretical understanding of the affordances provided by the methods.

Keywords Workplace learning · Research · Complexity · Discussion

28.1 Introduction

Research on learning that takes place in professional contexts poses challenges different from those faced when studying learning that occurs in intentionally structured learning contexts such as schools. One fundamental challenge is to

---

E. Lehtinen (✉)
Department of Teacher Education, University of Turku, Turku, Finland

Education Academy, Vytautas Magnus University, Kaunas, Lithuania
e-mail: erno.lehtinen@utu.fi

M. Goller et al. (eds.), *Methods for Researching Professional Learning and Development*, Professional and Practice-based Learning 33,
https://doi.org/10.1007/978-3-031-08518-5_28

observe learning when it takes place. It is common that learning is a by-product of new working methods or solving novel problems (Lehtinen et al., 2014). Incremental learning is often an imperceptible by-product of everyday work (e.g., Edmondson, 2002). School-based learning organises teaching and learning based on a curriculum, and the expectation is that students will learn certain skills and knowledge within a given period of time. It is possible to time data collection in this kind of environment in such a way that it focuses on certain teaching-learning processes and catches the increase in learning outcomes during that period. Only a small part of professional learning situations is organised in the same manner as school-based learning. In typical professional learning contexts, the researcher does not know in advance when crucial learning takes place and what exactly participants are supposed to learn.

The fact that workplace learning occurs in concrete work contexts leads easily to overstating the situated nature of learning in a trivial manner (Anderson et al., 1996). It is difficult to comprehend the richness of professional development if learning is defined and measured only in terms of immediate changes in work processes and self-reported experiences. As a result, a variety of long-term cognitive effects on individual minds, including preparation for future learning (Bransford & Schwartz, 1999), may not be apparent. Complex professional competences are multi-layered in nature (Lehtinen et al., 2020). Hence, professional learning can simultaneously include processes such as the unintentional formation of neural networks, the intentional construction of conceptual knowledge, and the gradual participation in shared professional practices (Lehtinen, 2012). Often, the most important "learning outcomes" do not occur with individuals, but can be observed in the collective level achievement of teams or organisations (e.g. Palonen, 2022, Chap. 22; David et al., 2022, Chap. 9). When studying professional learning, it is highly important–although challenging–to define appropriate units of analysis (see Säljö, 2009). As a consequence, it is also necessary to develop research methods which focus on various units of analysis, ranging from micro level processes in individuals (Silvennoinen et al., 2022, Chap. 7) to social ties in large organisations (Palonen, 2022, Chap. 22).

In spite of the many problems (such as common method variance and social desirability), self-reports have been widely used in social and behavioural sciences, including studies on workplace learning (Samuelstuen, 2003). It is reasonable to assume that these problems, particularly social desirability, are serious also when self-report data are collected during working life (e.g. Podsakoff & Organ, 1986). It is difficult to completely avoid the use of self-reports altogether in social and behavioural sciences, but it is important to find ways to combine self-reports with less subjective forms of data collection.

In order to address the above-mentioned challenges of workplace learning research, such as recognising long-term learning, catching relevant episodes, dealing with multi-layered complexity, and overcoming the limitations of self-reports, novel methods and a combination of methods are needed. The chapters of this book provide a variety of perspectives on these challenges. There are chapters devoted to research approaches, and others describe alternative methods of collecting and analysing data. Chapter classification is not always straightforward, and in this case there are some overlaps. Chapters on general research approaches also cover specific data collection methods. The papers in all these groups address directly or indirectly the challenges I raised at the beginning of this discussion.

28.2 How to Cope with the Complexity

The book presents several approaches and methods that would allow frequently repeated or continuous data collection over time. Temporal change is in focus in the change laboratory (Kajamaa & Hyrkkö, 2022, Chap. 24) and design-based research (Gerholz & Wagner, 2022, Chap. 23) approaches, which both include phases. In this book, David et al. (2022, Chap. 9), in their chapter on temporal approach to team learning, also emphasise long-term learning. Each of these chapters focuses on the collective level changes of organisations, systems, or teams. A number of other chapters describe specific data collection methods such as diaries (Rausch et al., 2022, Chap. 3), experience sampling (Seifried & Rausch, 2022, Chap. 2), and long-term case studies on self-regulation (Cuyvers et al., 2022, Chap. 26) that are aimed at addressing the matter of recognising individual learning. All of these examples illustrate the importance of continuous or frequent data collection, which enables a better understanding of work processes than traditional cross-sectional research or longitudinal studies based on repeated testing with long intervals.

Digital systems play a big role in modern work. The advantage of computers and other digital devices is their ability to keep a detailed log (Spliethoff & Abele, 2022, Chap. 8). Log data is widely used in digital training and testing environments to analyse learning. In principle, log data collected in digital platforms during regular work processes could be used to perform an indirect learning analytics, which could uncover gradual learning from everyday work. In particular, log data, but also other types of data based on small pieces of information, makes different kinds of learning analytics possible in workplace learning (Littlejohn et al., 2022, Chap. 25). Big data and learning analytics measures are quite attractive but also problematic approaches for research on learning (Buckingham et al., 2013; Mittelmeir et al., 2017). Easy methods of collecting large amounts of data can lead to “random empiricism”. Even the most sophisticated analysis methods will produce meaningless results if the input data is meaningless. It is important to consider the theoretical relevance of the collected data in all of these data collection methods.

The use of visual elements in many professions has given rise to a completely new area of research (Gegenfurtner et al., 2011; Lehtinen et al., 2020). Eye tracking has been an important method for this research field. In her chapter, Jossberger (2022, Chap. 21) gives an example of how this method can be used to gain a detailed insight into the professional development of radiologists. Medical imaging has been a field that has used eye tracking extensively as a tool to analyse learning trajectories from novices to experts as well as a tool for professional training (Gegenfurtner et al., 2017). The majority of the studies on visual expertise and professional learning in medical imaging have used so-called remote eye tracking connected to desktop computers. Recent developments in eye tracking technology have made it possible to measure data from many types of phenomena and in many situations (e.g., teachers' professional vision in the classroom) with mobile eye tracking glasses (Pouta et al., 2021; Stürmer et al., 2017). The analysis of eye movement data can be challenging even in very controlled situations (Holmqvist et al., 2011) because there is a huge amount of data and noise, and there is no direct relationship between gaze and thinking. Mobile eye tracking in natural situations such as classrooms complicates data analysis further. Additionally, mobile eye tracking produces a new type of intensive video data that requires powerful analysis methods (see Filliettaz et al., 2022, Chap. 19; Steffen & Pouta, 2022, Chap. 13).

Following a long anti-biological period in social and behavioural sciences, there has been a trend over the past two to three decades to analyse humans from the perspective of psychophysiology (Grings & Dawson, 1978) and to consider the embodied nature of human behaviour (Lakoff, 2012). In the present book, this trend is also evident. There are several chapters that present ways to study participants' activities at work using a variety of multimodal physiological measures. There is a chapter in this book that discusses the role of physiological measures in following learning processes itself (Silvennoinen et al., 2022, Chap. 7) but most of them focus on the measurement of learning-related emotions during work processes (e.g. Kärner & Sembill, 2022, Chap. 6). A few chapters in this book present the possibility of simultaneously analysing physiological and behavioural data with self-reports (e.g., Paloniemi et al., 2022, Chap. 5). By using brain imaging and physiological measures, we will be able to analyse processes that are not possible to analyse with mere behavioural measures or self-reports.

Experimental psychology has a long tradition of using physiological measurements to answer very specific questions in laboratory studies. The results of these studies can be used to interpret the data of physiological methods used in workplace situations. Nonetheless, the biggest challenge in using multimodal data collection including physiological measures is in combining and interpreting them within a coherent theoretical framework. As far as emotions are concerned, previous laboratory studies and theories developed on their basis may provide at least some basis for interpreting data in less controlled workplace situations. When multimodal measures are used in studies to understand complex cognitive processes, the challenges are much greater.

There are very few well-established theoretical explanations based on experimental evidence relating specific processes of learning to physiological measures, including brain imaging. Without this type of evidence-based theory, conclusions and recommendations about workplace learning or professional learning and development in general and improvement suffer from the same limitations that Bowers (2016) outlined for neuroscience-based recommendations for teaching. "Unfortunately, a careful reading of the literature reveals that the successes are either (a) trivial, in the sense that the recommendations are self-evident, (b) misleading, in the sense that the recommendations are already well established (based on behavioural studies), or (c) unwarranted, in the sense that the recommendations are based on misrepresentations of neuroscience or the conclusions do not follow from the neuroscience" (Bowers, 2016, p. 2). This does not mean that multimodal methods cannot be used to investigate learning. On the contrary, I think that it is very important to study and understand the embodied nature of learning, as well as to develop a theoretical grasp of how various physiological processes affect learning. However, very much basic research is needed before neuroscience can substantially contribute to research on complex learning in workplaces.

During the past few years, there have been a strong trend towards enriching the dominant variable oriented statistical methods with person oriented approaches. Latent profile and class analyses (LPA, LCA) are promising approaches for describing phenomena in workplaces in novel ways (Bauer, 2022, Chap. 11). If combined with latent transition analyses (Collins & Lanza, 2010), the person oriented approach can also provide important insights into longitudinal processes. Additionally, this book provides a variety of at least partly new or less often used qualitative and quantitative methods for analysing complex data. Because data collected from complex workplace situations does not always meet the criteria of classical statistical analysis, new methods such as data mining (Ifenthaler, 2022, Chap. 14), visual analysis (Kyndt & Aerts, 2022, Chap. 15), PLS-modeling (Goller & Hilkenmeier, 2022, Chap. 12), and Bayesian statistics (Nokelainen et al., 2022, Chap. 10) are needed for analysis. The Bayesian approach and PLS-modelling can both help to solve the challenge posed by the size and distribution of the data.

PLS-modelling can provide features such as flexible ways to use mediation and moderators, as well as formative latent constructs, which can be very useful for studies on workplace learning. PLS-modelling has long tradition in business studies, but it is not widely used in educational research, and there are no strong established publication conventions, as there are in covariance-based SEM. Even though Bayesian statistics is becoming more well-known within the educational research community, strong research examples that illustrate the strengths of this approach are still lacking. Chapters in this book that discuss PLS-modelling and Bayesian statistics present good examples that may encourage researchers to expand their methodological repertoire. The chapter on data mining (Ifenthaler, 2022, Chap. 14) deals with the topic on a more general level rather than providing examples of concrete analyses, but it does provide an interesting overview of the affordances and challenges of this approach.

Using technology, complex data can be visualized in a variety of ways. Visualizations are often used to complement numerical information when presenting results. However, visualizations can also be used as means to analyse phenomena from the data which are difficult or impossible to capture using statistical procedures (Kyndt & Aerts, 2022, Chap. 15; Palonen, 2022, Chap. 22). The challenge in publishing findings based on visual analysis is that in behavioural and social sciences are accustomed to relying on quasi-exact measures such as p-values. In research on workplace learning, it is essential to discuss the exploratory and confirmatory nature of research in more detail.

28.3 Conclusion

Those planning research on workplace learning will find the chapters in this book to be an inspiring collection of approaches and concrete methods for data collection and analysis. It is true that this rich variety of methods can be very valuable, but we need to be careful not to misuse them. This richness in available methods means that a researcher who formulates strongly theory framed research questions could find methods that are perfectly suited to the research questions. On the other hand, the large variety of appealing data collection methods may tempt a researcher toward eclectic method-driven research, in which methods are used without coherent theoretical justification. The majority of readers of this book may choose the first option.

References

Anderson, J. R., Simon, H. A., & Reder, L. M. (1996). Situated learning and education. *Educational Researcher, 25*, 5–11.

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 243–268). Springer.

Bowers, J. S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review, 123*(5), 600–612. https://doi.org/10.1037/rev0000025

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61–100.

Buckingham Shum, S., Hawksey, M., Baker, R., Jeffery, N., Behrens, J. T., & Pea, R. (2013). *Educational data scientists: A scarce breed*. Paper presented at the Learning Analytics and Knowledge (LAK), Leuven, Belgium.

Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley.

Cuyvers, K., Van den Bossche, P., & Donche, V. (2022). Longitudinal case study research to study self-regulation of professional learning: Combining observations and stimulated recall interviews throughout everyday work. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 579–600). Springer.

David, L., Endedijk, M., & Van den Bossche, P. (2022). Investigating interaction dynamics: A temporal approach to team learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 187–209). Springer.

Edmondson, A. C. (2002). The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science, 13*(2), 28–146. https://doi.org/10.1287/orsc.13.2.128.530

Filliettaz, L., Garcia, S., & Zogmal, M. (2022). Video-based interaction analysis: A research and training method to understand workplace learning and professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 419–440). Springer.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education, 13*, 212–225. https://doi.org/10.1016/j.compedu.2017.06.001

Gerholz, K.-H., & Wagner, A. (2022). Design-based research – Grounding, understanding and empirical illustration in the context of vocational education. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 513–534). Springer.

Goller, M., & Hilkenmeier, F. (2022). PLS-based structural equation modelling: An alternative approach to estimate complex relationships between unobserved constructs. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 269–292). Springer.

Grings, W. W., & Dawson, M. E. (1978). *Emotions and bodily responses: A psychophysiological approach*. Academic.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Ifenthaler, D. (2022). Data mining and analytics in the context of workplace learning: Benefits and affordances. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 313–327). Springer.

Jossberger, H. (2022). Eye tracking in professional learning and development: Uncovering expertise development among residents in radiology. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 467–489). Springer.

Kajamaa, A., & Hyrkkö, S. (2022). Change laboratory method for facilitating transformative agency and collective professional learning – Case from a Finnish elementary school. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 535–556). Springer.

Kärner, T., & Sembill, D. (2022). Multimodal stress assessment in working and learning contexts: Using physiological, observational and experience-based data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 111–136). Springer.

Kyndt, E., & Aerts, J. (2022). Addressing 'wicked problems' using visual analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 329–348). Springer.

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in Cognitive Science, 4*(4), 773–785.

Lehtinen, E. (2012). Learning of complex competences: On the need to coordinate multiple theoretical perspectives. In A. Koskensalo, J. Smeds, A. Huguet, & R. de Cillia (Eds.), *Language: Competencies – Contact – Change* (pp. 13–27). LIT Verlag.

Lehtinen, E., Hakkarainen, K., & Palonen, T. (2014). Understanding learning for the professions: How theories of learning explain coping with rapid change. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International handbook of research in professional practice-based learning* (pp. 199–224). Springer. https://doi.org/10.1007/978-94-017-8902-8

Lehtinen, E., Gegenfurtner, A., Helle, L., & Säljö, R. (2020). Conceptual change in the development of visual expertise. *International Journal of Educational Research, 100*, 101545. https://doi.org/10.1016/j.ijer.2020.101545

Littlejohn, A., Kennedy, E., & Laurillard, D. (2022). Professional learning analytics: Understanding complex learning processes through measurement, collection, analysis, and reporting of MOOC data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 557–578). Springer.

Mittelmeier, J., Edwards, R. L., Davis, S. K., Nguyen, O., Murphya, V. L., Brummer, L., & Rienties, B. (2017). 'A double-edged sword. This is powerful but it could be used destructively': Perspectives of early career education researchers on learning analytics. *Frontline Learning Research, 6*(2), 20–38.

Nokelainen, P., Aldahdouh, T. Z., & Aldahdouh, A. A. (2022). Bayesian statistics in the research field of professional learning and development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 213–241). Springer.

Palonen, T. (2022). Seeing workplaces from a social network analysis (SNA) approach. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 491–512). Springer.

Paloniemi, S., Penttonen, M., Eteläpelto, A., Hökkä, P., & Vähäsantanen, K. (2022). Integrating self-reports and electrodermal activity (EDA) measurement in studying emotions in professional learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 87–109). Springer.

Podsakoff, P. M., & Irgan, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531–544.

Pouta, M., Palonen, T., & Lehtinen, E. (2021). Student teachers' and experienced teachers' professional vision of students' rational number concept understanding. *Educational Psychology Review, 33*(1), 109–128. https://doi.org/10.1007/s10648-020-09536-y

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering informal workplace learning by using diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 43–70). Springer.

Säljö, R. (2009). Learning, theories of learning, and units of analysis in research. *Educational Psychologist, 44*(3), 202–208.

Samuelstuen, M. S. (2003). Psychometric properties and item-keying direction effects for the learning and study strategies inventory-high school version with Norwegian students. *Educational and Psychological Measurement, 63*, 430–455.

Seifried, J., & Rausch, A. (2022). Applying the experience sampling method to research on workplace learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 19–41). Springer.

Silvennoinen, M., Parviainen, T., Malinen, A., Karjalainen, S., Manu, M., & Vesisenaho, M. (2022). Combining physiological and experiential measures to study adult learning. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 137–164). Springer.

Spliethoff, L., & Abele, S. (2022). Measuring professional competence using computer-generated log data. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 165–186). Springer.

Steffen, B., & Pouta, M. (2022). Participant's video annotations as a database to measure professional development. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for researching professional learning and development: Challenges, applications, and empirical illustrations* (pp. 293–312). Springer.

Stürmer, K., Seidel, T., Müller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft, 20*, 75–92.

Erno Lehtinen (PhD) is emeritus professor of education at the University of Turku and visiting professor at the Vytautas Magnus University (Lithuania). He was vice-rector of the University of Turku, Academy professor (A Finnish distinguished professor funded by the Academy of Finland) and the head of the Finnish national "Doctoral Programme on Multidisciplinary Research on Learning Environments" and the doctoral programme on "Research on learning, teaching and learning environments" of the University of Turku. Lehtinen has published about 400 scientific publications including many articles in international refereed journals of educational sciences and educational psychology. Lehtinen is selected member of Finnish Academy of Science and Letters and International Academy of Education. He was president of European Association for Research on Learning and Instruction (EARLI 2001–2003) and founding editor-in-chief of the Frontline Learning Research. In 2009 he got the Oeuvre Award of the EARLI.