Educational Linguistics

Matthew E. Poehner
Ofra Inbar-Lourie *Editors*

Toward a Reconceptualization of Second Language Classroom Assessment

Praxis and Researcher-teacher Partnership

Springer

Educational Linguistics

Volume 41

Series Editor
Francis M. Hult, University of Maryland, Baltimore County, Baltimore, USA

Editorial Board
Marilda C. Cavalcanti, Universidade Estadual de Campinas, Campinas, Brazil
Jasone Cenoz, University of the Basque Country, Leioa, Spain
Angela Creese, University of Stirling, Stirling, United Kingdom
Ingrid Gogolin, University of Hamburg, Hamburg, Germany
Christine Hélot, Université de Strasbourg, Strasbourg, France
Hilary Janks, University of the Witwatersrand, Johannesburg, South Africa
Claire Kramsch, University of California, Berkeley, USA
Constant Leung, King's College London, London, United Kingdom
Angel Lin, Simon Fraser University, Burnaby, Canada
Alastair Pennycook, University of Technology, Sydney, Australia

Educational Linguistics is dedicated to innovative studies of language use and language learning. The series is based on the idea that there is a need for studies that break barriers. Accordingly, it provides a space for research that crosses traditional disciplinary, theoretical, and/or methodological boundaries in ways that advance knowledge about language (in) education. The series focuses on critical and contextualized work that offers alternatives to current approaches as well as practical, substantive ways forward. Contributions explore the dynamic and multi-layered nature of theory-practice relationships, creative applications of linguistic and symbolic resources, individual and societal considerations, and diverse social spaces related to language learning.

The series publishes in-depth studies of educational innovation in contexts throughout the world: issues of linguistic equity and diversity; educational language policy; revalorization of indigenous languages; socially responsible (additional) language teaching; language assessment; first- and additional language literacy; language teacher education; language development and socialization in non-traditional settings; the integration of language across academic subjects; language and technology; and other relevant topics.

The *Educational Linguistics* series invites authors to contact the general editor with suggestions and/or proposals for new monographs or edited volumes. For more information, please contact the Associate Editor: Natalie Rieborn, Van Godewijckstraat 30, 3300 AA Dordrecht, The Netherlands.

More information about this series at http://www.springer.com/series/5894

Matthew E. Poehner • Ofra Inbar-Lourie
Editors

Toward a Reconceptualization of Second Language Classroom Assessment

Praxis and Researcher-teacher Partnership

Springer

Editors
Matthew E. Poehner
Department of Curriculum and Instruction, College of Education
The Pennsylvania State University
University Park, PA, USA

Ofra Inbar-Lourie
The Jaime & Joan Constantiner, School of Education
Tel Aviv University
Tel Aviv, Israel

ISSN 1572-0292 ISSN 2215-1656 (electronic)
Educational Linguistics
ISBN 978-3-030-35080-2 ISBN 978-3-030-35081-9 (eBook)
https://doi.org/10.1007/978-3-030-35081-9

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Contents

Contributors

Beverly Baker Official Languages and Bilingualism Institute, University of Ottawa, Ottawa, ON, Canada

Tineke Brunfaut Department of Linguistics and English Language, Lancaster University, Lancaster, UK

Cheri Chan Teacher Education and Learning Leadership, Faculty of Education, University of Hong Kong, Hong Kong, China

Kristin J. Davin Cato College of Education, University of North Carolina at Charlotte, Charlotte, NC, USA

Chris Davison School of Education, Faculty of Arts and Social Sciences, University of New South Wales, Sydney, Australia

Ana Maria Ducasse School of Global, Urban and Social Studies, RMIT Melbourne Institute of Technology, Melbourne, Australia

Martin East School of Cultures, Languages and Linguistics, Faculty of Arts, The University of Auckland, Auckland, New Zealand

Joyce Germain Alaqsite'w Gitpu School, Listuguj Mi'gmaq First Nation, Pointe-à-la-Croix, QC, Canada

Luke Harding Department of Linguistics and English Language, Lancaster University, Lancaster, UK

José D. Herazo Facultad de Educación y Ciencias, Universidad de Córdoba, Montería, Colombia

Kathryn Hill La Trobe University, Melbourne, Australia

Ofra Inbar-Lourie The Jaime & Joan Constantiner, School of Education, Tel Aviv University, Tel Aviv, Israel

Constant Leung Centre for Language, Discourse & Communication, School of Education, Communication & Society, King's College London, London, UK

Tziona Levi Ministry of Education, Jerusalem, Israel

Michael Michell School of Education, Faculty of Arts and Social Sciences, University of New South Wales, Sydney, Australia

Matthew E. Poehner Department of Curriculum and Instruction, College of Education, The Pennsylvania State University, University Park, PA, USA

Angela Scarino Research Centre for Languages and Cultures, School of Creative Industries, University of South Australia, Adelaide, Australia

Rémi A. van Compernolle Department of Modern Languages, Carnegie Mellon University, Pittsburgh, PA, USA

Chapter 1
An Epistemology of Action for Understanding and Change in L2 Classroom Assessment: The Case for Praxis

Matthew E. Poehner and Ofra Inbar-Lourie

Abstract Although the potential for classroom assessment to support L2 teaching and learning has long been recognized, conceptual frameworks for appraising specific assessment practices and determining how they may be developed has remained elusive. In this chapter, we outline a potential way forward in L2 classroom assessment research and practice that envisions them as existing in a bidirectional, reciprocal relationship. We discuss this through the prism of *praxis*, as brought to the attention of L2 researchers in the work of Lantolf and Poehner (Sociocultural Theory and the pedagogical imperative in L2 education. Vygotskian praxis and the research/practice divide. Routledge, London, 2014). We begin by identifying the epistemological foundations of much assessment scholarship to date, arguing that these have led, on the one hand, to prescribing assessment practices to teachers (a unidirectional relationship between theory/research and practice), or, on the hand, to documenting existing classroom assessment practices without any attempt to consider them in a broader context or to improve upon them. Following this, our discussion turns to praxis, which represents a unity of theory/research and practice wherein these inform one another and change together. Specifically, theory offers principles and concepts that teachers may draw upon to construct practices in a reasoned manner that is responsive to but that goes beyond firsthand experience. Practice, for its part, serves to identify ways in which theory may need to be revised and expanded. Understood as praxis, general conceptual frameworks of classroom assessment and the ways in which it is practiced in particular contexts must be developed in tandem. We argue that this can be productively pursued through

M. E. Poehner (✉)
Department of Curriculum and Instruction, College of Education, The Pennsylvania State University, University Park, PA, USA
e-mail: mep158@psu.edu

O. Inbar-Lourie
The Jaime & Joan Constantiner, School of Education, Tel Aviv University, Tel Aviv, Israel
e-mail: ofrain@tauex.tau.ac.il

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_1

collaboration between assessment researchers and teachers. The chapter concludes with an overview of how praxis is taken up and realized in each of the chapters in the book.

Introduction

Assessment as a regular feature of second language (L2) classroom activity has long been understood as indispensable to monitoring student progress and informing instructional decisions (Genesse and Upshur 1996; Harlen and James 1997). Over the past 20 years, L2 language teaching researchers, as well as general education researchers, have debated whether assessment conducted in classroom learning contexts should align with criteria created for formal standardized tests or should follow its own principles and frameworks (e.g., Assessment Reform Group 2002; Moss 2003; Poehner 2011; Teasdale and Leung 2000). One might argue that assessment researchers and professional test developers hold the expertise required to design and interpret reliable and valid instruments and procedures. According to this perspective, teacher participation in this process may include administering the assessment and, in some instances, following established procedures – provided they have appropriate training – to interpret assessment outcomes and possibly to even develop assessments of their own. Another view proposes that assessments undertaken in classrooms tend to differ from externally created, large-scale tests in certain fundamental ways. For instance, Moss (2003), working outside the L2 field, argued that classroom assessments need not adhere to the many standards that characterize formal tests and that it is perhaps even desirable for them to depart from such standards. In her view, imposing time limits, restricting access to resources, and disallowing interaction with peers make little sense for classroom assessments given that the aim of classroom assessments is typically gain insights into how learners are performing or will perform in classroom settings. Following this line of reasoning, Moss and others (e.g., Torrance and Pryor 1998) proposed that classroom-based assessments intended to support processes of teaching and learning ought to be developed and evaluated according to their own criteria rather than the adoption of those created for formal testing, a notion that has also gained some traction within the L2 field (e.g., Teasdale and Leung 2000).

Encouraging teachers to create their own assessments and to establish their own standards for determining their quality and appropriateness may be appealing to many, but it is not without its own problems. One question that immediately arises is whether teachers currently have the expertise required for this undertaking, and indeed whether they are aware of the knowledge that might be relevant to ensuring that assessments function the way they wish them to. In this regard, the notion of *language assessment literacy* (Inbar-Lourie 2017) has emerged as a way of understanding the knowledge base teachers likely require in order to engage in assessment in a manner that advances their instructional goals. Another issue concerns the role of assessment researchers. What kind of involvement ought they to have in the

design of classroom-based assessments? Should they provide instruments and directions for implementing them to teachers? Is it more appropriate to simply document how teachers are assessing learning and to withhold any evaluative commentary in favor of simply relaying what is done, the stated reasons for those practices, and teachers' and learners' experiences with the assessments? Are there ways in which assessors might engage in collaboration with teachers? What might this look like?

Recent years have seen a number of models or sets of principles that teachers might draw upon to inform their assessment activities. *Assessment-for-learning* (Black and Wiliam 2009), *dynamic assessment* (Poehner 2008), *learning-oriented assessment* (Turner and Purpura 2015), and other related proposals each share a commitment to moving beyond the accepted norms of standardized assessments to arrive at a framework for classroom assessment that is better aligned to the interests and needs of teachers and learners. While this work has made important advances toward that goal, it is not always clear to what extent the proposals themselves have been elaborated as a result of applications to classrooms, nor are teacher experiences with the various models fully accounted for in research. In our view, for these frameworks to realize their potential to support L2 teaching and learning, researchers must partner with classroom practitioners, offering theoretical and methodological expertise and conceptual guidance, but also learning from teachers in order to evaluate their frameworks and continue to revise them. We propose that through this process we can arrive at a vantage from which to determine the features and practices from standardized assessments that might be relevant to classrooms, those that need to be revised or even abandoned, and what new insights from teachers and learners can inform the ongoing development of appropriate frameworks.

The position we are proposing is not simply the result of our own experiences collaborating with teachers, although we have both done so and found this to be essential to our work. Rather, this position reflects an epistemology that runs counter to the more conventional stance that research and theory building (including the elaboration of principles and models) might be pursued exclusively by researchers and then conveyed to teachers for them to apply it (Delanty and Strydom 2003). At the same time, our position does not forego any interest generating principles and models that might be applied widely in favor of an exclusive interest in the experiences of individual teachers and learners. The stance adopted throughout this book values teachers' and learners' insights and understands them in relation to broader theoretical proposals, as both are ultimately necessary for the advancement of research and practice. The epistemology to which we are referring is known as *praxis* and has its roots in a Marxian critical approach to science (Kemmis et al. 2008). According to this view, a scientific enterprise – including research and the development of theory as well as the design of practices to impact daily life – occurs only when these two domains are recognized as interrelated and mutually informative. More specifically, praxis regards theory as providing principles and concepts that allow teachers to build their practice in a reasoned, reflective manner that goes beyond firsthand experience. At the same time, practice serves as a testing ground for theory, pointing to areas in need of revision and expansion. In praxis then, theory

and research are accountable to practice; praxis represents a unity of theory and practice wherein they inform one another and change together.

This chapter introduces the concept of praxis and considers its implications for thinking about and doing language classroom assessment in support of teaching and learning. We argue that praxis represents a shift away from tendencies that, on the one hand, prescribe assessment practices to teachers (a unidirectional relationship between theory/research and practice), or that, on the other hand, simply document existing classroom assessment practices without any attempt to revise them. Our position calls for the dialogic, interactive development of how classroom language assessment is conceptualized and how it is 'practiced', an undertaking that we propose may be pursued through researcher-teacher partnership. We conclude the chapter by previewing how praxis functions as an orienting basis for each of the papers included in the book.

Approaches to 'Doing' Science

Much of the history of thinking concerned with epistemology from the mid-Nineteenth Century onward has been an effort to deal with growing skepticism toward the Enlightenment belief that science is *the* mode of inquiry and that through science we can arrive at absolute knowledge of everything (Berlin 2013). Garratt and Li (2005) explain that the shift from earlier conceptualizations of science to a more contemporary one can be captured as moving from a *foundationalist* to *non-foundationalist* epistemology. While the former view holds that there is a solid, irrefutable basis upon which knowledge claims can be made, the latter maintains that all knowledge is subject to critique and challenge, especially as new insights emerge. Earlier *Positivist* conceptualizations of science have given way to general acceptance that the best that human knowledge can achieve is not absolute certainty but increased probability of the accuracy of explanations (see also Peters and Burbules 2004). As Rorty (2009) observes, while the shift away from *Positivism* has had implications for the natural sciences – specifically, the scope of what science might achieve, recognition of the need for vigilance against universalizing claims, and the importance of ongoing reflection on the part of scientists and openness to change in light of new discoveries – the consequences of the decline of *Positivism* have been most acutely felt in the social and human sciences, where the object of study itself (human beings and their worlds) has been a focus of much debate (Latour 1999).

As social scientists have struggled to determine how best to proceed to construct theories, establish research methods, define the objects of their inquiry, and determine the ultimate goals of their work, several distinct traditions have emerged. Delanty and Strydom (2003) count among the major re-orientations to science *post-Positivism, Interpretivism, Critical Theory, post-Structuralism,* and *Pragmatism*, as well as perspectives that draw upon one or more of these traditions or that exist at their intersection (e.g., various feminisms, post-Colonial theories, queer theory,

etc.). Discussion of each of these philosophies of science and their relevance to L2 assessment is well beyond the scope of the present chapter, and indeed it would require a book of its own. Furthermore, given that our primary focus is how the field of L2 assessment has generally understood the relation between theory/research and practice, we limit our remarks to how this topic is conceptualized within three intellectual traditions: post-Positivism, Interpretivism, and Critical Theory. While recognizing that the L2 assessment research literature includes instances of scholars drawing from a wide range of theories and philosophies, our view of the field is that it has been heavily dominated by work that falls within a post-Positivist orientation, with some research also informed by Interpretivist perspectives. With this in mind, in what follows we foreground these two philosophies, briefly outlining some of their major characteristics and drawing attention specifically to how they conceptualize the relation between theory/research and practice. For both these traditions, we also provide examples from the L2 assessment research literature of studies that have productively investigated classroom assessment from the vantage of those orientations. Of course, this does not mean that the authors of the studies we discuss explicitly announce a commitment to either post-Positivism or Interpretivism, and it may be the case that those authors would not readily embrace either label. Be that as it may, our purpose in referencing them is because the questions they ask and how they attempt to explore those questions aligns in important ways with the traditions we are concerned with, as we will explain. This discussion should provide a useful contrast with our interest in praxis, which as we mentioned emerges from a Critical tradition. We also hasten to add that our comments that follow should not be read as either an attack on work framed according to other philosophies or an argument for the superiority of praxis. Indeed, we remain convinced of the value of each of these orientations to L2 assessment researchers. Nonetheless, we regard praxis as an orientation that has not been adequately explored in the L2 field to date and that provides a potential basis for enriching our understanding of classroom-based assessment and helping teachers and learners meet their assessment needs.

Post-positivism and L2 Classroom Assessment

As its name suggests, post-Positivism retains many of the commitments of its foundationalist predecessor albeit softened to be less absolute and universal. For instance, while positivist science is motivated by a pursuit of knowledge, generated through rigorous inquiry, that appears to be true in all instances, post-Positivism regards knowledge along the lines of what Dewey described as *warranted assertibility* (Boyles 2006). Warranted assertions, from this perspective, are beliefs arrived at through competent inquiry and that are strongly enough supported to provide a basis for confident action. As Phillips and Burbules (2000, pp. 86–87) put it,

> The postpositivist approach to research is based on seeking appropriate and adequate *warrants* for conclusions, on hewing to standards of truth and falsity that subject hypotheses (of

whatever type) to test and thus potential disconfirmation, and on being open-minded about criticism.

The language here of hypotheses, truth and falsity, and conclusions clearly bears the imprint of Positivism. However, what is equally clear is the importance of openness and change, a recognition of *fallibilism* in human knowledge production, that is, it is always and necessarily subject to revision and even rejection as new insights are obtained.

Readers are likely to recognize in post-Positivism much of what is generally accepted as 'scientific.' This includes the centrality of testing hypotheses and falsifying assumptions or claims through examination of empirical data. Indeed, within post-Positivism knowledge construction has come to be understood as a process of turning assertions supported by warrants into conjectures that can be tested. These conjectures, or hypothesis, typically concern relations among measurable variables that can be verified through experimentation, during which one variable or set of variables may be manipulated in order to observe and measure changes to other variables (Latour 1999). It is through this process that causal relations may be proposed, although again these are described in probabilistic rather than absolute terms (Phillips and Burbules 2000). It is worth adding that while measurement of the strength of relations among variables is frequently an important consideration in post-Positivist research designs, this does not preclude the use of qualitative methods. Rather, it is the commitment to representing observable phenomena in the world, untainted by the values and perspectives of individuals, that characterizes a post-Positivist orientation, regardless of whether the data that informs the representations are drawn from an experimental design, observations, large-scale surveys, or interviews.

As the 'conventional' model for scientific inquiry, post-Positivism has enjoyed a privileged status in nearly all areas of the human and social sciences, and the field of L2 classroom assessment research is no exception. For purposes of illustration of the productive and valuable research that has been generated by L2 assessment scholars, we briefly mention two examples. The reader will note that neither of these studies employed an experimental or quasi-experimental design, randomly assigning teachers and learners to a control group or to a group that receives some treatment or intervention so that results – changes over time – might be measured and compared. Following Phillips and Burbules (2000), such a design is often upheld within post-Positivism as an ideal, but social science researchers also recognize various obstacles to adhering to that model, particularly when investigations are undertaken in contexts beyond the researchers' control, such as classrooms.

The recent work of Sawaki and Koizumi (2017) offers a useful instance of the concern within post-Positivist research designs over the careful delineation of constructs to be studied and the investigation of assertions to be empirically evaluated. These authors set out to understand the impact of test feedback on subsequent instructional activity. Noting that some large-scale English language tests have begun to offer not only overall test results but also more subtle and detailed feedback, Sawaki and Koizumi highlighted the kinds of feedback provided by two

Japanese standardized English language tests: the Global Test of English Communication for Students (GTECfS) and the Eiken Test in Practical English Proficiency (Eiken). The researchers first conducted an analysis of these feedback reports and coded their specific features. They then carried out interviews with Japanese secondary school teachers and students in order to ascertain how the feedback – with attention to specific features – impacted teaching and learning practices. Turning to a project that relied heavily on qualitative research methods, we note the influential work of Rea-Dickins (2001). One of the earliest attempts to understand the dynamics of L2 classroom assessment, this study sought to gain access to "the interactional patterns between a teacher and his or her pupils, and between pupils" (Rea-Dickins 2001, p. 433). Data were taken from a larger study of English language learners and included classroom observations, recording of class interactions, and interviews with teachers and support teams. Rea-Dickins's analysis shed light on the rich repertoire of assessment strategies that a teacher can use while also raising questions regarding the implementation of formative assessments.

Taken together, the studies by Sawaki and Koizumi (2017) and Rea-Dickins (2001) highlight the valuable insights that may become available through research designs that prioritize documentation of observable behaviors and practices and that may employ a range of research methods. Before moving on, two observations are in order. In neither project was the focus on the experiences of the participants and the meanings that they derived from those experiences. This is not to say that the authors were not interested in this, but rather that the data that are reported concern what the participants were actually *doing* rather than an in-depth investigation of what it *meant* to them. As we explain, this focus, much more than selection of a particular research method, helps to distinguish research influenced by post-Positivism from work that reflects a more Interpretivist orientation. Moreover, neither the Sawaki and Koizumi (2017) nor the Rea-Dickins (2001) study sought to change the phenomena under studying. To be sure, the authors may well have regarded their work as contributing to the field's knowledge based concerning L2 classroom assessment and therefore helping to provide a basis upon which others might act to bring about change. This, however, differs sharply from the commitment to provoking change *as an inherent part of research*, which is the position taken in the Critical tradition, as we discuss later.

Interpretivism and L2 Classroom Assessment

Delanty and Strydom (2003) trace the origins of Interpretivism to the writings of Dilthey in the early Nineteenth Century and in particular his argument for the necessary disambiguating of the *human sciences,* or *Geisteswissenschaften,* from the natural sciences. In Dilthey's formulation, the latter is concerned with explanation (*erklarung*) of phenomena, whereas the proper goal of a human science should be *verstehen*, or understanding (see also Berlin 2013). Thus, even before the decline of Positivism, Dilthey had proposed the epistemological and methodological

independence of the *Geisteswissenschaften* from the natural sciences. Precisely how these human sciences ought to proceed, what they should take as their priorities, and which methods might be most appropriate has been the focus of subsequent 'generations' of scholarship in the Interpretive tradition.

According to Van Manen (1997, p. xi, emphasis added), what unites each of these approaches within Interpretivism is that they "place *human situatedness* central and are based on the belief that we can best understand human beings from the *experiential reality of their lifeworlds.*" During the early Twentieth Century, and in response to historic events including the First World War, scholars who had grown disenchanted with Positivist and were engaged in dialogue concerning possibilities for how a human science might be constructed, turned toward phenomenology and ethnography as, respectively, a theory of what constitutes meaning for individuals and a method for reading cultural artifacts and practices to arrive at an understanding of the meanings they carry for the communities in which they are found. From this perspective, a thing does not *mean* anything outside of its relation to real people; the goal of the researcher is to faithfully represent what things mean for the human beings with whom they engage in the course of their investigations.

Following Delanty and Strydom (2003), it is possible to discern multiple iterations or generations of Interpretivist thinkers. They explain that during the early Twentieth Century Interpretivist ideas were developed largely in the fields of cultural anthropology and sociology through the writings of Husserl, Malinowski, Geertz, and others. The latter half of the Twentieth Century saw this work extended to other disciplines while at the same time a so-called 'hermeneutic turn' began to develop in the work of scholars including Heidegger, Gadamer, Habermas, and Ricoeur. Delanty and Strydom (2003) explain that for these thinkers, "pure" phenomenology seeks only to describe, as purely and fully as possible, human lived experience while hermeneutics is concerned with the interpretation of experience via symbolic representation, including the production of texts.

Returning to Van Manen's (1997) characterization of each of these iterations of Interpretive science, they share an interest in the world as it is experienced by people, that is, the *essence* of experience rather than the way experience is refracted through processes of categorization, labeling, generalizing, etc. According to Van Manen (1997), essence within phenomenology is the universal quality that defines a thing, without which it would no longer be what it is. The essence of phenomena can only be apprehended through lived experience of particular instances or examples of it. He continues that engaging in research within this tradition, one aims for precision and exactness not through measurement and statistical analysis but through fully detailed descriptions of human experience. The researcher knows that "The essence or nature of an experience has been adequately described in language if the description reawakens or shows us the lived quality and significance of the experience in a fuller or deeper manner" (van Manen 1997, p. 10). Ultimately, this human science is not concerned with solving problems but with asking *meaning questions*, which can be understood more deeply but never fully; better, deeper understanding allows for more thoughtful action moving forward.

As L2 assessment researchers have turned their attention to processes of classroom assessment, it is unsurprising that Interpretivist traditions have gained some influence and have led to valuable insights. Again, given space constraints we limit ourselves to briefly commenting upon two studies that reflect Interpretivist commitments. Harding et al. (2011) set out to understanding the decision-making process that raters engage in as they mark examinee performance on an English language listening comprehension test. Importantly, the focus of research for Harding and colleagues was not whether the raters were 'correct' or 'accurate' in their ratings or even their degree of consistency. Rather, it was the process itself, with an emphasis on value judgments and personal considerations in determining the markings they gave. The researchers employed a stimulated recall procedure to form a taxonomy of the types of decisions made and to arrive at the reasons provided by the assessors for their decisions based on their marking experiences. Working under the rubric of language assessment literacy, Scarino (2013) examined the vital role of teachers' deep understanding of what assessment is as well as how they construe their role as assessors when developing their language assessment literacy. Analysis of teacher reflections, produced in a context of their professional development, revealed the importance of their preconceived beliefs and worldviews on their acquisition and practice of assessment literacy. Findings indicate the need to raise the awareness of both teachers and teacher educators as to the importance of giving a voice to individual (often tacit) beliefs, when engaging with the complexities of the assessment process and the knowledge base required for interpretive assessment and decision-making.

To this point, our discussion has made explicit the often-tacit sets of assumptions and priorities that guide appeals to theoretical frameworks, choice of research questions and topics, and selection of methods for data collection and analysis. As explained, post-Positivism and Interpretivism are not the only two philosophies of science that have gained influence in the L2 field, but their aims and arguments are likely immediately recognizable to L2 assessment researchers. Perhaps less familiar is the concept of praxis, which emerges from the tradition broadly described as Critical Theory. Despite its name, this orientation does far more than to critique other philosophies, although it does engage with them in a manner that challenges and problematizes. Equally important is the effort within a Critical tradition to 'carry forward' elements from both post-Positivism and Interpretivism as part of the emergence of something new. This carrying forward, or *aufhebung*, is crucial to our own position, which recognizes considerable value in various approaches to researching L2 classroom assessment but that is also interested in exploring new directions.

Praxis: Toward a Philosophy of Knowing Through Action

We acknowledge from the outset that there are precedents in the L2 assessment field to our present concern with praxis. For example, Pennycook (2001), considering the broader field of applied linguistics, deals overtly with Critical Theory and emphasizes the importance of interrogating the discipline's post-Positivist assumptions, constructs, and methods while also urging applied linguists to attend to the social consequences of their work, in particular the extent to which existing power structures and hierarchies are reinforced or contested. A similar argument is developed by Shohamy (2001) in her seminal book on what she refers to as critical language testing. Tracing the history of widespread standardized testing in general and language testing in particular, Shohamy lays bare the instrumental role such procedures have played in advancing political agendas, imposing values and ways of thinking, and limiting access to opportunities. Shohamy's work challenges language testers to become self-reflexive vis-à-vis the agendas that are being served by testing policies and practices. Lynch (2001), also working within the domain of language assessment, invokes a critical perspective in expressing concern over what he sees as the field's default embrace of post-Positivist research. As an alternative, Lynch emphasizes the importance of engaging directly with issues of social justice and equity through language testing.

Other examples of critical work can be found in the applied linguistics and language testing and assessment research literatures. We applaud the value and ambition of this research while we also maintain that it differs from the work in the present volume. Each of the chapters in this book adheres to the premise that knowledge of the world is arrived at through engagement in activity, and that processes of *doing* and of *understanding* exist in relation to one another. Concerns with issues of power, authority, and educational equity identified by critical applied linguists and critical language testers are very likely shared by many of the contributors to this book. However, those themes are not necessarily brought to the fore in the projects reported. Instead, the focus remains squarely on partnerships between researchers and teachers as they cooperatively endeavor to simultaneously develop conceptual understandings and practices, that is, as they engage in praxis.

With that said, perhaps the crucial starting point in Critical Theory is the view that human beings are meaning-making agents who are not only shaped by their cultural and historical environments but are actively involved in creating and reshaping those environments. This position is strongly associated with the writings of Marx, for whom the world in its present state can only be apprehended historically, that is, through analysis that proceeds backward from the present to identify how phenomena came to be what they are. While Marx is most well-known for his work in economics and specifically his application of this historical method to a critique of capitalism, his methodological contributions amount to no less than a philosophical position for understanding the human world and attempting to harness and direct its ongoing change. The orientation Marx elaborated is referred to as *dialectical materialism*.

Novack (1978) characterizes materialism as the general view that counters earlier claims concerning universality among human beings and instead understands individuals as fundamentally reflecting the cultural and historical environments in which they live. Novack explains that this view was in fact challenged by Marx as overly deterministic, not allowing for human agency, and inadequately accounting for processes of change. To resolve the issue, Marx moved to reconceptualize materialism in light of dialectics (Illyenkov 2012). According to Novack (1971), dialectical logic has its roots in Greek thought, particularly the work of Heraclitus, who postulated that despite the appearance of relative stability, the world exists in a continual state of movement. To apprehend this change, it is important to understand phenomena as processes that exist in relation to one another rather than as discrete entities that can be studied in relative isolation. Dialectical logic thus challenges the tendency in many cultures to define reality according to immediately observable characteristics of phenomena, arriving at easy conceptualizations of A or B. Instead, dialectics compels us to conceive of A in relation to B (Novack 1971). Equally important, Novack explains, is that this relation is neither one of identity (A is the same as B) nor of opposites (A is A and B is B and A can never be B). Instead, the relation is one of A *with* B, raising questions of how A influences B, how changes to B affect A, and what greater unit is comprised of A and B. While dialectics is perhaps most often associated with Hegel's *thesis-antithesis-synthesis* formulation, it is Marx's work that elaborated dialectics for the human and social sciences. Significantly, dialectics enabled Marx to understand human beings as not only able to apprehend processes of change and their role in supporting particular kinds and directions of change but also as uniquely positioned to direct change toward desired ends (Ollman 2003). Indeed, Willis et al. (2007) explain that the starting point of Critical Theory is that individuals and the meanings they carry and embody are socially and historically conditioned, and awareness of these conditions is the necessary prelude to social transformation. To this, Kemmis et al. (2008) add that it is through the activity of changing the world that people realize their humanity, that is, they recognize themselves as not simply objects of history but as makers of history.

In the L2 field, the most in-depth discussion of dialectical materialism to date is provided by Lantolf and Poehner (2014). Those authors situate Vygotsky's Sociocultural Theory within a Marxian philosophy of science, explicating Vygotsky's efforts to establish a scientific psychology following Marx's methodology. We will not repeat their arguments here, and we remind the reader that the focus of our chapter is at the level of philosophy of science rather than concern over theories within a given domain. Indeed, as we point out later in this chapter, some of the contributions to this volume proceed from a Vygotskian theoretical perspective but many do not. Regardless of the theoretical orientation of the authors – and whether this is explicitly stated or merely implied – all share a commitment to praxis. According to Carr and Kemmis (1986), praxis again has its origins in Greek philosophy, and specifically in Aristotle's formulation of the three general categories of human endeavor. These are the theoretical, the productive, and the practical. Of these, the latter two are both concerned with action. The difference between them is that productive activity, for which Aristotle employed the term *poietike*

(making action), is guided by an ideal image (*techne*) that is not changed by the activity. For example, the production of a piece of furniture is guided by an image of what the furniture ought to be like; the laborer produces it, rendering the ideal image material, but the ideal is not changed in any way. The activity is non-reflexive. In contrast, Aristotle employed the term *praxis* (doing action) for the practical domain of activity, and here reflexivity is central. Praxis is dialectical in that the ideal, the images and knowledge base that guide activity are continually revisited and revised, as are specific actions, over the course of the activity. The only fixed, or unchanging, element in praxis according to Aristotle is *phronesis*, the disposition to act truly and rightly.

As Carr and Kemmis (1986) explain, it is the dialectic of knowledge guiding action but also being revised, extended, and enriched through action that marks Critical Theory's major epistemological break from both post-Positivism and Interpretivism. Following conventions from the natural sciences, post-Positivist researchers in the social and human sciences hold action to affect change as only a distal goal, one that can be considered after sufficient research has been conducted that one can act with confidence according to empirically supported theories and models. The potential for the researcher to influence the object of study is treated as a threat to the objectivity and validity of the research in post-Positivism. Moreover, Carr and Kemmis (1986, p. 79) point out that

> …insofar as it studies educational situations in the same way as natural scientists study natural phenomena, [post-Positivist] scientific research inevitably assumes that these situations operate according to a set of 'general laws' that regulate the behaviour of individuals. Furthermore, because these laws are assumed to be independent of the purposes of the individuals whose actions they determine, it follows that the only way to affect practice is by discovering what these 'laws' are and manipulating educational situations accordingly.

The authors continue that inherent in this way of approaching educational research is the problem of generating a base of knowledge that is widely agreed upon in the field and can meaningfully orient practice. In addition, Carr and Kemmis note concern over the role that this approach allows for practitioners (i.e. teachers), namely that they are limited to functioning as technicians armed with general laws to be applied to their specific situations.

With regard to Interpretivism, Critical Theorists also charge that this tradition is limited in its capacity to bring about change. According to Delanty and Strydom (2003), Critical Theorists contend that inquiry that focused exclusively upon identification of the meanings and perspectives of individuals overlooks the reality that the historical and ideological reasons for present circumstances may remain invisible, and so actors are limited in the range of perspectives and meanings that are open to them. Acceptance of circumstances (social systems, resources, divisions of labor, goals, and values) as simply being as they are and not open to question or critique means that the risk in Interpretivist research is that individuals resign themselves to the current social reality and simply alter the discourse they employ for contemplating their actions.

Within Critical Theory, and especially as it has been brought into education, work that follows a praxis orientation and seeks to address difficulties arising from

post-Positivism and Interpretivism is most often referred to as *action research* and sometimes *emancipatory action research* to emphasize the aim of helping individuals move beyond acceptance of present situations to imagine preferred alternatives (Bambino 2002). This latter point is emphasized by Carr and Kemmis (1986, p. 222) in their discussion of the importance schooling as a social institution with particular potential for bringing about change:

> Too often, schools take the structure of society for granted rather than treat it as problematic, even though it is a human and social construction, the product of many decisions and expectations. For schools the accept the assumption that our social structure is 'natural' or 'given' is to rob education of its critical function and to deprive schools of their critical role.

In addition, a hallmark of action research is that the role of the researcher is fundamentally transformed from that of either an impartial scholar documenting perspectives and measuring behaviors or even an authority prescribing models for practitioners to apply. Instead, the researcher functions as a resource, sometimes referred to as a *critical friend* (Bambino 2002), who engages in dialogue with practitioners for the purpose of offering expertise that can be drawn upon as practices are examined, critiqued, and reformulated. As a dialectic, of course, the process continues with additional reflection and dialogue when revised practices have been trialed, and in this way the base of expertise itself continues to be refined. In this way, neither the researcher nor the cooperating practitioner knows in advance precisely where their partnership will lead, but they are guided in their collective action by phronesis, in this case, a shared commitment to continuing to improve educational practices.

As we turn now to the chapters in this volume, we will see that they are in fact quite diverse in several respects. The authors themselves are geographically dispersed, conducting work in Asia, Australia, Europe, New Zealand, North America, South America, and the United Kingdom. The partnerships these researchers describe include pre-service and practicing teachers of English as a second language as well as other commonly taught European languages (French and Spanish), Chinese, Hebrew, and indigenous languages (Mi'gmaq). The educational contexts in which these collaborations have developed also vary, including both novice and very advanced language learners and adult learners in university settings as well as younger learners in primary and secondary schools and an outdoor nature-based environment. As mentioned, some of the contributors to this volume make explicit reference to theoretical perspectives on L2 development, but others do not. Their work makes use of research methods ranging from discourse analysis and think-aloud protocols to score reporting and from narrative inquiry interviews to ethnography. Each chapter, however, maintains a research focus upon enriching the field's frameworks and conceptual models for understanding L2 classroom assessment and pursues this goal through engaging with teachers to help develop assessment practices in a direction that identifies the abilities and needs of all learners and that helps teachers and learners move toward their language goals. Taken together, these chapters showcase several of the possibilities – though certainly not all – that are available to our field through a praxis orientation.

About this Book

While the contributors to this book each share a commitment to praxis, precisely how it is realized differs from one context to another and reflects participants' goals, institutional histories, and available resources. While some studies report of existing collaborations and their aftermath in terms of the reciprocal theory-practice understandings, others lead towards awareness of the necessity and value of praxis for making future theoretical and practical strides. Even though the studies strive for symmetry between practitioners and researchers in their aim to reject the conventional power relations whereby theory dictates practice, such symmetry is not always achieved. In many cases, the initiation for the research is that of the researcher who guides the course of events. However, in all the reported cases the results illustrate the potential and enacted advantages of praxis in gaining better understandings, in enriching theoretical and practical knowledge, and in legitimizing but also critically examining existing worldviews and beliefs. What follows is a brief description of each chapter. The reader will note that the book has been organized into four parts that each highlight different dimensions of the praxis initiatives reported in their respective chapters. We caution that readers should not interpret this organization to imply that these are the only dimensions of praxis on display in a given chapter. In reality, efforts to redefine researcher-teacher roles and to jointly explore responses to issues in theory and practice, for example, are common to each of the chapters. Our division of the chapters reflects only our identification of an important feature of praxis that particular chapters address in interesting ways, and it is purely for the convenience of the reader.

Part I offers three exemplars of praxis on a relatively large scale. In the chapter entitled "'Bringing the teacher back in': Toward L2 assessment praxis in English as an additional language education", Michael Michell and Chris Davison describe the iterative nature of knowledge construction in praxis and how it may be mediated by particular tools or resources. In this case, the mediating tool is an online resource system referred to as *Tools to Enhance Assessment Literacy for Teachers of English as an Additional Language (TEAL) I*. Importantly, TEAL is a toolkit that is at once a product of researcher/teacher cooperation (i.e. it was created through praxis and continues to be updated to reflect ongoing teacher work), as well as a resource that teachers can turn to for professional development. The system was developed and validated collaboratively as part of the reconceptualising of the assessment of English Additional Language learners (EAL) in Australia, with the teachers gaining practical informed assessment pedagogy in the process. As Michell and Davison explain, the teachers are regarded as key players in the assessment act. Guided by the kit they utilize their professional expertise in situated tool-mediated activity that takes place within a knowledge creation system, one that involves on-going change stimulated by reflective practices. In addition to assessing EAL development TEAL is also a formative tool that guides development in the different skills of L2 classroom assessment literacy and praxis.

In Angela Scariono's chapter, "Mediation in the assessment of language learning within an interlingual and intercultural orientation: the role of reciprocal interpretation", we learn of a 3-year collaborative phenomenographic case study, conducted in Australia, with researchers and teachers of different languages working in cooperation with one another. Scarino provides background to the conceptual complexity of interculturality, and hence to the particular considerations involved in the reciprocal interpretative meaning-making of exchanges in interlingual and intercultural settings. For Scarino, this requires not only a re-thinking of L2 curricula but also the adoption of situated assessment approaches which both reveal and promote interlingual and intercultural meaning-making. Scarino documents an ongoing cyclic interplay of various project phases, including the theorizing and planning phases during which assessment was reconceptualized in an interlingual and intercultural framing to the process of implementation, feedback, and validation. Findings reinforce the need for collaborative dialogue and for teachers and researchers to move towards a dialectical relationship between planning, action, and reflection, underscoring the situated nature of praxis.

The third chapter in Part I documents praxis in an especially high-stakes assessment context. In their chapter entitled "Trajectories of language assessment literacy in a teacher-researcher partnership: Locating elements of praxis through narrative inquiry", Luke Harding and Tineke Brunfaut describe an on-going collaboration between themselves as language testing experts and teaching/assessment practitioners from Luxemburg involved in redesigning a national end-of-secondary-school English exam. Harding and Brunfaut share an analysis of a set of narratives produced, respectively, by the two researchers and by two of the teachers/emerging language testing experts from the Luxembourgian team. Analysis of the narratives is informed by Pavlenko's (2007) framework that includes the interplay between content, context and form. The narratives bring to light the essentially fertile process of the collaboration in terms of knowledge building in the area of language assessment literacy for the teachers. However, the narratives also aid in surfacing areas for improvement, difficulties experienced by team members, and missing 'pieces of the puzzle', specifically the knowledge required for political engagement that includes communication with stakeholders empowered to propose and implement policy, which Harding and Brunfaut submit may ultimately decide the success and sustainability of the project.

Together, the projects reported in these three chapters, each of which is established but also ongoing, provide a solid frame of reference for appreciating some of the forms L2 assessment praxis may take. With that as background, the remaining parts of the book each bring into focus particular aspects of praxis work and where it may lead L2 assessment researchers and practitioners. Part II concerns specifically the nature of the relation between researchers and teachers in praxis and the importance of upending traditional hierarchies that position teachers, for example, as mere 'consumers' of research whose responsibility is limited to 'implementing' theory. Working in a context of teacher education (an MA TESOL program in the UK), Constant Leung reports a phenomenological investigation into teacher candidates' experiences receiving and interpreting written feedback. Leung's study is an

initial stage of a larger praxis project concerned with elaborating formative assessment practices in the MA program as well as mediating teacher candidates' assessment experiences and, ultimately, their understanding of the role of assessment in supporting teaching and learning. Leung's chapter, entitled "Learning-Oriented Assessment: More Than The Chalkface", probes complexities in the Learning Oriented Assessment (LOA) framework and identifies the need for ongoing collaborative dialogue among teachers and learners, a dialogue that considers curricular, analytic, and conceptual perspectives. Leung argues that such a situated approach to assessment fosters theoretical understanding of formative assessment processes while simultaneously serving the purpose of illuminating student learning, including how their learning interacts with cultural background and affective factors.

The central place in praxis that must be afforded to teacher experiences, expertise, and goals for re-envisioning practice moves to the forefront in the next chapter, which reports a project jointly conducted by a language assessment researcher (Beverly Baker) and an indigenous language teacher (Joyce Germain) in the unique setting of a Mi'gmaq language immersion classroom in Listuguj, Canada. Baker and Germain's chapter, "Narrative inquiry as praxis: Examining formative assessment practices in a nature-based Indigenous language classroom", probes the range of assessment practices employed by the teacher and their role in supporting learning within a discovery-based experiential learning setting. This includes a focus on formal and informal assessment episodes in-class and outdoors. Through narrative inquiry and collaborative dialogue both participants (i.e. researcher and teacher) were able to capture and reflect on formal and informal, in-class and outdoors 'assessment moments'. What clearly emerges is the benefits of the joint collaborative inquiry which allows for different perspectives as well as the significance of harmonious relations of mutual trust and respect between the two partners. Following Baker and Germain, this open flow of knowledge takes on even greater importance in the context of Indigenous language instruction, as attention must also be given to dynamics between assessment/pedagogy and cultural values and practices.

The relevance of local culture is a theme reprised in the chapter "Learning from each other: School-university collaborative action research as praxis" by Cheri Chan and Chris Davison. Chan and Davison take us to Hong Kong to consider how socio-political and socio-cultural discourses and norms apply to research-teaching praxis following a large-scale assessment reform initiative. The authors report an action research study focusing on English language teachers in three Hong Kong secondary schools with the aim of improving strategies for providing formative feedback on oral assessment tasks. Discourse and narrative analysis showed that the teachers were able to adapt and modify their approach to feedback to more closely resemble assessment for learning strategies, at the same time continuing to meet the expectations of their own schools and students. Chan and Davison approach their project as a two-way process comprising, on the one hand, the input, guidance and support offered by the university team to the teachers to help them understand and apply the reform, and on the other hand the reality of the teachers functioning as co-inquirers who contributed to improving the feedback and assessment for learning

models. The authors' discussion raises possibilities for re-envisioning teachers as researchers.

The chapters in Part III detail uses of praxis to explore potential responses to problems in theory/research or practice or both. In the chapter "Advancing written feedback practice through a teacher-researcher collaboration in a university Spanish program", Kathryn Hill and Ana Maria Ducasse began their inquiry from a shared interest in written corrective feedback, Hill from the perspective of how it can provide insights into student learning as part of assessment while for Ducasse the question was how well her current feedback practices aligned with her instructional goals for learners. Their partnership made use of Ducasse's written feedback to students as well as recordings and transcriptions of unstructured think-aloud protocols, semi-structured interviews and reflective journals. Through a process of probing questions reflective analysis, they determined that Ducasse had been unconsiously applying an external standard for grading the class assignment, hence emphasizing accuracy beyond the program requirements. Relevant literature was identified by Hill to help Ducasse reinforce practices that resonated more strongly with her overall pedagogical orientation (e.g., progressive feedback). While this collaborative inquiry also informed Hill's conceptual understanding of classroom assessment (specifically, the impact of contextual factors on a proposed broader framework), the authors noted a certain asymmetry whereby the researcher "had essentially retained primary control of the shape of the collaboration", an insight that has prompted them to aim for a more equal relationship as the project progresses.

The chapter "Reconsidering time and process in L2 Dynamic Assessment" by Matthew E. Poehner & Rémi A. van Compernolle takes Dynamic Assessment (DA), a specific framework for integrating teaching as part of assessment, and considers its usefulness in the context of a university L2 French program. Specifically, Poehner and van Compernolle are responding to a need within that program to better identify learners' readiness to progress from basic language instruction to more advanced courses concerned with literature and culture studies. Initially intended to provide diagnostic information regarding learners' L2 abilities to teachers and program administrators, piloting of the DA procedure pointed to the importance of refining approaches to capturing and representing learner performance during DA. For Poehner and van Compernolle, the project ultimately served a threefold purpose of trialing the DA procedure that will eventually inform the design of a computerized assessment tool, improving assessment reporting procedures, and bringing the curricular foci of beginning level French courses in line with the expectations for learners in higher level courses.

In some respects, the focus on using DA to address a problem of practice gets inverted in the chapter "Reconceptualizing classroom dynamic assessment: Lessons from teacher practice" by Kristin J. Davin and José David Herazo. The question motivating these authors was how teachers integrated DA as a conceptual framework with their pedagogical practices following a professional development workshop on that topic. Davin and Herazo, who also led the DA training, revisited three of the participating teachers 2 years after the workshop to see what insight can be gained into how the teachers developed their understanding and use of DA principles

and frameworks. Their findings point to the situated nature of the teachers' practices and to the importance of understanding local dynamics as influencing what they refer to as a 'recontextualization' of conceptual frameworks (in this case, DA recontextualized into a theory of teachers' own practices). Davin and Herazo explain that teachers found creative ways to overcome some of the challenges in implementing DA (e.g., class size and time resources), but that they also went through meaningful changes with regard to defining the purpose of their assessments and the time they allocated to assessment as well as to specific topics in the curriculum.

Finally, Part IV of the book brings to the fore ways in which praxis initiatives may offer opportunities to transform ways of thinking about and engaging with assessment more broadly. In his chapter "Addressing the possibilities and limitations of implementing a new classroom-based assessment of oral proficiency", Martin East follows a French teacher's enactment of a new assessment standard, *Interact,* as part of curricular reforms in New Zealand that include an emphasis on assessment for learning. The new standard stipulates the need to collect evidence of peer-to- peer interaction in French, in line with assessment for learning and authentic communicative language use principles. Through individual and focus group interviews, East captures teacher and student perspectives on the introduction of the standard, noting in particular the view that "*Interact* has taken the students 'from French *learners* to French *speakers.*" East explains that the teacher's previous participation in a theory- and research-informed pre-service teacher education programme is believed to partially account for his successful ability to mitigate some of the accountability constraints and to reflect on his practice. Nonetheless, despite the teacher's efforts to lessen anxiety and formatively cater to the students' needs in introducing the new requirements, the formal assessment aspect still hovers in the background.

The chapter by Ofra Inbar-Lourie and Tziona Levi, entitled "Assessment literacy as praxis: Mediating teacher knowledge of assessment-for-learning practices", also looked at praxis as part of assessment reform but from the perspective of school culture and its alignment with broader organizational beliefs and assumptions that guide assessment practices. The research looked at the degree to which assessment for learning culture was processed and integrated in schools with differential school cultures following a professional development course on assessment, introduced as part of national assessment reforms and administered by one of the researchers. After a 6-month period school administrators and language teachers who had participated in the assessment course were interviewed to gauge whether and to what degree the language teaching community was engaged with the assessment themes acquired. Findings show that schools with a progressive school culture which fostered autonomy, teacher collaboration, innovation and decentralization, tended to adopt the assessment for learning conceptual framing and assessment practices more readily than their more centralized and traditional counterparts. The themes overall were linked to four perspectives or domains identified previously: the cultural, technical, post-modern and political. Though most of the respondents seemed to agree with assessment for learning culture encouraged in the course, interviewees in all schools regardless of their orientation expressed concern over the dominance

of the external matriculation exam. Similar to the previously reported chapters, here, too, interactions between the field and the assessment experts facilitated insights into practice-bound theory trajectories. This was especially true with regard to teacher development of language assessment literacy.

As will be clear, there is no single prescription for 'doing' praxis but rather a willingness to engage in and with practice and practitioners as an integral part of research. This engagement is defined by mutual trust and by an interest in critically examining what *is* as part of determining what *might be*. Put another way, researcher-teacher efforts to affect change will certainly build upon what currently exists, and so continuities are to be expected, but a process of critical examination also means that disruption and discontinuity are necessary as well. This process of course continues as researchers and teachers evaluate where their efforts have taken them and new problems that may have arisen, as well as changing conditions that require them to focus on other issues and possibilities. To be sure, there is much work to do, and it can certainly appear daunting. However, as the chapters in this book attest, the potential to learn and improve through praxis is considerable.

References

Assessment Reform Group. (2002). Assessment for learning: 10 principles. *Research-based principles to guide classroom practice.*

Bambino, D. (2002). Critical friends. *Educational Leadership, 59*(6), 25–28.

Berlin, I. (2013). *Three critics of the enlightenment: Vico, Hamann, and Herder*. Princeton: Princeton University Press.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21*(1), 5.

Boyles, D. R. (2006). Dewey's epistemology: An argument for warranted assertions, knowing, and meaningful classroom practice. *Educational Theory, 56*(1), 57–68.

Carr, W., & Kemmis, S. (1986). *Becoming critical*. New York: Routledge.

Delanty, G., & Strydom, P. (2003). *Philosophies of social science. The classic and contemporary readings*. Philadelphia: Open University Press.

Garratt, D., & Li, Y. (2005). The foundations of experimental/empirical research methods. *Research Methods, 198*.

Genesse, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.

Harding, L., Pill, J., & Ryan, K. (2011). Assessor decision making while marking a note-taking listening test: The case of the OET. *Language Assessment Quarterly, 8*(2), 108–126.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy, & Practice, 4*(3), 365–379.

Illyenkov, E. (2012). Dialectics of the ideal (2009). *Historical Materialism, 20*, 149–193.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment Encyclopedia of language and education* (3rd ed., pp. 257–270). Berlin: Springer.

Kemmis, S., Mattsson, M. A. T. T. S., Ponte, P. E. T. R. A., & Rönnerman, K. A. R. I. N. (2008). Series introduction. In *Enabling praxis: Challenges for education*. Berlin: Springer Publishing.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education. Vygotskian praxis and the research/practice divide*. London: Routledge.
Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351–372.
Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*(4), 13–25.
Novack, G. (1971). *An introduction to the logic of Marxism*. Atlanta: Pathfinder Press.
Novack, G. (1978). *Polemics in Marxist philosophy. Essays on Sartre, Plekanov, Lukacs, Engels, Kolakowsi, Trotsky, Timpanaro, Colletti*. New York: Pathfinder.
Ollman, B. (2003). *Dance of the dialectic: Steps in Marx's method*. Urbana-Champaign: University of Illinois Press.
Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics, 28*(2), 163–188.
Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Mahwah: Lawrence Erlbaum.
Peters, M., & Burbules, N. (2004). *Poststructuralism and educational research*. Lanham: Rowman and Littlefield.
Phillips, D. C., & Burbules, N. (2000). *Postpositivism and educational research*. Lanham: Rowman and Littlefield.
Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer.
Poehner, M. E. (2011). Validity and interaction in the ZPD: Interpreting learner development through L2 dynamic assessment. *International Journal of Applied Linguistics, 21*(2), 244–263.
Rea-Dickins, P. (2001). Mirror mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.
Rorty, R. (2009). *Philosophy and the mirror of nature* (2nd ed.). Princeton: Princeton University Press.
Sawaki, Y., & Koizumi, R. (2017). Providing test performance feedback that bridges assessment and instruction: The case of two standardized English language tests in Japan. *Language Assessment Quarterly, 14*(3), 234–256.
Scarino, A. (2013). Language assessment literacy as self-awareness: *Understanding* the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309–327.
Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing, 17*(2), 163–184.
Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in the classroom. In *Handbook of second language assessment*. Berlin: DeGruyter Mouton.
van Manen, M. (1997). *Researching lived experience* (2nd ed.). London: The Althouse Press.
Willis, J. W., Jost, M., & Nilakanta, R. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Thousand Oaks: Sage.

Part I
Realizing Praxis: Exemplars of Researcher-Teacher Partnerships

Chapter 2
'Bringing the Teacher Back In': Toward L2 Assessment Praxis in English as an Additional Language Education

Michael Michell and Chris Davison

Abstract School assessment systems in Anglophone countries typically reflect monolingual 'native speaker' norms of language and literacy development and fail to capture the distinct pathways and milestones of language minority students' social and academic English language learning in the curriculum. Such systems limit teachers' access to appropriate and useful assessment tools and advice, devaluing their role in students' assessment based on a contextualised and detailed understanding of their language learning needs. This chapter 'brings the teacher back in' to assessment by reconceptualising L2 classroom assessment as tool-mediated assessment praxis (assessment-focused teacher knowledge practice) developing within a teacher community of assessment practice. The nature and role of the Tools to Enhance Assessment Literacy for Teachers of English as an Additional Language (TEAL), an online resource system developed to help Australian primary and secondary teachers assess the stage of development for EAL students in speaking listening, reading and writing, and improve learning and teaching, is then considered within this social system of developing assessment praxis. The Assessment for Learning (AfL) and Vygotskian learning theory design of the online 'toolkit' is highlighted. By reconceptualising L2 classroom assessment as language assessment praxis, the chapter supports a view of teachers of EAL learners as prime assessment agents, re-imagining and revaluing them as trustworthy language assessment practitioners.

M. Michell · C. Davison (✉)
School of Education, Faculty of Arts and Social Sciences, University of New South Wales, Sydney, Australia
e-mail: m.michell@unsw.edu.au; c.davison@unsw.edu.au

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_2

Introduction

Assessing English as a second or an additional language (EAL) learners in schools is a particularly challenging area for teachers. These students require specific assessment tools and frameworks that capture the distinct pathways and milestones of their social and academic English language learning in the curriculum in ways that can inform systematic language-based teaching support. Most school assessment systems in Anglophone countries, however, are predicated on monolingual 'native speaker' norms of literacy development and imposed on teachers of EAL learners, rather than negotiated with them based on a contextualised and detailed understanding of their needs. Assessment systems developed with the English learning needs of these students in mind and the teacher practices that surround them are therefore key sites for understanding and conceptualising L2 classroom assessment praxis that underpins these systems.

As Poehner and Inbar-Lourie (Chap. 1) explain in their introduction to this volume, the rich concept of *praxis* – which positions theory and research in a relation with practice such that they mutually inform one another – has attracted attention among scholars working in the human and social sciences, and it is especially productive when applied to teachers and teaching. Following Poehner and Inbar-Lourie, praxis offers an alternative to views that understand knowledge production to occur through theory and research and only later to be applied to practice. In the field of education, this arrangement removes teachers from the activity of knowledge generation and situates them instead as technicians left to implement generalized research findings in their local contexts. Praxis, in contrast, 'brings the teacher back into' the heart of the process of knowing and doing. With its moral and 'pedagogical imperative' (Lantolf and Poehner 2014) underpinnings, the concept of teacher praxis also offers a timely response to concerns about the bureaucratisation and deprofessionalisation of education that 'seem to be eroding the moral, social and political commitments that have always informed pedagogical practice' (Smith et al. 2010, p.1). Moves towards reconceptualising EAL assessment as an exemplar of L2 classroom praxis therefore stands to revitalise thinking about education, pedagogy and assessment and offer teachers of EAL students renewed scope of action.

This chapter offers a conceptual examination of praxis and its application to teaching in general and EAL assessment in particular. As such, the present chapter differs from others in this volume in that it does not report data from a project undertaken by researchers and teachers and that is presented as illustrative of L2 assessment praxis. Rather, we seek to elaborate discussions of praxis in the educational research literature by bringing them into contact with Vygotskian approaches to psychological development. This area of research has also been proposed by Lantolf and Poehner (2014), although our aim differs in that we are specifically concerned with reconceptualising EAL assessment as a tool-mediated assessment praxis. Our theoretical discussion is followed by presentation of a specifically designed EAL assessment system, which we refer to as the 'TEAL toolkit'. The toolkit represents the result of praxis within a particular community of researchers and teachers

engaging around assessment issues, while at the same time it serves as a resource to promote and sustain praxis in the future. In this context, reconceptualising EAL assessment as L2 classroom assessment praxis brings teachers of EAL learners back in as prime assessment agents, re-imagining and revaluing them as trustworthy language assessment practitioners.

What Is Teacher Praxis?

An adequate understanding of the concept of *praxis* requires understanding its meanings from within two philosophical traditions – its original meaning in Aristotelian philosophy and its further development in Marxist thought. Today, (neo) Aristotelian notions of praxis emphasising the individual practice of ethical deliberation and Marxian praxis emphasising the collective practice of social transformation are elaborated in Anglo-phone and European countries respectively (Kemmis 2012a). Bringing together Aristotelian understandings of praxis as 'morally committed action' with its accompanying disposition, *phronesis* (practical wisdom), and Marxian praxis as 'world-changing action,' Kemmis and colleagues (Kemmis and Smith 2008; Kemmis et al. 2014) seek to reclaim ethical agency within the larger moral purpose of education as the project of helping 'people live well in a world worth living in' (Kemmis and Smith 2008). In doing so, they regard both these personal and collective meanings of praxis

> as dialectically related to one another, that is, as mutually constituted in that the personal praxis of the individual expressed in moral intentions, agency and capabilities is not something formed entirely by individuals on their own. Rather, their praxis is also shaped, formed, enabled and constrained collectively through historically formed and transformed cultural-discursive, material-economic and social arrangements (practice architectures) which in turn pre-form and prefigure the possibilities of personal praxis. (Smith et al. 2010, p.5)

For its part, Aristotelian *praxis* describes the morally informed but uncertain, deliberative actions of practical reasoning (*phronesis*) directed towards 'right action' in contrast to the defined, instrumental, means-ends rationality (*techne*) that characterises craft activity (*poietike*). From this perspective, teacher praxis is therefore viewed as situated ethical knowledge work conducted in particular teaching contexts within a practice tradition (Kemmis 2012a, b). Arising from this moral purpose of reclaiming the ethical agency of teaching is a renewed education research agenda focused on praxis (Kemmis and Smith 2008; Edwards-Groves and Kemmis 2016) and praxis development (Kemmis et al. 2014), and the ecological arrangement of teaching practices – 'practice ecologies' (Kemmis 2010) and 'practice architectures' (Edwards-Groves and Grootenboer 2015). Praxis has been used as a lens to understand the 'sayings', 'doings' and 'relatings' inherent in the teaching of mathematics (Grootenboer and Edwards-Groves 2013, 2014) and English (Edwards-Groves and Grootenboer 2015) as well as their implications for developing students' mathematics identities (Grootenboer 2013). This holistic, *phronesis*-based

ontology of praxis has particular appeal to teachers as a way of describing, recognising and legitimizing their tacit knowledge practice in contrast to 'theory' (Wubbels et al. 1997) and as an alternative to dominant technical, instrumental, approaches to education.

The hallmark of the Marxian tradition is its concept of praxis as an emancipatory project of social transformation, 'the action and reflection of men [sic] upon the world in order to transform it' (Freire 1970, p.66). Indeed, Gramsci's reinterpretation of Marxism as the 'philosophy of praxis' places the concept of praxis at the very heart of the Marxian project (Thomas 2015). Both Aristotelian and Marxian traditions together inform a notion of praxis as self-formation:

> …praxis as 'right conduct' and as 'history-making action' is inseparable from the person or persons performing it; it is always a process of *self-formation*. It might be added that praxis is also a process of self-formation in both an individual and collective sense – praxis forms the person, the identity, of the ones who act and the communities of which they are a part, and these persons and communities are, as Marx observes, both products and producers of history (Kemmis 2010, p.21, original italics)

A Marxian view of praxis is outlined by the Russian psychologist L. S. Vygotsky (1926/2006) as he emphasises the centrality of knowledge in 'world-changing action' and its origins and validation in practical activity:

> All knowledge ultimately, has arisen, and will arise always, out of some sort of practical need or necessity, and if, in the course of the development of knowledge, it loses touch with the practical problems that give rise to it, at its final points of its development it will again be oriented towards praxis and find in praxis its ultimate justification, confirmation, and verification (Vygotsky 1926/2006, p. 200).

This view is consistent with Vygotsky's better known articulation of the theory-practice relationship in his work, *The historical meaning of the crisis in psychology* (Vygotsky 1997a). There, Vygotsky in fact echoes Engel's mundane understanding of Marxian praxis, an application-verification understanding that Bottomore (1983, p. 387) characterizes as the 'proof of the pudding is in the eating'. Specifically, Vygotsky explains

> Practice pervades the deepest foundations of the scientific operation and reforms it from beginning to end. Practice sets the tasks and serves as the supreme judge of theory, as its truth criterion. It dictates how to construct the concepts and how to formulate the laws (Vygotsky 1997a, p. 305–306).

Vygotsky's depiction of praxis in this statement, however, hardly does justice to the research practice he pursued in his cultural-historical project of 1926–1934. His research praxis during these years is better described as 'the dialectical unity of methodology and practice' (Vygotsky 1997a, p. 310) where research theory, i.e. scientific knowledge as appropriated and created conceptual tools, is mediated and shaped by a Marxian, goal-directed philosophy of practice:

> "Method" means "way", we view it as a means of knowledge acquisition. But in all its points the way is determined by the goal to which it leads. That is why practice reforms the whole methodology of the science (Vygotsky 1997a, p. 306).

Such praxis is evident in development of his 'philosophical method' in which theory and method are integrated. This integration is especially notably in the double stimulation method that underpinned Vygotsky's early psychological experiments (Engestrom 2011; Sannino 2015; Sannino and Laitinen 2015; Vygotsky 1997a, b, c, 1999a, b, c). Driving Vygotky's praxis and made explicit in his writings about the Marxist goals of his psychology project (Vygotky 1994) is the moral purpose and motive of his research activity and the emancipist values that underpin and drive it. Key characteristics of a Vygotskian notion of praxis can therefore be described as epistemic, instrumental cultural-historical, and ethical, and these reflect the Marxian view of praxis as conscious, emancipatory, world-changing, 'history-making action'.

Vygotskian research praxis offers a model for teacher praxis that highlights the specific role that 'scientific', theoretical knowledge and reflection (Karpov and Bransford 1995; Vygotsky 1987) and social and material processes play in its formation. It draws attention to the role of conscious knowledge and knowledge production in practices and delineates teacher praxis as knowledge or knowledgeable practice, or what Kemmis terms 'knowing doing' (Kemmis 2010, p.9). A similar knowledge-driven view of praxis is evident in Park et al.'s (2010) research on the pedagogical development of novice science teachers where they argue for a conception-led notion of teacher praxis as practice that results from interaction between teacher conceptions (of their subject discipline, learning and learners) and their specific environmental constraints, ideally leading towards their integration in teaching practice:

> …conceptions and praxis are reciprocally related, with each in turn influencing the other. Because, however, teaching ultimately happens within school and society, there are environmental constraints on the interactions between conceptions and practice. This means that teachers' conceptions are not necessarily consistent with their praxes. From this perspective, teachers would ideally learn to teach by addressing all three of these aspects – conceptions, praxis and environment – and the complex ways in which they interact with each other (Park et al., p.719).

From this perspective then, teacher praxis may be understood as the intentional pursuit of the alignment or integration of pedagogical thinking and practice in specific teaching and learning situations.

Research drawing on knowledge-building and knowledge-creating classrooms (Scardamalia and Bereiter 1999, 2003) also provides a useful way of conceptualising knowledge-oriented nature of teacher praxis. With its focus on developing enquiry-based pedagogies that foster students' epistemic agency, this literature also highlights key elements of epistemic practice for *teachers as learners*: namely, the identification and exploration of ill-defined 'problem spaces' within a teacher community (Teo 2014), the role of Pedagogical Content Knowledge in the development of new pedagogic practices (Law 2014) and the development of transformational, tool-mediated pedagogic activity through design-based research (Yeo 2014). This literature emphasises teacher praxis as 'educational knowledge work' or 'knowledge practice' reflecting the practical, theory-building focus of Vygotskian research praxis. Here, teacher praxis is viewed as 'routine personal and social activities

related to working with knowledge [and representing] deliberate efforts to expand one's intellectual resources by creating and building epistemic artifacts' (Ritella and Hakkarainen 2012, p.240).

The knowledge-oriented nature of teacher praxis draws attention to the key role of theoretical knowledge and reflection in and on action. Ertsas and Irgens (2017) offer a comprehensive model of professional theory-making which elucidates the nature and influence of theory in teacher praxis. Their graded theory model outlines three degrees of knowledge that avoids the problematic theory-practice dichotomy. These are: the first degree (T1) of 'hidden', non-articulated, tacit, 'theory-in-use' (Argyris and Schön 1978) implicit in teacher practice; the second degree (T2) of articulated, 'espoused theory' (Argyris and Schön 1978) about teaching practice; and the third degree (T3) of generic, well-articulated theory having meta-theoretical and reflective functions. 'Professional theorizing' describes the process by which the second degree draws on and overcomes the myopia and context restrictions of primary experience (T1) and applies the expansive meta-cognitive resources of theory (T3). This epistemic process describes the mechanism of teacher learning and development as a dialectic interaction between theory and practice:

> a teacher must develop his/her T2 by drawing on T1 as well as T3. Neither practice nor theory should be allowed primacy, rather attention should be placed on the theorizing process that interrelate different grades of theory and construct professional awareness and knowledge. This we propose is a way of developing professional judgement: a critical assessment, not only of one's own practice, but also of theories of the third degree (Ertsas and Irgens, p.347).

This model sheds light on the meta-cognitive nature of teacher reflection in and on action that underpins teacher praxis. Here, professional reflection is the activity of T2 that appraises the lived experience of T1 using the transformational knowledge resources of T3. Reflection then is 'a metacognitive mechanism that teachers can use to regulate their own practice, before during and after teaching' (Hoffman-Kipp et al. 2003, p. 251) and the practice of reflection is itself a tool within the larger praxis of teacher learning and development. This model contrasts with the *phronesis*-based practice epistemology proposed by Korthagen and colleagues (Kessels and Korthagen 1996; Korthagen et al. 2001; Wubbels et al. 1997) which views teacher learning as a 'bottom-up' process emerging from classroom-based gestalts to schemas to theory (Korthagen and Kessels 1999; Korthagen et al. 2001). The model attempts to offer a solution to the ubiquitous theory-practice gap in education (Kessels and Korthagen, 1996; Korthagen and Kessels 1999) by devaluing the role and value of external theoretical knowledge and valorising 'useful' theory developed by and for teachers themselves.

Different understandings of teacher praxis reflect varied responses to the theory-practice gap. What might be called theory-to-practice and practice-to-theory perspectives are forms of educational praxis that propound the integration of theory and practice. In this context, Tsui (2009) outlines a model of teacher expertise development that bridges the theory-practice gap through a two-way process of 'practicalising theory and theorising practice' which has been taken up and applied in AfL implementation research (Tang 2010). In her model of teacher-researcher

collaboration, Gade seeks to bridge the theory-practice divide by implementing 'close to practice' research (Edwards et al. 2005, p.123) that promotes 'steered practice'(Gade 2012) through both 'theory-which-informs and theory-being-built' (Gade 2015b, p. 605). In the area of initial teacher education, argue that this divide can be bridged through university-school partnerships as 'communities of praxis' which 'applies theory to practical situations' (p.363) where 'the unrecognised work of teacher educators is the development of pedagogies, models and structures that innovatively close the gap between practice and theory, making praxis possible' (p.362).

Research in the Vygotskian theoretical tradition supports a holistic view of teacher praxis as situated, tool-mediated activity within a knowledge-creating activity system. Applied to developing knowledge practices in workplace settings, such research focuses on the social material processes involved in the formation and development of professional expertise (Fenwick et al. 2012; Mäkitalo 2012; Miettinen and Virkkunen 2005). Taking pedagogic activity as the unit of analysis, sociocultural education research highlights the conscious goal-oriented development of teacher practice directed towards an 'epistemic object' (Lantolf and Poehner 2014; Lee 2014). Such knowledge objects of teaching activity are formed, pursued and transformed as part of a dynamic process of expanded learning which is both mediated by, and an outcome of, sociomaterial resources (Engeström 2011; Lee 2014). This process describes the professional learning that accompanies the sustained, systematic pursuit of situated pedagogical problems.

Foregrounded in this research is the knowledge-carrying nature and function of mediating tools (Mäkitalo 2012) in creating knowledge-action affordances and knowledge transformation through their deliberate incorporation within the activity system. Mäkitalo (2012) highlights this process as one of *explicit* mediation (Wertsch 2007), where 'the material-semiotic tool becomes the object to which the participants orient themselves [organising] their perception and action in the ongoing activity' (p.62). The practice of explicit mediation furthers understandings of teacher knowledge praxis as much more than the unthinking appropriation and adaption of socio-material resources in the classroom. It points to the intentional, sustained and agentive transformation of artifacts into instruments of teacher activity (Spicer 2011), a process Ritella and Hakkarainen (2012) call 'instrumentization', or 'the emergence and evolution of the artefact to support activity in a local cultural context … adapted to local needs and purposes of the activity (p. 240). This process of instrumentization is described as follows:

> technological artifacts become instruments of human activity and sustained iterative efforts of using them in practice, a process through which the cognitive-cultural operating system of activity gradually transforms and adapts according to evolving practices of using technologies. This evolution is reflected in deep level changes in mental processes, such as new capabilities being, in effect, cognitive prostheses adapted to changed modes of learning and creating knowledge. (p.248)

In classrooms, especially second language classrooms, the sociomaterial resource for learning, the mediating tool, *par excellence,* is classroom discourse (Gibbons 2006, 2007).

From the perspective of the evolving sociomaterial resources themselves, the evolution of tools in activity over time is described as 'instrumental genesis' (p.246). Such praxis can thus be characterised as 'knowledge-creating inquiry' which 'is mediated by deliberate construction of epistemic artefacts that crystallise the participants' intellectual processes, the evolving network guides subsequent participants' inquiry efforts' (p. 251).

Vygotskian-informed research also offers a situated, systemic perspective on the *social relations* of teacher praxis and its development. This research draws attention to the social dynamics of knowledge praxis through different forms of collective knowledge-sharing and production in a teaching activity system. It is these social relations surrounding the tools, and not the tools in themselves, that are transformational for action and cognition (Hakkarainen 2009). This social perspective is reflected in different forms of teacher collaboration that range from research partnerships between researchers and teachers to joint classroom activity among teachers. Kemmis (2012a, b) identifies teacher praxis (both emic and etic perspectives) as the object of study in teacher-researcher action research. Gade (2015a) emphasises teacher praxis as the expansive learning that occurs through teacher-researcher participation and transformation of the teaching activity system. Roth and Tobin (2004) highlight the epistemic and transformative nature of teacher praxis through teacher engagement in co-teaching and cogenerative dialogue. Similarly, Martin-Beltran and Peercy (2012, 2014) highlight joint tool-mediated praxis by class and specialist teachers as they go about using and developing organisational and pedagogical tools to sustain collaborative communication and create a language-inclusive classroom. Each of these collaborations exemplifies knowledge-creating, classroom-based communities of collaborative praxis.

In these contexts, sociocultural research pays particular attention to the key mediating and epistemic role of teacher dialogue. For example, Roth and Tobin highlight the critical role of to what they call 'cogenerative dialogue', teacher talk about actions taken in teaching. Such talk about actions (praxis) is 'both context specific, first person and put forward as an alternative to the abstract theory which is then supposed to apply in any context' (Gade 2015b, p. 622). Extending Bereiter and Scardamalia's work, van Aalst (2009) identifies three modes of classroom learning discourse which may be readily applied to teacher epistemic praxis - knowledge-sharing discourse, knowledge-construction discourse and knowledge-creation discourse, reflecting knowledge transmission, constructivist and social constructivist epistemologies respectively. Alternatively, narrative understanding of teaching practice as 'personal practical knowledge' is emphasised by Clandinin and Connelly (1998). Here narrative 'grasps not only past, present and future scenarios, but also the personal, professional and practical nature of teachers' experiences (Gade 2015b, p.607) Narrative itself is used as a unit of analysis for understanding teaching praxis and practitioner action in Gade's (2011, 2012) 'close to practice' model of teacher-researcher collaboration.

Drawing on these strands of research, the sociocultural nature of educational praxis can be characterised as epistemic (knowledgeable and knowledge-producing practice) through both goal- and tool-oriented activity; collective (small and large collaborative communities of practice) and cultural-historical (developing practice within practice traditions). The term 'teacher funds of knowledge' (Moll et al. 1992; Hedges 2012) describes the pedagogic knowledge resources that accrue from the interaction of epistemic ends and means in teacher praxis.

Developing Language Assessment Praxis

Having outlined the nature and focus of teacher praxis, it remains to consider its assessment component – teacher assessment praxis. A starting point would be to say that teacher assessment praxis is nothing other than the mobilisation of the social, mediational and epistemic processes of teaching praxis described above directed towards assessment activity. To be sure, this recognizes the need for teacher assessment literacy (see Chap. 4 by Harding and Brunfaut and Chap. 12 by Inbar-Lourie and Levi, this volume), but it also goes well beyond that, emphasizing not merely teacher understanding of externally constructed principles and frameworks of assessment but their bringing these understandings into contact with their situated practices. In the assessment praxis that concerns us here, namely the development of Assessment for Learning (AfL) practice, this praxis, both in its object and moral purpose, is directed towards ensuring that assessment activity serves the requirements of learning (Assessment Reform Group 1999; Black and Wiliam 1998; Black et al. 2007). Assessment praxis then is the active process by which teachers develop a practical but coherent 'assessment pedagogy' (Fleer 2015). Developing a language assessment praxis based on AfL principles addressing the language learning needs of English as additional language learners is a further response to an additional 'pedagogical imperative' (Lantolf and Poehner 2014).

Language assessment praxis is therefore inextricably tied to the AfL project - the putting into practice AfL theory as articulated in its assessment reform rationales, ideals, goals, principles and strategies (Assessment Reform Group 1999; Black and Wiliam 1998, 2005a, 2009). The success or otherwise of this project is seen in the accumulated funds of knowledge practice arising from AfL implementation studies. Overall, these studies have highlighted that the translation of AfL theory into practice is by no means straightforward. Research has shown the explicit application of AfL objectives and strategies can narrow curriculum and learning (Davies and Ecclestone 2008; Hume and Coll 2009; Torrance 2007); effective AfL practice reflects the 'spirit' rather than the 'letter' of implementation (Dixon et al. 2011; Marshall and Drummond 2006); and changes to assessment practice are fundamentally constrained by the organisational and pedagogic structures in which they are embedded (Black and Wiliam 2005b; Davison 2007, 2013; Hill 2011; Webb and Jones 2009). 'Deep' changes to assessment practices therefore require deep changes to school structures and classroom pedagogy, highlighting that a key task of any

praxis is to engage and overcome the constraints of the given teaching environment by transforming it. In this context, the AfL project exemplifies researcher-led praxis development, or a process of AfL 'practice-in-the-making'.

Hermansen and colleagues' research on AfL classrooms, in particular, highlights the situated, problem-solving nature of AfL praxis 'as a problem complex that needs to be explored and developed locally' (Hermansen and Nerland 2014, p.187). Her research draws attention to an understanding of assessment praxis as teacher 'knowledge work' that involves the active negotiation of existing practice and re-contextualisation of new practice in local settings. Against a background of established routines and conventions 'initiatives like AfL cannot simply be inserted into educational organisations, but need to be 'worked in' to the organisation for them to become part of the collective process (p.188). Assessment praxis therefore involves 'processes of reworking practice through making sense of the 'new' and the 'old', through investing artefacts with meaning, and through reworking relationships with student and parents' (p.195).

At a micro-interactional level this assessment knowledge work is transacted through localised and effortful appropriation, use and transformation of AfL-related epistemic and organisational tools such as assessment criteria, templates and generic guidelines. (Hermansen and Nerland 2014). As these tools, 'are historically laden and carry with them specific affordances and constraints for use' (Hermansen 2014, p.473) they may be used in isolated, limiting ("letter") or creative, expansive ("spirit") ways showing

> how AfL tools and procedures may become narrowly instrumentalist rather than supportive of holistic learning processes, or how previously existing assessment approaches may influence the ways in which AfL tools are taken up (Hermansen and Nerland 2014, p.190).

Moreover, the interaction and negotiation surrounding AfL tool appropriation, use and transformation constituting language assessment praxis is accomplished *dialogically* in collaborative teacher discourse, specifically through key discursive moves of elaboration and specification, validation, and adaptation (Hermansen and Nerland 2014).

From this AfL implementation research, three *planes of praxis* and praxis development can be identified with their respective mediational dynamics: the *cultural plane* (macro- professional level) where knowledge practice is mediated through 'knowledge relations and epistemic infrastructures' in AfL (Hermansen 2017); the *institutional plane* (meso-school level) where knowledge practice is mediated through organisational roles and routines, and the *situational plane* (micro-interactional level) where knowledge practice is mediated through localised pedagogical tools and resources. AfL implementation research provides a model of and basis for research into L2 assessment praxis.

It is possible to identify five collaborative models of 'close to classroom research' (Cochran-Smith and Donnell 2006) that foster L2 pedagogy and assessment praxis development through particular teacher-researcher roles. These may be described

as follows: (1) *researcher as experimenter* as exemplified in the theory-driven Galperin-inspired systemic-theoretical instruction model elaborated by Negueruela and colleagues (Negueruela 2003; Lantolf 2008, Lantolf and Poehner 2014); (2) *researcher-teacher partnerships,* that is, researcher-led classroom-based research or action research (e.g., Gibbons 2008, 2009; Hammond 2006, 2009; Hammond and Gibbons 2005; Hill and Ducasse, Chap. 8, this volume; Michell and Sharpe 2005); (3) *teacher as researcher,* or teacher-led action research such as that undertaken by 'Tracy' in 'practicalising' and operationalising dynamic assessment theory (Lantolf and Poehner 2011); this model is also exemplified in Leung's (Chap. 5, this volume) discussion of what he terms 'practitioner research' in the context of a teacher education program; (4) *researcher as co-teacher,* where researcher and teacher are 'learning from each other' as co-equals (e.g., Baker and Germain, Chap. 6, this volume; Gade 2015a, b); and (5) *co-teachers as researchers,* where teachers (e.g., generalist and specialist teachers) learn from each other through collaborative teaching (e.g., Martin-Beltran and Peercy 2014; Peercy and Martin-Beltran 2012).

The TEAL Toolkit as a Pedagogical Artefact Embodying L2 Assessment Praxis

Reflecting both its praxis origins and praxis development purpose, the online resource system, *Tools to Enhance Assessment Literacy for Teachers of English as an Additional Language (TEAL)*, http://teal.global2.vic.edu.au/, was developed through researcher-teacher collaboration involving the active input and ideas of hundreds of EAL specialists from selected government, Catholic and independent schools in the Australian states of Victoria and New South Wales. EAL teachers' assessment practice needs were used as the starting point for the design of the toolkit (Davison and Michell 2014). The project resulted in the development of a validated online assessment advice and 'toolkit' for use by all teachers to help assess the stage of development for EAL students in speaking and listening, reading and writing, with a view to improving learning and teaching. The assessment resource is not only a design artefact reflecting the distributed assessment praxis of teachers and researchers, it is also a mediating tool to support the ongoing work and continued innovation of both groups, that is, to guide continued L2 classroom assessment praxis in the future.

Drawing on Assessment for Learning (AfL) principles and a Vygotskian theory of learning, the TEAL toolkit includes four main components: (1) a set of sequenced teacher professional learning resources about EAL and AfL (including self-assessments) designed for small group or self-directed study; (2) an assessment tool bank containing a range of assessment tools and tasks organized around the three broad macro-skills (oral, reading and writing), three macro-functions (informative,

persuasive, imaginative), three stages of schooling (early elementary, mid to upper elementary, and secondary) and a range of EAL proficiency levels; (3) a range of assessment-for- learning and teaching exemplars including a selection of annotated units of work across a range of subject areas and year levels showing assessment tasks with formative feedback embedded within a teaching/learning cycle; and (4) an online teacher discussion forum, including a password-protected area for teachers to share problems and strategies and to moderate or benchmark work samples.

EAL assessment praxis is firstly reflected in the *content* of the TEAL toolkit. EAL teacher praxis informs the design of the TEAL resource including through key linguistic assessment design features such as language macro-functions, text framework (informative, imaginative and persuasive texts), and assessment criteria, and pedagogical design features such as task types, task-based assessment, task difficulty and scaffolded tasks. Secondly, EAL assessment praxis is also reflected in the TEAL toolkit development *process*. Teacher trialling, questionnaires, focus groups and teacher-based classroom observation data were used to help teachers implement assessment principles and theories, discourses and practices, and, importantly, elicit teacher feedback to capture current assessment praxis to inform toolkit development. This process at times forced researchers to engage with a number of new issues to build a more context-sensitive assessment model that took into account Australian EAL teachers' specific sociocultural, technical and practical needs. These considerations focused on issues of 'user friendliness', such as the structure and format of the TEAL elements and confirmation or revision of the resource in relation to practicality and use in context. The content and development of the toolkit therefore highlights the central and generative role of EAL teacher assessment praxis. Figure 2.1 provides an overview of the range of assessment tools and frameworks that informed development of the TEAL toolkit and the interaction of elements in use that comprises the multiple, tool-mediated system that guides development of teacher L2 assessment praxis. TEAL therefore reflects a researcher-led model of teacher-developer partnership involving strong forms of distributed praxis.

In its design and content, the TEAL toolkit therefore reflects three applications of praxis and praxis development. As a pedagogical artefact, the toolkit represents AfL praxis applied to EAL education; as an assessment artefact 'systematising' EAL assessment practices, it draws on and formalises EAL teacher assessment praxis; and as an 'epistemic infrastructure' (Hermansen 2017) for the profession, it embodies EAL teacher assessment praxis and advances its development. In its intended use, the TEAL toolkit aims to impact on three planes of praxis: developing L2 assessment praxis on the situational plane at the micro-interactional classroom level; developing L2 assessment organisational praxis on the institutional plane at the school level, and developing L2 assessment communities of praxis on the cultural plane at the macro-professional level.

The study of teacher L2 assessment praxis development through use of the TEAL toolkit highlights the nature of the assessment activity system and its role in

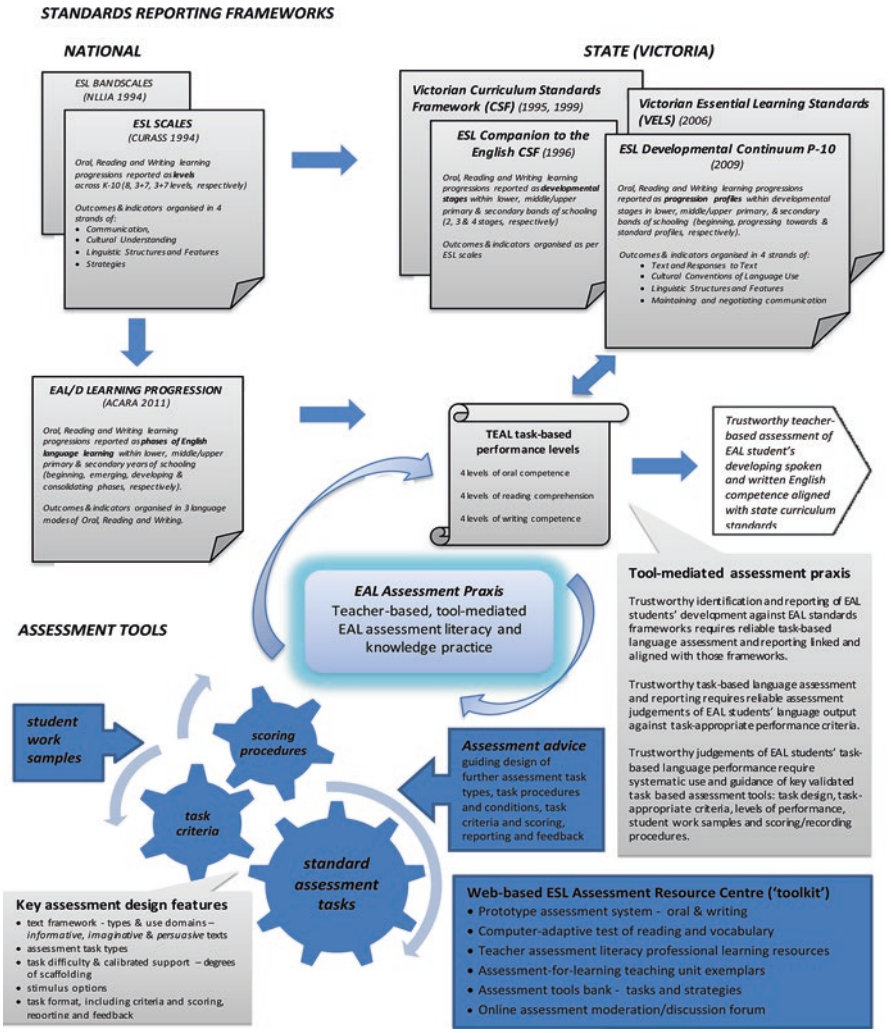


Fig. 2.1 Tools to enhance assessment literacy for teachers of EAL (TEAL) as a tool-mediated assessment system

developing L2classroom praxis. As shown in Fig. 2.2, the *proximal goal* of the system is trustworthy language assessment and reporting and teacher development of EAL and context-appropriate assessment tools, guided by the multiple mediation of a range of assessment tools, rules and roles within an emerging community of assessment practice. The *ultimate outcome* of this system, however, is L2 classroom assessment praxis and a community of assessment praxis with the possible emergence of leadership in the area of EAL assessment.

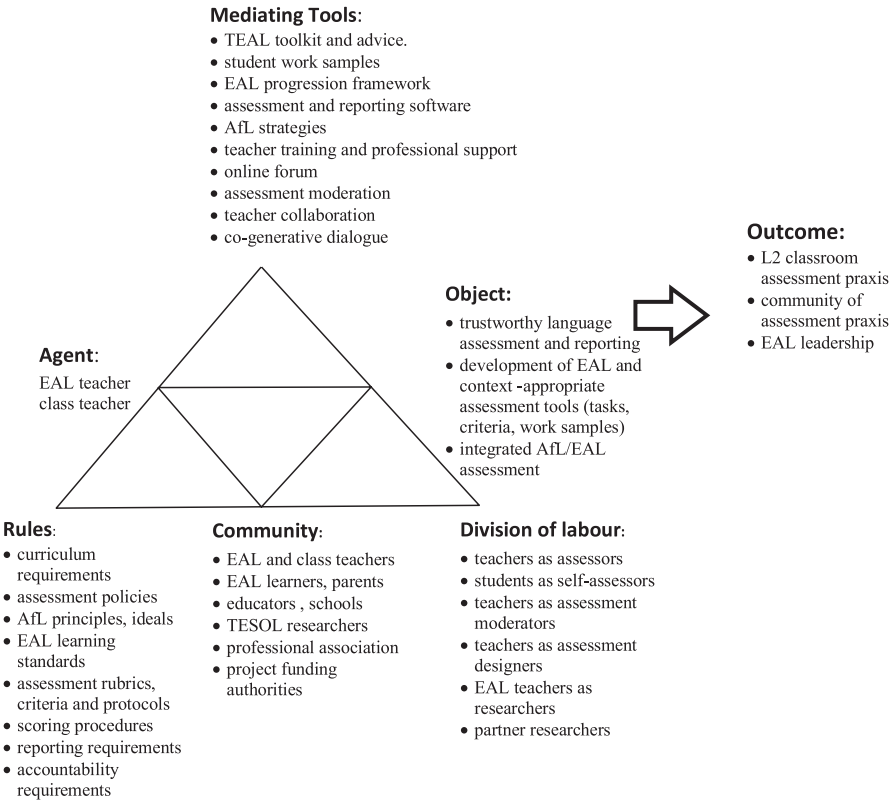


Fig. 2.2 TEAL within an emerging community of praxis

Conclusion

Teacher praxis always involves the development of some form of 'new' practice. Examples include beginning teachers establishing a pedagogic repertoire; experienced teachers changing practice to meet the demands of a new student group or teaching context; and improving or innovating practice in response to feedback, policy, 'reforms', or application of theory in classrooms. Praxis involves the translation of new practice into given contexts or existing practice into new ones.

The appropriate unit of analysis of assessment praxis is not individual teacher assessment behaviours but the assessment activity system as a whole. These activity systems promote object-, other- and self-regulated assessment praxis through multiple mediations in assessment tool use. Such assessment activity systems embody and encourage both individual and collective teacher assessment praxis at situational, institutional and professional levels and are catalytic in the development of communities of EAL teacher assessment practice and learning.

The TEAL toolkit is an artefact and product of a teacher-developer/researcher praxis that in turn fosters individual and collective language assessment praxis. With its focus on tool-mediated agency, developing knowledge practice and transforming teaching selves and classrooms, this conceptualisation of L2 classroom assessment praxis 'brings the teacher back in', putting them back into the heart of assessment and humanising it in the process.

References

Anderson, M., & Freebody, K. (2012). Developing communities of praxis: Bridging the theory practice divide in teacher education. *McGill Journal of Education, 47*(3), 359–377.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading: Addison-Wesley.

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge, School of Education.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (2005a). *Inside the black box: Raising standards through classroom assessment*. London: Granada Learning.

Black, P., & Wiliam, D. (2005b). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal, 16*(2), 249–261.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2007). *Assessment for learning*. Maidenhead: McGraw-Hill International (UK) Limited.

Bottomore, T. (1983). *A dictionary of Marxist thought*. Oxford: Blackwell.

Clandinin, D. J., & Connelly, F. M. (1998). Stories to live by: Narrative understandings of school reform. *Curriculum Inquiry, 28*(2), 149–164.

Cochran-Smith, M., & Donnell, K. (2006). Practitioner inquiry: Blurring the boundaries of research and practice. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 503–518). Mahwah: LEA.

Davies, J., & Ecclestone, K. (2008). 'Straitjacket'or 'springboard for sustainable learning'? The implications of formative assessment practices in vocational learning cultures. *The Curriculum Journal, 19*(2), 71–86.

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37–68.

Davison, C. (2013). Innovation in assessment: Common misconceptions and problems. In K. Hyland & L. Wong (Eds.), *Innovation and change in English language education* (pp. 263–275). Oxon: Routledge.

Davison, C., & Michell, M. (2014). EAL assessment: What do Australian teachers want? *TESOL in Context, 24*(2), 51.

Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice, 18*(4), 365–379.

Edwards, A., Gilroy, P., & Hartley, D. (2005). *Rethinking teacher education: Collaborative responses to uncertainty*. London: Routledge.

Edwards-Groves, C., & Grootenboer, P. (2015). Praxis and the theory of practice architectures: Resources for re-envisioning English education. *Australian Journal of Language and Literacy, 38*(3), 150.

Edwards-Groves, C., & Kemmis, S. (2016). Pedagogy, education and praxis: Understanding new forms of intersubjectivity through action research and practice theory. *Educational Action Research, 24*(1), 77–96.

Engeström, Y. (2011). From design experiments to formative interventions. *Theory & Psychology, 21*(5), 598–628.

Ertsas, T. I., & Irgens, E. J. (2017). Professional theorizing. *Teachers and Teaching, 23*(3), 332–351.

Fenwick, T., Nerland, M., & Jensen, K. (2012). Sociomaterial approaches to conceptualising professional learning and practice. *Journal of Education and Work, 25*(1), 1–13.

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural–historical perspective. *Assessment in Education: Principles, Policy & Practice, 22*(2), 224–246.

Freire, P. (1970). *Pedagogy of the oppressed*. New York: Continuum.

Gade, S. (2011). Narrative as unit of analysis for teaching-learning praxis and action: Tracing the personal growth of a professional voice. *Reflective Practice, 12*(1), 35–45.

Gade, S. (2012). Close-to-practice classroom research by way of Vygotskian units of analysis. In *The 12th International Congress on Mathematics Education (ICME-12) July 8–15, Seoul, Korea (TSG21, Research on classroom practice)* (pp. 4312–4321).

Gade, S. (2014). Praxis and phronesis as units of analysis: Realising a social science that matters in practitioner inquiry. *Reflective Practice, 15*(6), 718–728.

Gade, S. (2015a). Unpacking teacher–researcher collaboration with three theoretical frameworks: A case of expansive learning activity? *Cultural Studies of Science Education, 10*(3), 603–619.

Gade, S. (2015b). Teacher-researcher collaboration as formative intervention and expansive learning activity. In *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3029–3035).

Gibbons, P. (2006). *Bridging discourses in the ESL classroom: Students, teachers and researchers*. London: Continuum.

Gibbons, P. (2007). Mediating academic language learning through classroom discourse. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 701–718). Boston: Springer.

Gibbons, P. (2008). 'It was taught good and I learned a Lot': Intellectual practices and ESL learners in the middle years. *Australian Journal of Language and Literacy, 31*(2), 155.

Gibbons, P. (2009). *English learners, academic literacy, and thinking: Learning in the challenge zone*. Portsmouth: Heinemann.

Grootenboer, P. (2013). The praxis of mathematics teaching: Developing mathematical identities. *Pedagogy, Culture and Society, 21*(2), 321–342.

Grootenboer, P., & Edwards-Groves, C. (2013). *Mathematics education as a practice: A theoretical position*. Adelaide: Mathematics Education Research Group of Australasia.

Grootenboer, P., & Edwards-Groves, C. (2014). *Mathematics teaching as praxis*. Adelaide: Mathematics Education Research Group of Australasia.

Hakkarainen, K. (2009). A knowledge-practice perspective on technology-mediated learning. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 213–231.

Hammond, J. (2006). High challenge, high support: Integrating language and content instruction for diverse learners in an English literature classroom. *Journal of English for Academic Purposes, 5*(4), 269–283.

Hammond, J. (2009). High challenge, high support programmes with English as a second language learners: A teacher-researcher collaboration. In *Culturally and linguistically diverse classrooms: New dilemmas for teachers* (pp. 56–76). Bristol/Buffalo: Multilingual Matters.

Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospects, 20*(1), 6–30.

Hedges, H. (2012). Teachers' funds of knowledge: A challenge to evidence-based practice. *Teachers and Teaching, 18*(1), 7–24.

Hermansen, H. (2014). Recontextualising assessment resources for use in local settings: Opening up the black box of teachers' knowledge work. *Curriculum Journal, 25*(4), 470–494.

Hermansen, H. (2016). Teachers' knowledge work in collective practice development: Approaches to introducing assessment for learning at the school level. *Scandinavian Journal of Educational Research, 60*(6), 679–693.

Hermansen, H. (2017). Knowledge relations and epistemic infrastructures as mediators of teachers' collective autonomy. *Teaching and Teacher Education, 65*, 1–9.

Hermansen, H., & Nerland, M. (2014). Reworking practice through an AfL project: An analysis of teachers' collaborative engagement with new assessment guidelines. *British Educational Research Journal, 40*(1), 187–206.

Hill, M. F. (2011). 'Getting traction': Enablers and barriers to implementing assessment for learning in secondary schools. *Assessment in Education: Principles, Policy & Practice, 18*(4), 347–364.

Hoffman-Kipp, P., Artiles, A. J., & Lopez-Torres, L. (2003). Beyond reflection: Teacher learning as praxis. *Theory Into Practice, 42*(3), 248–254.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice, 16*(3), 269–290.

Karpov, Y. V., & Bransford, J. D. (1995). L.S. Vygotsky and the doctrine of empirical and theoretical learning. *Educational Psychologist, 30*(2), 61–66.

Kemmis, S. (2010). Research for praxis: Knowing doing. *Pedagogy, Culture and Society, 18*(1), 9–27.

Kemmis, S. (2012a). Researching educational praxis: Spectator and participant perspectives. *British Educational Research Journal, 38*(6), 885–905.

Kemmis, S. (2012b). Phronesis, experience, and the primacy of praxis. In E. A. Kinsella & A. Pitman (Eds.), *Phronesis as professional knowledge* (pp. 147–161). Rotterdam: Sense Publishers.

Kemmis, S., & Smith, T. J. (2008). *Enabling praxis: Challenges for education*. Rotterdam: Sense Publications.

Kemmis, S., Wilkinson, J., Edwards-Groves, C., Hardy, I., Grootenboer, P., & Bristol, L. (2014). *Changing practices, changing education*. Singapore: Springer.

Kessels, J. P., & Korthagen, F. A. (1996). The relationship between theory and practice: Back to the classics. *Educational Researcher, 25*(3), 17–22.

Korthagen, F. A., & Kessels, J. P. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational Researcher, 28*(4), 4–17.

Korthagen, F. A., Kessels, J., Koster, B., Lagerwerf, B., & Wubbels, T. (2001). *Linking practice and theory: The pedagogy of realistic teacher education*. London: Routledge.

Lantolf, J. P. (2008). Praxis and classroom L2 development. *ELIA, 8*, 13–44.

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11–33.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York: Routledge.

Law, N. (2014). Knowledge building pedagogy and teachers' technological pedagogical content knowledge. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 227–247). Singapore: Springer.

Lee, Y. J. (2014). Knowledge creation in the mangle of practice: Implications for educators. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 167–187). Singapore: Springer.

Mäkitalo, Å. (2012). Professional learning and the materiality of social practice. *Journal of Education and Work, 25*(1), 59–78.

Marshall, B., & Jane Drummond, M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21*(02), 133–149.

Martin-Beltran, M., & Peercy, M. M. (2014). Collaboration to teach English language learners: Opportunities for shared teacher learning. *Teachers and Teaching, 20*(6), 721–737.

Michell, M., & Sharpe, T. (2005). Collective instructional scaffolding in English as a second language classrooms. *Prospects, 20*(1), 31–58.

Miettinen, R., & Virkkunen, J. (2005). Learning in and for work, and the joint construction of mediational artefacts: An activity theoretical view. In E. Antonacopoulou, P. Jarvis, V. Andersen, B. Elkjaer, & S. Høyrup (Eds.), *Learning, working and living* (pp. 154–169). London: Palgrave Macmillan.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice, 31*(2), 132–141.

Negueruela, E. (2003). *A sociocultural approach to the teaching and learning of second languages: Systemic-theoretical instruction and L2 development.* Unpublished doctoral dissertation. The Pennsylvania State University, University Park.

Park, H., Hewson, P. W., Lemberger, J., & Marion, R. D. (2010). The interactions of conceptions of teaching science and environmental factors to produce praxis in three novice teachers of science. *Research in Science Education, 40*(5), 717–741.

Peercy, M. M., & Martin-Beltran, M. (2012). Envisioning collaboration: Including ESOL students and teachers in the mainstream classroom. *International Journal of Inclusive Education, 16*(7), 657–673.

Rahmawati, Y., Koul, R., & Fisher, D. (2015). Teacher–student dialogue: Transforming teacher interpersonal behaviour and pedagogical praxis through co-teaching and co-generative dialogue. *Learning Environments Research, 18*(3), 393–408.

Ritella, G., & Hakkarainen, K. (2012). Instrumental genesis in technology-mediated learning: From double stimulation to expansive knowledge practices. *International Journal of Computer-Supported Collaborative Learning, 7*(2), 239–258.

Roth, W. M., & Tobin, K. (2004). Coteaching: From praxis to theory. *Teachers and Teaching, 10*(2), 161–179.

Sannino, A. (2015). The principle of double stimulation: A path to volitional action. *Learning, Culture and Social Interaction, 6*, 1–15.

Sannino, A., & Laitinen, A. (2015). Double stimulation in the waiting experiment: Testing a Vygotskian model of the emergence of volitional action. *Learning, Culture and Social Interaction, 4*, 4–18.

Scardamalia, M., & Bereiter, C. (1999). Schools as knowledge-building organizations. In D. H. Keating & C. Hetzman (Eds.), *Today's children, tomorrow's society: The developmental health and wealth of nations* (pp. 274–289). New York: Guildford.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building environments: Extending the limits of the possible in education and knowledge work. In A. DiStefano, K. E. Rudestam, & R. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 269–272). Thousand Oaks: Sage.

Smith, T., Edwards-Groves, C., & Brennan Kemmis, R. (2010). Pedagogy, education and praxis. *Pedagogy, Culture and Society, 18*(1), 1–8.

Spicer, D. H. E. (2011). From artefact to tool: teachers' collective agency in school reform. *Pedagogies: An International Journal, 6*(4), 359–380.

Tang, S. Y. F. (2010). Teachers' professional knowledge construction in assessment for learning. *Teachers and Teaching: Theory and Practice, 16*(6), 665–678.

Teo, C. L. (2014). Conceptual shifts within problem spaces for knowledge-building practice within a teacher community. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 205–226). Singapore: Springer.

Thomas, P. D. (2015). Gramsci's Marxism: The 'philosophy of praxis'. In M. McNally (Ed.), *Antonio Gramsci* (pp. 97–117). London: Palgrave Macmillan.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education, 14*(3), 281–294.

Tsui, A. B. (2009). Distinctive qualities of expert teachers. *Teachers and Teaching: Theory and Practice, 15*(4), 421–439.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 259–287.

Vygotsky, L. S. (1926/2006). *Educational psychology* (R. Silverman, Trans.). New Delhi: Pentagon Press.

Vygotsky, L. S. (1987). Chapter 6; development of scientific concepts in childhood. Thinking and speech. In R. Rieber, A. Carton, & N. Minick (Eds.), *The collected works of L.S. Vygotsky: Problems of general psychology* (Vol. 1, pp. 43–285). New York: Plenum.

Vygotsky, L. S. (1994). The socialist alteration of man. In R. Van der Veer & J. Valsiner (Eds.), *The Vygotsky Reader* (pp. 175–184). Oxford/Cambridge: Blackwell.

Vygotsky, L. (1997a). The historical meaning of the crisis in psychology: A methodological investigation. In *The collected works of LS Vygotsky* (Vol. 3, pp. 233–343). New York: Plenum.

Vygotsky, L. S. (1997b). The methods of reflexological and psychological investigation. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L.S. Vygotsky: Problems of the theory and history of psychology* (Vol. 3, pp. 35–49). New York: Plenum.

Vygotsky, L. S. (1997c). The instrumental method in psychology. In R. W. Rieber (Ed.), *The collected works of L.S. Vygotsky: Problems of the theory and history of psychology* (Vol. 3, pp. 85–89). New York: Plenum.

Vygotsky, L. S. (1999a). Chapter 2: Research method. In A. S. Carton & R. W. Rieber (Eds.), *The collected works of LS Vygotsky: The history of the development of higher mental functions* (Vol. 4, pp. 27–63). New York: Plenum.

Vygotsky, L. S. (1999b). Chapter 3: Analysis of higher mental functions. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L.S. Vygotsky: The history of the development of higher mental functions* (Vol. 4, pp. 65–82). New York: Plenum.

Vygotsky, L. S. (1999c). Chapter 9: Mastering attention. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L.S. Vygotsky: The history of the development of higher mental functions* (Vol. 4, pp. 153–179). New York: Plenum.

Vygotsky, L. S. (1999d). Tool and sign in the development of the child. In R. W. Rieber (Ed.), *The collected works of L.S. Vygotsky: Scientific legacy* (Vol. 6, pp. 3–68). New York: Plenum.

Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice, 16*(2), 165–184.

Wertsch, J. (2007). Mediation. In H. Daniels, M. Cole, & J. Wertsch (Eds.), *The Cambridge companion to Vygotsky*. Cambridge/New York: Cambridge University Press.

Wubbels, T., Korthagen, F., & Brekelmans, M. (1997). Developing theory from practice in teacher education. *Teacher Education Quarterly, 24*(3), 75–90.

Yeo, J. (2014). From problem-based learning to knowledge creation. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 143–166). Singapore: Springer.

Chapter 3
Mediation in the Assessment of Language Learning Within an Interlingual and Intercultural Orientation: The Role of Reciprocal Interpretation

Angela Scarino

Abstract Over the past two decades, consideration of the nature and role of culture in foreign language learning has gained renewed prominence in the educational work of teachers. The expanded understanding of language learning to include an interlingual and intercultural orientation to language teaching and learning has presented major challenges for assessment on the part of teachers. In this chapter I discuss the collaborative objectives and processes of a three-year collective case study that investigated teacher assessment of student language learning within this orientation. A team of researchers worked with 15 highly experienced teachers of a range of languages in both primary and secondary school settings (i.e., kindergarten through twelfth grade) over the course of 2 year-long assessment cycles. The researchers accompanied the teachers through their work on conceptualising the nature of language learning within an interlingual and intercultural orientation, designing assessment experiences to elicit language learning; gathering classroom learning and assessment data; analysing and judging samples of student work; and evaluating the overall process. Through ongoing *facilitated dialogues* researchers mediated collaborative, interpretive analyses together with the teachers as a group. These analyses focused on identifying and explaining the emergent characteristics of the teachers' assessment designs and the group's evolving understanding of the expanded construct of language learning and specifically the phenomenon of assessment of language learning within an interlingual and intercultural orientation.

A. Scarino (✉)
Research Centre for Languages and Cultures, School of Creative Industries, University of South Australia, Adelaide, Australia
e-mail: Angela.Scarino@unisa.edu.au

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_3

Introduction

Over the past two decades, consideration of the nature and role of culture in foreign language learning has gained renewed attention in the educational work of teachers of languages (Byrnes 2010; Kramsch 1993, 2004). This is a part of an expanded understanding of language itself as personal, creative, and expressive (Shohamy 2006), of communication as going beyond the simple exchange of information in interaction to the reciprocal exchange of symbolic meanings (Kramsch 2006, 2010), and of learning as the mutual, interpretive process of making sense of participant contributions (the subject matter) and, at the same time, each other (the person) (Gallagher 1992; Scarino 2014; Douglas Fir Group 2016). It is this expanded understanding that is captured in an *interlingual and intercultural orientation* to language teaching and learning (Byram 1997, 2014; Kramsch 2014; Liddicoat and Scarino 2013). Within this orientation language learning itself is understood as both an interlingual and intercultural process whereby students learn to mediate between languages and cultures. Mediation is central to the processes of interpreting, creating and exchanging meanings across languages and cultures. Students learn to become intercultural mediators, analysing the meanings of others, which are constructed within particular cultural framings. The process provides those who do not share a cultural framing with the means to understand diverse others (Gohard-Radenkovic et al. 2004). This orientation foregrounds a relational view of language learning and knowing in which four kinds of knowledge come together within experiences of language use and learning over different time scales. These are: disciplinary knowledge (subject matter or content); linguistic and cultural knowledge; procedural knowledge (processes of interpretation and analysis); and metacognitive knowledge (developed through reflection and reflexivity—see Byrd-Clark and Dervin 2014). It is also an orientation that recognises and develops the capability of students to integrate in communication an understanding of themselves as already situated in language/s and culture/s, and an understanding of the same in others. Assessment of language learning within this orientation is an under-researched dimension of language/s education and presents major challenges for assessment on the part of teachers.

In this chapter, I first describe the complex nature of assessing language/s learning within an interlingual and intercultural orientation, and the need to reconceptualise assessment in order to capture intercultural practices and capabilities. I then consider learning-oriented assessment (Leung 2004; Turner and Purpura 2015; Purpura 2016) as a way of approaching assessment that offers possibilities for capturing the expanded capabilities in language learning. Next, I discuss the collaborative objectives and processes of a three-year collective, phenomenographic case study that investigated – *with* teachers of different languages – the assessment of student language learning within an interlingual and intercultural orientation. I foreground the process of ongoing *facilitated dialogues* as the process through which the researchers mediated collaborative, interpretive analyses undertaken jointly by

researchers and teachers that focused on identifying and explaining the emergent characteristics of the teachers' assessments and the group's evolving understanding of both the expanded construct of language learning and the phenomenon of the assessment of language learning within an interlingual and intercultural orientation. Put simply, the project was one of praxis, where "theory and research are accountable to practice; praxis represents a unity of theory and practice wherein they inform one another and change together" (Poehner and Inbar-Lourie, Chap. 1, in the introduction to this volume, p. 5). It simultaneously sought to foster teachers' understandings of the expanded interlingual and intercultural conceptualisation of language learning and support their efforts to develop their assessment practices according to these new understandings while simultaneously documenting these processes as they offer an important point of departure for future work and for the continued elaboration of the role of assessment in supporting language learning.

Assessing Language Learning Within an Interlingual and Intercultural Orientation

For students learning a language in addition to their primary one, that is, a second language (L2), the learning process itself, as well as the process of using language/s in communication to interpret and exchange meanings, involves processes of 'moving between' languages, cultures and knowledge systems. For each learner a different configuration of languages may be involved, and each learner will have a different affiliation with the specific language being learnt and be challenged in different ways by the additional language and its proximity to or distance from his/her primary language. As such, intercultural language learning may be understood as a hermeneutically inspired view of language learning that takes account of the linguistic and cultural background of the learner and the role of languages and cultures in the act of learning (Liddicoat and Scarino 2013; Scarino and Liddicoat 2016).

In recent years there has been an increasing interest in assessing intercultural practices and capabilities, both in language learning and in education in general. However, as Byram (2014) noted in a recent 25-year review of foreign language learning, the question of assessment specifically in the context of intercultural language learning remains insufficiently developed. Kramsch (2009), who developed the most elaborated and complex theorisation of culture and the intercultural in language/s learning, questioned whether or not these capabilities, which she conceptualises as 'symbolic competence', can be assessed:

> [S]ymbolic competence based on discourse would be less a collection of … stable knowledges and more a savviness i.e. a combination of knowledge, experience and judgment … Trying to test symbolic competence with the structuralist tools employed by schools … is bound to miss the mark. Instead symbolic competence should be seen as the educational horizon against which to measure all learners' achievements (p.118).

Kramsch highlighted that assessment of this kind of language learning cannot be an assessment of knowledge, understood as facts or skills, but of a bringing together of linguistic, cultural, and 'world' knowledge, experiential knowledge (derived from direct participation and experience of the phenomenon) and judgment (which might be understood as the capability to be reflective and reflexive). She also recognised that in a school context, wherein assessment is traditionally conceived and institutionally prescribed, that is, as measurement, depersonalised, standardized and objective (Shohamy 2001), the nature of intercultural language learning cannot be captured. However, alternative assessment paradigms that foreground learning and development and that are open to diversity of interpretation may provide possibilities (Scarino 2010, 2017). Equally, Shohamy (2011) has called for assessing 'multilingual competences' and Block (2014) has highlighted the need in language teaching and learning to go 'beyond lingualism' to include dimensions such as subjectivity and identity. Language learning and use can be seen as inherently subjective; culture and the intercultural are subjectively experienced and construed (Sercu 2002). This kind of expansion of the dimensions that are enmeshed in contemporary understandings of language learning renders the assessment of language/s learning more challenging than previously conceived. The challenge is in assessing expanded goals of learning language/s (Leung and Scarino 2016) and the opening up of understandings of language learning itself as a complex interdisciplinary endeavour (Douglas Fir Group 2016), together with increased attention to the historicity and subjectivity of language learners (Kramsch 2009) and to reflectivity and reflexivity (Byrd-Clark and Dervin 2014). Furthermore, capturing processes of mediation and interpretation in the assessment of language learning presents additional challenges.

This conceptual complexity is intensified by the legacy of communicative language teaching and its tendency to frame communication as no more than a transactional exchange, a conceptual framing of language learning (or rather, language teaching) that persists in the minds of teachers. In contrast, within an interlingual and intercultural orientation to the assessment of language learning, there is a need to capture:

- observation, description, analysis and interpretation of phenomena shared when communicating and interacting
- active engagement with the interpretation of self (intraculturality) and 'other' (interculturality) in diverse contexts of exchange
- understanding of the ways in which language and culture come into play in interpreting, creating and exchanging meaning
- the recognition and integration into communication of an understanding of self (and others) as already situated in one's own language and culture when communicating with others (Liddicoat and Scarino 2013, 130–131).

Thus, in the design of assessment of language learning that seeks to elicit students' developing interlingual and intercultural capabilities it becomes necessary to capture participation in or performance of an exchange that involves the interpretation and exchange of meaning in a reciprocal process of simultaneously making sense of diverse participant contributions to the subject matter and of the persons

making those contributions (Gallagher 1992). In operational terms, the assessment design needs to (1) incorporate multiple linguistic and cultural frameworks, (2) include differing positions, understandings or perspectives that need to be mediated (compared, interpreted, connected, juxtaposed, etc.), (3) draw upon the students' own experiences in which students are personally invested, and that make use of their own positioning and assumptions, and (4) invite interaction, analysis and reflection. This analysis and reflection is directed both to the experience of communication or exchange and the self. Gallagher argued that 'understanding' in learning is not an abstract, mental act but a linguistic – and cultural – event, because language has a central role to play in coming to understand the world. In this sense, learning is understood from the learner's point of view as a process of interpretation, recognising that all interpretation is governed by history and all interpretation itself is linguistic and cultural. In the assessment of language learning within an interlingual and intercultural orientation, therefore, it becomes necessary to take a process-oriented view, with a focus on assessment to promote learning.

Learning-Oriented Assessment

Traditional conceptions of assessment with its orientation towards objectivity and with notions of fixed tasks, rubrics, criteria and scales are not sufficient to capture the assessment of language learning within an interlingual and intercultural orientation. An assessment process that does capture this orientation needs to be in tune with processes such as experiencing, analysing and reflecting, and with the development of these processes over time. In other words, it needs to capture both the experience of communicating, and metacognitive and metalinguistic development. It also requires greater synergy between teaching, learning and assessment. This means that assessment cannot be separated from the curriculum and its shaping force on student learning.

Many proposals are available for the use of assessment to support the multidimensional and dynamic process of language/s learning. These include insights from *assessment for learning* (Black and Wiliam 2009), learning-oriented assessment (Turner and Purpura 2015), and dynamic assessment (Poehner 2007, 2011, 2013). These developments in assessment all point to an opening up of assessment towards a long-term perspective on assessment, as personalised and developmental. It is an assessment that moves beyond 'finished products' (Moss et al. 2008) and includes consideration of students' progression in learning, their capacity to make sense of disparate information and multifaceted meanings from diverse perspectives, and to use such knowledge in constructive ways. In this sense, assessment itself becomes a process of inquiry (Delandshere 2002) and research, going beyond the initial experience of exchange to include, through student decentring, analysis, reflection and reflexivity (Byrd-Clark and Dervin 2014; Leung, Chap. 5, this volume). This view of assessment presents challenges for teachers. This is not because such assessment paradigms are not perceived to be closely aligned to their work, but

because many teachers have come to internalise an 'objective' view of assessment and are reluctant to embrace an alternative that might not be understood as legitimate by their school leaders and as such run counter to the institutional requirements of assessment, which they feel they must respect.

For teachers of languages who seek to assess language learning within an interlingual and intercultural orientation, a dual change is needed. On the one hand, they need to work with the expanded language/s learning constructs that this orientation brings, and on the other, they need to work with alternative paradigms of assessment (McNamara 2003). They need to do this in a context where assessment remains that part of the educational endeavour which is least amenable to change, and where, equally, it is teachers who are uniquely positioned to work with and assess students' learning and development over time. In both areas there is value for teachers to partner with researchers who can provide the conceptual expertise that is needed and who can work with teachers to facilitate making sense of these shifts in redesigning, enacting and reflecting on their assessment work. Examples of such partnerships may be found in the Chaps. 6 and 8 in this volume by Baker and Germaine and Hill and Ducasse. While these studies do not take precisely the interlingual and intercultural orientation as the present chapter, they do illustrate some of the possibilities for how both assessment practices and research may be enriched through teacher-researcher partnerships. Baker and Germaine report a collaboration focused on embedded formative assessments in a nature-based Mi'gmaq language learning environment while Hill and Ducasse examine the use of a research-informed assessment resource as a mediating tool to guide reflection on assessment feedback practices in a Spanish language classroom while relying on this process to simultaneously elaborate the resource itself.

To investigate what might be entailed in the assessment of interlingual and intercultural capabilities, and because of the complex shifts in conceptualisation that are required in assessment within an interlingual and intercultural orientation to language teaching and learning, a small team of researchers at the Research Centre for Languages and Cultures, University of South Australia designed a collaborative, phenomenographic study with teachers of languages to address the question of assessment within this orientation. The reciprocal nature of the study was such that theoretical insights from the researchers would inform teacher practice and afford opportunities for teachers to explore their practices reflectively and reflexively, inviting an engagement with praxis; and at the same time, teacher practice would offer a basis for experimenting with practice to inform ongoing theorisations.

The Study and the Cycle of Assessment

The three-year study involved a researcher–teacher partnership, focused on generating insights into the ways in which assessment of language learning within an interlingual and intercultural orientation might be undertaken. Fifteen experienced teachers of diverse languages in both primary and secondary schools (i.e., kindergarten through to twelfth grade) participated in a collective case study over 2 year-long

cycles of assessment. The researchers accompanied the teachers in the interrelated processes that included the following: (1) conceptualising with the teachers assessment within an interlingual and intercultural framing; (2) eliciting an exemplar instructional unit and accompanying assessment plan from the teachers; (3) judging, again conducted jointly with the teachers, of student work and capabilities; and (4) validating as the team reflected on the assessment process and planned ahead.

Effort to reconceptualise assessment as interlingual and intercultural occurred during facilitated workshops in which the researchers worked to mediate discussion among the entire group of participants. The facilitated workhops were designed first to problematise existing, conventional notions of assessment, language, and learning, and to proceed from this to then open up discussion of how assessment might be construed within an interlingual and intercultural orientation to language/s learning. A central component of the workshops were *facilitated dialogues* wherein researchers endeavored to further mediate participant thinking through roundtable discussions, returning continuously to group discussion of the construct of interlingual and intercultural capabilities. It must be highlighted here that for the researchers too, the process was a highly exploratory one. They recognised that they needed to work with teachers in practice as a means towards further theorizing the nature of the construct.

For the second stage of the assessment cycle, teachers designed a unit of work to provide a context for the design of assessment procedures, including all the accompanying resources in the design. They provided a written *rationale* for each aspect of their designs, specifying the (multiple) kinds of data they would collect as evidence of interlinguistic and intercultural capabilities. Although teachers do not normally provide a written rationale as part of their planning, the researchers invited such an explanation as a means for capturing the teachers' thinking, specifically in relation to the construct. The teachers also paid specific attention to the *specificity* of each particular language and its learning, *contextualising* the assessment within their own teaching and learning programs and contexts.

Researchers provided individual, written feedback on each teacher's unit of work and assessment procedures. Based on understandings that emerged from the feedback process, the teachers were invited to shift from talking about 'assessment tasks' to 'data points', signaling procedures and times along the trajectory of teaching and learning when assessment data would be gathered. This shift, as noted by the researchers, arose from the discussions with teachers when it became clear that they were focusing on 'tasks' and 'products' of learning, and thereby potentially missing evidence that might arise from classroom interactions as part of the process of teaching and learning. In roundtable discussion forums, teachers each provided a narrative account of their unit of work and their experience of teaching, learning and assessing, considering the processes and evidence, as well as processes of reflection and reflexivity. Finally, they revised all documentation based on facilitated group discussions and implemented the planned units of work and related assessment procedures.

The third and fourth stages of the cycle were concerned, respectively, with forming judgements of student performance and validation of the entire assessment

process. Teachers reported to the group as a whole through narratives of experiences of making judgments of students' language learning within an interlingual and intercultural orientation. Researchers and teachers participated in *mediated, collaborative, interpretive analysis* to capture evolving understandings of the phenomenon of interlingual and intercultural capabilities, unravelling along the way various assumptions about both the construct and assessment practices held by teachers and the researchers themselves, rationales, desires/intentions, expectations, reflections, and understandings (As indicated above these discussions were recorded and fully transcribed to capture the interplay of knowledge/theory and practice on the part of all participants). Finally, researchers and teachers participated in collaborative reflection on the activity of assessment and on the research process, connecting experience to the conceptualization of language/s learning and its assessment as an interlingual and intercultural endeavour, in preparation for subsequence cycles.

Study Design

In keeping with an overarching commitment to praxis, the design of the project recognised that the processes of conceptualising, eliciting, judging and validating are interrelated through the ways in which the teacher-assessors interpret and make inferences, based on their evolving understanding of what is relevant to assessment of language learning within an interlingual and intercultural orientation. The work of the researchers with the teachers paralleled the work of the teachers of language/s in mediating the learning and assessment of their students through the central process of researcher mediation and the facilitated (and recorded) dialogues. Just as the teachers focused both on their students' participation, analysis and reflection in the teaching, learning and assessment processes and on 'probing their students' (self and other) meanings, positioning, reactions, responses, use or choice of language/s and culture/s and identities enacted, so too, the teachers in the study participated in designing, enacting and judging in the assessment processes and responded to the researchers' and each other's probes related to evolving understandings, meanings, judgments, and so on.

Ongoing points of focus throughout the project included teacher rationales for their designs; the collaborative analysis of multiple data sources, including teachers' written commentaries to evidence their understanding; and reflections on the diverse perspectives and contributions to the evolving discussion. Equally, the researchers were engaged in ongoing reflection, fine-tuning their own understandings of the construct, processes for eliciting students' learning and the nature of evidence of learning within this orientation. Through this iterative process, repeated in cycles over 2 years, the researchers and teachers captured their evolving conceptualisation of the construct and processes for eliciting and judging students' development of interlingual and intercultural capabilities.

To help the reader appreciate the constant interplay of conceptualisation and planning, action ←→ reaction ←→ debriefing/feedback ←→ theorizing that

characterised the project, itself understood as an exchange of meanings, the next section offers an example of the assessment process focused on one teacher. It must be acknowledged that no single example can represent the detail and nuance of the evolving conversation and learning of a community of collaborating teachers and researchers but limitations of space do not permit otherwise. The case described below was selected because the teacher was working in Chinese with students who were learning it as an L2. Given the complexity that the Chinese language presents for these learners, it also became the language learning and assessment case-study that presented the greatest challenge in the collaborative project.

A Focal Case: Interlingual and Intercultural Assessment and Female Identity

During the first year of the study, one of the teachers was working on a curricular unit, the aim of which, as the teacher expressed it, was 'to explore the students' female identities through examining examples of Chinese female identity.' The teacher (henceforth, Margie) was a Year 11 Chinese teacher in an all-girls Catholic school. The inquiry question Margie posed in her unit of work was "What does it mean to be a woman in today's world?" Her narrative account of her planning portrays her initial conceptualisation of the nature of interlinguistic and intercultural capabilities and her ways of developing and eliciting these capabilities. This is captured in the excerpt from her planning narrative represented in Fig. 3.1.

The guiding question will form a starting point of our explorations and we will keep referring to this question as we examine texts and will use this question as a stimulus for reflection throughout the topic. At this stage I see this question as our overarching focus for the learning program, however I envisage that as we progress through the tasks, sub-questions will arise either from myself or the students that we will add to our learning focus. The goal for this is to enable students to develop a deeper understanding of their own femaleness and that of young women in China and to develop students' skills in accessing online Chinese texts … While the end point of assessment will be in the form of an investigative task there will be several formative and summative tasks along the way to help students prepare for this final task and also provide a scope for assessing students' progress in intercultural thinking as I believe that I cannot assess the intercultural simply through performance in one task, but rather through monitoring individual students' evidence of deepening understanding from one task to the next. The tasks are also designed to help students connect the intracultural to the intercultural through looking at their own ideas of femaleness and then comparing those with some views held in China. My concern with only choosing one text is that the views will be very one-sided and I think it is important to present a range of views. The tasks are also designed with various communication forms in mind (mostly involving internet access) so that students are able to engage in various interactions with texts both through discussion and response as well as actual involvement in chat forums.

Fig. 3.1 Excerpt from Margie's narrative account of designing her unit

This extract reveals Margie's intention to ensure the exploration of a concept – 'femaleness'. Through the collaborative work with the teachers, the researchers came to appreciate that it is in working with 'concepts' rather than 'descriptions' (which have become typical of language/s learning in K–12) that interlinguistic and intercultural comparison becomes of interest and relevant. Margie also addresses personalisation, which again, the researchers came to understand as central to connecting the intercultural with the intracultural (which had been discussed in a preliminary workshop at the commencement of the study); she sees her students as posing questions in addition to her own and reflecting on themselves as young women. Moreover, she realises the need to draw upon a range of texts that are authentic and that present and allow for multiple perspectives that will invite multiple interpretations within and across cultures. Again, it is through this project that the researchers came to understand the way in which working with diverse perspectives is integral to interlinguistic and intercultural language learning.

As Year 11 in Australia represents the first year of a two-year senior secondary cycle that leads to the end of the secondary school external assessment, the teacher is mindful of the requirements, which include a culminating 'investigative task'. She realises that there will need to be an 'end point' assessment but, at the same time, she knows that she will include formative assessment 'to help students prepare for this final task. It will also provide scope for assessing students' progress in developing their interlinguistic and intercultural capabilities, not through performance on one task alone, but through monitoring individual students' evidence of deeper understanding from one task to the next. Margie recognises the developmental character of interlingual and intercultural capabilities as another characteristic of language learning. She specifically seeks to connect 'the intracultural to the intercultural', followed by comparison, echoing concepts that had been discussed in the initial conceptualising workshops. She expresses her concern at selecting only one text, recognising that the presentation would then be biased, when in interlingual and intercultural language learning, it becomes necessary to present and discuss on multiple perspectives. As seen in the extract in Fig. 3.1, Margie articulates her own considerations in designing the unit of work, including assessment processes, for interlingual and intercultural language learning and assessment. It must be highlighted, however, that at this stage of the study, neither the researchers nor the teachers had fully articulated a set of considerations that might be taken into account in designing assessment in this context. In line with the praxis orientation that guided this project throughout, researchers refrained from providing what they thought ought to be the necessary considerations for the teachers. Instead, it was through their collaborative work that all participants came to understand and formulate this set of considerations.

The unit of work on Female Identity involved a range of 'tasks', with an accompanying rationale for each. By inviting the teachers to provide a simple, written rationale for each assessment procedure, it became possible to access their thinking, specifically in relation to students' interlingual and intercultural language learning. This simple mechanism, designed to render teachers' thinking external and available for discussion, also invited teachers to think more deeply about the interlingual

(1) Individual written reflection in English to the following question: "What does it mean to be a woman in today's world?". This will be a starting point for assessment so that the students and I can see what their responses to this question are before any deeper exploration is begun.
(2) Class concept map to bring the different ideas together from reflection in 1. This will be conducted in English and will begin hopefully to identify some common views and also some diverse views from which to begin opening students to new ideas and ways of thinking about this topic.
(3) Summative Task—Present a speech in Chinese to the class about a female who inspires them, providing reasons and examples as to what makes this female important or respected. This will lead students to a deeper reflection on what they value/consider positive qualities in the female identity, first individually, and then as a group when the speeches are presented and the class is invited to ask questions in Chinese/English.
(4) Summative Task—Read a series of authentic texts on famous Chinese woman from the past and present. Respond in English to questions focusing on what it is that makes these women valued/respected (or even despised). This will give students the opportunity to compare and contrast what is valued in Chinese and their own cultures regarding female identities. This will also increase student confidence in accessing authentic texts.
[(5), (6), (7)—more textual work]
(8) Summative Task—Investigative Task. Using the texts discussed as a stimulus, write a reflection, as a response to a chat forum, on what does it mean to be a woman in China today. This should enable students to develop their own meanings of the Chinese context for female identity with some comparison to their own views.

Fig. 3.2 Excerpt from Margie's plan of a unit of work, including assessment procedures

and intercultural significance and value of each task. The description of some of the sequences of 'tasks' that Margie designed, and the related rationale, taken from her curriculum planning, are provided in Fig. 3.2.

A consideration of the rationale for each of the tasks as opposed to just a description of the tasks themselves further reveals some of the characteristics of the interlingual and intercultural capabilities that Margie had thought to include within the scope of her assessment of student learning. The first two tasks, for example, are intended to provide an understanding of students' prior understanding as a basis for further teaching and learning. This is not only in terms of students' knowledge, but also their assumptions, positions and values. The focus of these two tasks is naturally on student responses, both in terms of substance and language or expression. At the same time, the tasks also invite students to notice the diversity of responses offered by their peers as a resource for further classroom exploration, elaboration, explanation and mediation.

The descriptors 'summative task' and 'investigative task' given to tasks 3, 4, and 8 reflect the requirements and terminology of the assessment authority, which the teacher needs to respect. These tasks reflect the value of (1) comparison in developing interlingual and intercultural capabilities; (2) reflection; and (3) enabling students to 'develop their own meanings' in relation to their evolving understanding of the target language and culture and its use. The learning and assessment tasks and the class discussion mediated by the teacher also indicate the cultural situatedness of the exchanges. The teacher as mediator seeks to and invites her students to

- Students' individual reflections in Task 1 will be collected as examples of what students already bring to this topic in terms of their own cultural views and observations.
- Concept map from shared discussion in Task 2 will be an example of a shared starting point.
- Speech to class will be an example of students' ability to express their values in relation to female identity in Chinese; this can also be used as a starting point for me to assess their ability to reflect in Chinese but also will help them develop their own views on this in Chinese.
- Text analysis task will be an example of students' ability to analyse features of language and culture in authentic texts in the Chinese context and make comparisons with their own cultural contexts.
- Sharing of comparisons on famous female identities in Task 5 will be recorded by taking notes.
- Shared response to chat forum will be an opportunity for the class to come to a shared understanding of what it means to be a woman today and discussions to arrive at this will be recorded for further reflection and analysis. Responses to our own posting will also be printed out.
- Investigative Task English and Chinese components will be used as evidence of the journey the students have taken and the changes or conclusions they have made in their own thinking.

Fig. 3.3 Excerpt of Margie's assessment plan

connect ideas, thoughts and knowledge across repertoires of knowledge, language/s and culture/s.

As part of the planning, Margie made available to the researchers all texts and resources used to develop students' learning. These were all authentic texts (with minor modifications), selected to expressly offer different (Chinese/Australian) perspectives on 'beauty' over time, famous women, changing roles and perceptions of women, and so on. As indicated, through the ongoing individual and group discussion in the context of the study, both teachers and researchers came to understand that evidence of students' development of interlingual and intercultural capabilities in language learning could take different forms, and that this evidence would extend well beyond the final products prepared as the culmination of a particular set of experiences of language/s learning. Teachers were therefore invited to consider in advance, as part of their planning work described above, the data that they intended to collect that might yield evidence of students' developing interlingual and intercultural capabilities. An extract from a plan devised by Margie to assess her students' learning is provided in Fig. 3.3.

This example shows the dual process of analysis that was at play. On the one hand, the teacher planned and indeed used the diversity of the students themselves as a source of input into local discussions of the concept. On the other hand, the teacher also planned, collected and made available for analysis a range of data that would inform her own (and the researchers' and research participants') evolving understanding of the nature and elicitation of interlingual and intercultural capabilities in language learning in the context of the study.

The teacher and all participants in the study experienced some difficulties in analysing and judging students' work and their development of interlingual and intercultural capabilities. In the following extracts taken from her extended account in the roundtable discussion of her work, Margie reflects on seeking to identify and judge her students' language learning and development of the capabilities of interest.

> 'One of the things I guess as far as understanding that I really liked was that they started off very much, almost crediting Western influence as changing a lot of the ideas in China and it was good to see them move away from that to see that it was much more than just Western influence that while that might had some impact in some areas, there were a lot of other forces that played and they looked at Communism and a lot of other areas where they could see that, that had changed…'

She appreciated the shift in her students' perceptions of 'Western influence on social attitudes in China'. In discussion, the focus on language had been highlighted. Indeed, the researchers had explained that in some modelling of intercultural language learning, language and cultural/intercultural understandings had been separated but that in the study we were interested in their integration.

> 'But I think that most students, one of the things, that did impress me was the level of language that they can actually achieve, like it was quite amazing, some of the things, some of the sentences, some of the things they were saying in language, I was like wow, I haven't taught them this. And they were making up their own and sentences really because they obviously had something they wanted to say. So that was really good, 'cos I found their language just went up a notch, which was fantastic …'

In reflecting on the students' language, and specifically their higher level of language use, she also noted the personalisation of language and ideas in the students' responses. Both aspects exceeded her expectations.

> '… it was hard putting marks to it, I think if I had written a descriptor on this is you know, is where the student has come, this is what they've done really well at, I think it would have been easier, but having to attach a mark was really hard…'

For the purposes of assessment in the context of the assessment authority's requirements, she needed to assign grades. It was this dimension that she found particularly difficult, preferring to provide a descriptor. The difficulty arises no doubt also because of the problem of the absence of a stable frame of reference for judging interlingual and intercultural capabilities as integral to learning language/s, as she does have, for example, for L2 writing in Chinese.

> '… when I looked at the last final task, it was quite amazing with what they've actually said. That they've really come away and you know, I think (they have) come away with their own ideas; like one girl was talking about her struggle that she's a feminist and yet, you know, the fact that, there was one of the text was about choosing money rather than a man for Chinese women. And she was struggling with that. She said even though I am a feminist, you know, I want to have a partner and I don't want to, I want to be independent but, I don't want to choose money and not have a man …'
>
> I certainly got to know my girls better—where they sit, from what they value right through to their own culture, what they seem to like about some of the ideas from things they read about Chinese culture, what they liked and what they challenged, challenged them …'

Margie acknowledged the value of 'getting to know [her] girls better' and their engagement with the concept.

> 'I think it's really for me the only way I found that I could assess it was the reflection, getting them to write; even the Chinese writing text was a very reflective text. I don't think I could have set something like write a letter to your friend about this issue. It wouldn't have been the same result as something that's just purely personal like a blog'.

Finally, Margie noted the value of reflection in interlingual and intercultural language learning. As both the teachers and researchers came to understand, this is because it is through reflection that students articulate their evolving understanding of the target language and culture and its use, and through reflexivity that they consider the phenomenon of interest (here, the concept of femaleness) and its personal meaningfulness to them. In this way, they articulate their own perceptions in relation to those they have experienced (through personal/social interaction and through interaction with diverse texts) in relation to the target language and culture. In Margie's work in the second year of the study she was able to incorporate in a more deliberative way the characteristics of interlingual and intercultural capabilities and ways of eliciting and judging evidence of such language learning. And the same for all participating teachers, at least to some extent.

Extension Beyond the Focal Case: Group Facilitated Dialogue

From the collaborative experience of the first cycle of Margie's work, several shifts in assessment were made as we entered into the second cycle. These shifts, as summarized by the researchers, on the basis of their work in practice with teachers, included: (1) a shift from 'assessment tasks' to 'data points' (a coinage made during the project to capture a range of sources of data that would yield evidence of learning beyond the final production); (2) from stand-alone instruments or procedures to the totality of formal and informal judgments made in interaction; and (3) the need for a range of means for gathering evidence that includes both snapshots and developmental processes. Teachers and researchers also began to articulate features of assessment intended to capture interlingual and intercultural capabilities, including a focus on personal meanings that students make of phenomena; the need for a range of interaction experiences, representations, resources/texts and tools; and the need to consider the positioning and perspectives of students and the identities enacted. We recognised the need for different kinds of evidence and analyses that focus not only on knowledge, skill and end-products of learning, but also on embodied experience and meanings-language/s-culture/s exchanged within that experience.

Each teacher worked through the process described and illustrated above and each presented his/her narrative account through the process of facilitated dialogues. Although the account was facilitated by the researchers, the discussion of each teacher's account involved all participants. It is difficult to capture the flow of

day-long discussions by way of illustration,[1] particularly of the way in which the discussion of each teacher's account raised issues that were then picked up as leitmotifs and built upon throughout the day. Towards the end of the morning when Margie's account of judging her students' work had been discussed, along with the work of another participant, another of the teachers of Chinese (Peter) in the group sought to summarise his understanding, no doubt also as a form of reflection on his own experience of the same process. His comments were also in response to the reflections of Martin, a teacher of English as an Additional Language, who had just given an account of his work with a class that included predominantly students whose primary language was Chinese. He explains:

> Peter: I've been thinking about this through the morning and scribbling in my manila folder. … I think that we're not talking about interculturality as a sort of thing. So I've been thinking it through like this. [He draws on the white board.] I mean, the arrangements of these [dimensions] probably could vary. I thought about this for a while and this is what it felt like to me… you have their first language, which is Chinese, then you've got language 2, English and there's the interaction that goes on between those two languages. I suppose the teacher's construct in some way and then there's the accidental [perhaps meaning 'incidental'] unintended ones, like go on a website and find something. There's the learner as sort of technician, based on their existing language, language 1. That's how they think that the world is constructed. And then they had perceptions about this language 2, you know, they had opinions about it. You present it to them, 'oh hell, that's stupid, why don't they write from Chinese characters, too hard.' That's their immediate sort of reception perception. So you've got this cognitive process going on, the perceptive, affective, so, hard in the mind. And the sort of interplay between all those things can bring them into this sort of little space which is that 'otherness' that we try to sort of get at, but don't quite really understand. So it's not just intercultural, but it's a whole interplay of factors and you might find that some people, you'll see a strength in the cognitive area or others have strength in the perceptive area or a strength in the way they engage in the interaction between the two things. It gives you an insight into how they're accessing that otherness in some way. Anyway, that would be my sort of thinking of what we're trying to do.
>
> Researcher 1: Lovely
>
> Researcher 2: One of the best discussions that's been going on here is going back to Margie's point that it would be so much easier to write a descriptor than to give a mark and that seems to be your point too, Martin, that it was 'levelling' somebody that was the problem. Not necessarily saying what they've learnt, but evaluating them in some way on a scale. And I think that's one of the big problems facing assessment. But that we've got something that's sort of difficult to sort of level in the same way that more grammatically correct word forms are better than fewer grammatically correct word forms or something like that. And perhaps the problems are not the assessment of the intercultural, but the way we're called on to report the assessing of the intercultural and that's something to think about and explore as well.
>
> Researcher 1: I think it's to do more with the larger culture of assessment, rather than, you know, the reporting per se. But, there's another level of which, you know, one could say that really the number is simply a summary that we attach to an impression of a number of things that we've looked at, perceive, … So it's a number, a description. But at the end of

[1] Transcription of discussions can be made available upon request.

> the day, what's really of interest is the process that yields both of those. And ok, one's got a few more sentences and one's got the sharpness of a number, but the process that actually yields that is the process that's of interest to me. So, you know, it's how we treat them, sure, and the description is going to be… it provides an avenue for us to declare more fully what the process has been. It an avenue for externalising the process of appraisal that we're making. But that still doesn't remove the fact that we've got something in our mind of what it is we're looking for … and everything else that's embedded in it.

In this exchange, prompted by Margie's reflection, Peter is seeking to represent and articulate his way of imagining the interaction of the two languages and cultures that are at play in learning, in the immediate instance of this case of learning and assessing, which involved English as an Additional Language on the part of Chinese-speaking students. As a way of explaining the 'inter', Researcher 1 expresses her appreciation. Researcher 2 then connects back to Margie's comment discussed earlier, about the difficulty of assigning a grade or score. Researcher 1 makes the point that the grade or score is a proxy for a description of what the assessor has observed, seeking to re-emphasise the focus on evidence of learning and seeking to capture characteristics of the capabilities of interest. Through this extract, which is no more than a fragment of the ongoing discussion throughout the 2 year-long cycles of work with teachers, I signal the layers of reciprocal discussion, interpretation, elaboration and extending of understandings and meanings. The teachers sought to understand their work and that of the students, and as a collective of teachers and researchers, we sought together to understand the assessment of interlingual and intercultural capabilities through the cycle of conceptualization, enactment in practice, reflection and reflexivity as a process of theorizing the assessment of these capabilities.

Conclusion

Just as with language/s learning within an intercultural orientation, Kramsch (2006) states that it is not sufficient for students to know how to communicate meanings; they have to understand the practice of meaning making. Similarly, it is not sufficient for teachers and researchers to know how to engage in assessment; they have to understand, question, problematise and (re)theorise the processes, meanings and meaningfulness of the processes. Just as students in learning languages are invited to decentre from their own primary linguistic and cultural world to consider their situatedness from the perspective of another, so too, teachers and researchers need to decentre from their primary practices and perspectives to engage with those of others through collaborative, facilitated dialogues. This decentring process involves the dialectical relationship between reflection (thinking back on practice), reflexivity (reflecting critically on self and one's own participation), practice (what teachers do) and praxis (both practice and thinking about practice, which is informed by understandings of theory).

The process was experiential, reflective and reflexive for all participants (both teachers and researchers), as each of us sought to integrate knowledge/s, under-

standings, perspectives (both multi-paradigmatic and multi-perspectival) and values within an evolving understanding of the assessment of language learning within an interlingual and intercultural orientation. The teachers and researchers came not only to a more deliberative, conscious understanding of language learning and its assessment within this orientation, but also the complex entailments of doing so. The focus on assessment as an integral part of the process of language/s learning opened up questions of processes and evidence. These questions arose in relation to the processes or experiences and participation, as well as the meta-awareness derived from the ongoing process of analysing and reflecting on them.

The collaborative study highlights the value of long-term contextualised approaches to teacher-research and learning. The research question was one that could only be addressed through contextualised, exploratory practice 'on the ground' in the area of assessment. The interlingual and intercultural orientation to language learning presented challenges; assessment within such an orientation presented an even greater challenge. The experience of cycles of (re)conceptualising → eliciting → judging →validating, and then repeating the same, enabled relearning to take place through collaborative praxis. It is through this process that we have begun to understand assessment of language learning within an interlingual and intercultural orientation through reflexive engagement with both practice and theory.

References

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Block, D. (2014). Moving beyond 'lingualism': Multilingual embodiment and multimodality in SLA. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 54–77). New York/London: Routledge.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.

Byram, M. (2014). Twenty-five years on – From cultural studies to intercultural citizenship. *Language, Culture and Curriculum, 27*, 209–225. https://doi.org/10.1080/07908318.2014.974329.

Byrd-Clark, J. S., & Dervin, F. (2014). *Reflexivity in language and intercultural education*. New York/London: Routledge.

Byrnes, H. (2010). Revisiting the role of culture in the foreign language curriculum. Perspectives column. *Modern Language Journal, 94*, 315–336. https://doi.org/10.1111/j.1540-4781.2010.01023.x.

Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record, 104*, 1461–1484.

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Modern Language Journal, 100*(Suppl 1), 19–47. https://doi.org/10.1111/modl.12030s.0026-7902/16/19-47.

Gallagher, S. (1992). *Hermeneutics and education*. New York: State University of New York Press.

Gohard-Radenkovic, A., Lussier, D., Penz, H., & Zarate, G. (2004). La médiation culturelle en didactique des langues comme processus. In G. Zarate, A. Gohard-Radenkovic, D. Lussier, & H. Penz (Eds.), *La La médiation culturelle en didactique des langues* (pp. 109–142). Strasbourg: Council of Europe Publishing.

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.

Kramsch, C. (2004). Language, thought and culture. In A. Davies & C. Elder (Eds.), *Handbook of applied linguistics* (pp. 235–261). Malden: Blackwell Publishing.

Kramsch, C. (2006). From communicative competence to symbolic competence. *Modern Language Journal, 90*, 249–252. https://doi.org/10.1111/j.1540-4781.2006.00395_3.x.

Kramsch, C. (2009). Discourse, the symbolic dimension of intercultural competence. In A. Hu & M. Byram (Eds.), *Interkulturelle Kompetenz und Fremdsprachliches Lernen. Modelle, empirie, evaluation* [Intercultural competence and foreign language learning: Models, empiricism, assessment] (pp. 107–122). Tűbingen: Gunter Narra Verlag.

Kramsch, C. (2010). *The multilingual subject*. Oxford: Oxford University Press.

Kramsch, C. (Ed.). (2014). Teaching foreign languages in the era of globalization. *Modern Language Journal, 98*(Special issue 1), 296.

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice and change. *Language Assessment Quarterly, 1*(1), 19–41.

Leung, C., & Scarino, A. (2016). Reconceptualising the nature and outcomes in language/s education. *Modern Language Journal, 100*(Supplement 1), 81–95. https://doi.org/10.1111/modl.123000026-7902/16/81-95.

Liddicoat, A. J., & Scarino, A. (2013). *Intercultural language teaching and learning*. Malden: Wiley-Blackwell.

McNamara, T. (2003). Tearing us apart again: The paradigm wars and the search for validity. *EUROSLA Yearbook, 3*(1), 229–238. https://doi.org/10.1075/eurosla.3.13mcn.

Moss, D. M., Osborn, T. A., & Kaufman, D. (2008). The promise of interdisciplinary assessment. In D. M, T. A. O. Moss, & D. Kaufman (Eds.), *Interdisciplinary education in the age of assessment* (pp. 1–6). New York/London: Routledge.

Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal, 91*(3), 323–340. https://doi.org/10.1111/j.1540-4781.2007.00583.x.

Poehner, M. E. (2011). Validity and interaction in the ZPD: Interpreting learner development through L2 dynamic assessment. *International Journal of Applied Linguistics, 21*(2), 244–263.

Poehner, M. E. (2013). Dynamic assessment in the classroom: Approaches and development. In *The companion to language assessment* (Part 6: Assessment and Learning). Wiley online. https://doi.org/10.1002/9781118411360.wbcla033.

Purpura, J. E. (2016). Second and foreign language assessment. *Modern Language Journal, 100*, 190–208.

Scarino, A. (2010). Assessing intercultural capability in learning languages: A renewed understanding of language, culture, learning and the nature of assessment. *Modern Language Journal, 94*(2), 324–329. https://doi.org/10.1111/j.1540-4781.2010.01026.x.

Scarino, A. (2014). Learning as reciprocal, interpretive meaning-making. A view from collaborative research into the professional learning of teachers of languages. *Modern Language Journal, 98*(1), 386–401. https://doi.org/10.1111/j.1540-4781.2014.12068.x.

Scarino, A. (2017). Reconsideration of the distinctive role of heritage languages in languages education in Australia. In P. Trifonas & A. Themistoklis (Eds.), *Handbook on research and practice in heritage language education* (pp. 465–478). Dordrecht: Springer.

Scarino, A., & Liddicoat, A. J. (2016). Reconceptualising learning in transdisciplinary languages education. *L2 Journal, 8*(4), 20–35.

Sercu, L. (2002). Autonomous learning and the acquisition of intercultural competence: Some implications for course development. *Language, Culture and Curriculum, 15*(1), 61–74.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.

Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. London: Routledge.

Shohamy, E. (2011). Assessing multilingual competences: Adopting construct valid assessment policies. *Modern Language Journal, 95*, 418–429. doi:https://doi.org/10.1111/j.1540-4781.2011.01210.x.0026-7902/11/418–429.

Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Baneerjee (Eds.), *Handbook of second language assessment* (pp. 255–272). Boston: De Gruyter Mouton.

Chapter 4
Trajectories of Language Assessment Literacy in a Teacher-Researcher Partnership: Locating Elements of Praxis Through Narrative Inquiry

Luke Harding and Tineke Brunfaut

Abstract In 2011, under the auspices of the Luxembourg Ministry of Education, but initiated and steered by a group of teachers, a researcher-teacher partnership was set-up between Lancaster University (UK) and a team of secondary school teachers in Luxembourg. The aim of the partnership was to (a) explore the potential for redesigning the national end-of-secondary-school English exam to ensure alignment with current approaches to language teaching in the classroom, and (b) help to develop the teaching team's language assessment literacy, and their capacity to carry out high-stakes language test development work. The partnership provides fertile ground for exploring the concept of "praxis" and its relationship to current understandings of language assessment literacy (LAL). This chapter will explore these issues through narrative inquiry; specifically, an analysis of narratives produced by two teachers and two researchers reflecting on their experiences of the project over the past 6 years. Through a discussion of narrative excerpts, we will demonstrate how narrative inquiry can provide evidence of trajectories of language assessment literacy over time, as well as reveal relations between key characters, and identify complicating factors within overarching plots. The chapter will conclude with a reflection on the usefulness of narrative inquiry as a method for exploring a praxis perspective on language assessment literacy.

Introduction

Following a period during which language assessment literacy (LAL) hovered at the fringes of language testing and assessment research, the area has now become a vibrant sub-field. Since the landmark publication of Inbar-Lourie's (2013) special

L. Harding (✉) · T. Brunfaut
Department of Linguistics and English Language, Lancaster University, Lancaster, UK
e-mail: l.harding@lancaster.ac.uk; t.brunfaut@lancaster.ac.uk

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_4

issue of *Language Testing*, a growing number of papers have investigated teacher assessment literacy across geographical and pedagogical contexts (e.g., Baker and Riches 2018; Hamp-Lyons 2017; Lam 2015; Vogt and Tsagari 2014; Xu and Brown 2017). Methods used in language assessment literacy have typically relied on questionnaire and survey approaches (Fulcher 2012; Kremmel and Harding 2019), although a more eclectic methodological approach has begun to emerge (e.g., Villa Larenas 2017) in tandem with a general shift in the field towards more contextualised investigations which require holistic understandings of teaching and learning contexts (see for example Scarino, this volume). One particular method is narrative inquiry, a term that applies to a wide range of approaches in practice, all of which focus on the stories told by those who have lived them. Narrative inquiry is still an uncommon method in language testing and assessment generally (Cumming 2004), and has seen only few applications in language assessment literacy research (Baker and Germain, this volume; Xu and Liu 2009). However, narrative inquiry provides great promise for tracing the development of language assessment literacy over time, and for exploring convergences and divergences between different narrators over the course of collaborative projects.

In this chapter, we will explore the utility of narrative inquiry as a method for tracing the development of language assessment literacy within a teacher-researcher partnership. Specifically, the chapter will focus on an ongoing English language exam reform project in Luxembourg in which we (the researchers) collaborated with a team of teachers by delivering language assessment literacy training and providing ongoing consultancy on test development work. We will draw on the concept of praxis to frame the study, focusing on two elements central to theories of praxis within educational research as explained in the introduction to this volume (Poehner and Inbar-Lourie; see also Lantolf and Poehner 2014): (1) that praxis is *iterative*, with teachers' emerging expertise playing a central role in a transformative process whereby theory informs practice and practice informs theory; and (2) that praxis connects with the concept of *informed action*, where educators function as agents of change in establishing more ethical practices in their teaching contexts.

The chapter is divided into six sections. The first has introduced the general aims of the chapter. In the second section, we describe the context of the wider study: an exam-reform project in Luxembourg. The third section will outline potential connections between language assessment literacy in an exam-reform scenario, and the concept of praxis. The fourth section will present narrative inquiry as a method, describing different approaches to narrative inquiry in language education research. In the fifth section, we provide illustrative examples of narrative inquiry analyses as applied to two teachers' narratives, and two researchers' narratives. Finally, we will reflect on the usefulness of narrative inquiry as a methodological approach for exploring language assessment literacy, and suggest ideas for further research.

Context – The Luxembourg Exam-Reform Project

In 2011, we were asked to form a researcher-teacher partnership between Lancaster University (UK) and a team of secondary school teachers of English in Luxembourg. The focus was to be on re-designing the national end-of-secondary-school English exam. Four years earlier, in 2007, the announcement of educational reforms in Luxembourg had signalled a move from a traditional grammar- and literature-oriented English secondary school curriculum to a more competence-based, communicative approach (MEN 2011a, b). Despite a gradual shift in classroom assessment towards competence-based testing (especially at the lower years of secondary school), however, the school-leaving examination for English remained largely unchanged in the years following the reform. As well as being misaligned to the new competence-based approach, subsequent research demonstrated limitations in terms of the exam's construct, equivalence across administrations, and scoring reliability (Brunfaut and Harding 2015, 2017). It is within this context that the Test Design and Evaluation (TDE) team was set-up: a collective of nine teachers working within the Luxembourg secondary school system. While the project was conducted under the auspices of the Luxembourg Ministry of Education, it was the TDE team who first contacted the authors at Lancaster University to explore the possibility of working collaboratively on the exam reform.

The aims of the researcher-teacher partnership went beyond planning the redesign of the end-of-secondary-school English exam. It was also agreed that we would concurrently develop the teaching team's language assessment literacy, and specifically their capacity to carry out high-stakes language test development work (see Brunfaut and Harding 2018). Given most teachers' lack of formal training in language testing, as well as the high-stakes nature of the school-leaving exam, an early decision was made – negotiated with the Ministry – to first work on the development of a lower-stakes English test for lower secondary school (the *Épreuve Commune* [EC]) as the teachers honed their skills. The EC would constitute only one of the many assessments contributing to lower-secondary students' end-of-year marks, and thus formed "a low-risk 'sandpit' in which to try-out the team's emerging expertise as test developers, and provide space for the TDE team to gain expertise in setting up and managing a national test system" (Brunfaut and Harding 2018).

Thus, during the first 3 years of the partnership, the teachers completed a blended training programme, taught by the researchers, and applied their knowledge directly in the development of the EC. Modules were partly adapted from the distance MA in Language Testing programme run through Lancaster University and partly developed from scratch according to the team's needs. Alongside, and following, each training module, the teachers performed concrete test development work on the EC (Brunfaut and Harding 2018). The teacher-researcher collaboration was managed through the online learning platform, via e-mail exchanges, over Skype meetings, and most importantly, in site visits to Luxembourg. An overview of the modules and parallel development tasks is provided below:

- **First year:** the teachers were offered a solid introduction to language test construction and evaluation (covering topics such as test specifications; evaluating and revising test items; the constructs and assessment of the four skills; item analysis; descriptive statistics; and reliability), wrote a background document on their educational context, set up a team and project management system, and drafted a test cycle, test specifications and a first set of test items for the EC – all in consultation with the researchers.
- **Second year:** the second module focused on broader concerns in language testing (historical views of validity and validation, standard setting, more advanced item writing skills, and the constructs and assessment of grammar and vocabulary). Simultaneously, the researcher-teacher partnership concentrated on developing the lower-secondary test and completed a full test development cycle (for more details, see Brunfaut and Harding 2018). This also included site visits by the researchers to conduct item moderation and standard setting with the teachers, and administration of the test in schools.
- **Third year:** the third module involved a course on theoretical, methodological and ethical issues in language testing research and development, and a second full run of the test development and administration cycle of the lower-secondary test.

As described in more detail in Brunfaut and Harding (2018), in the years following the initial three-year training, the researchers gradually stepped back to the point that the teachers fully managed the entire EC test cycle on their own (with the researchers only consulting on pilot data results). At that point, the partnership's attention began to shift to the original aim of a redesigned English school-leaving exam, which is the current focus of the collaboration.

Praxis and Language Assessment Literacy

The teacher-research partnership described above provides fertile ground for exploring the concept of "praxis" (Lantolf and Poehner 2014) and its relationship to current understandings of language assessment literacy. Language assessment literacy has been defined variously as "the knowledge and skills required for performing assessment-related actions" (Inbar-Lourie 2012, p. 1), "a repertoire of competences that enable an individual to understand, evaluate and, in some cases, create language tests and analyse test data" (Pill and Harding 2013, p. 382) and, most comprehensively, by Fulcher (2012, p. 125):

> The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals.

Harding and Kremmel (2016) make the point that Fulcher's definition is not only highly-detailed, but also a summation of the ultimate expertise that might be required of a language assessment specialist, involving knowledge of practice, theoretical principles, and broader social and ethical frameworks of language assessment use. We can therefore conceptualise LAL as a type of expertise in language assessment principles and practice, which may develop to become deeper and more comprehensive over time.

In the context of the Luxembourg exam reform project, the notion of expertise was focused specifically towards the ability to develop high-stakes, large-scale language examinations which would address the problems identified with existing assessment provision. The aim of the teacher education component of the partnership was to increase knowledge, skills, and confidence among the teacher group with respect to language assessment, as well as to work towards a tangible product in the form of the lower-stakes EC. At the same time, the researchers' growing understanding of the unique educational-policy context of Luxembourg, and the experience of seeing theoretical or research-based approaches constrained in practice, would lead to more tailored advice through the consultancy phase, and ultimately a deeper language assessment literacy among the research team. We might therefore characterise the approach to language assessment literacy training in this project as a praxis approach.

The concept of praxis has been covered extensively elsewhere in this book, however in this chapter we draw on the broad definition of praxis provided by Lantolf and Poehner (2010) as "the integration of theory with practice … whereby theory provides a basis to guide practical activity, but at the same time practice informs and shapes theory" (p.12). Lantolf and Poehner elaborate on this symbiotic relationship for teachers: "praxis does not position classroom teachers merely as consumers of research – and here we might add consumers of test scores and other outcomes from formal assessment procedures – but recognizes their expertise as central to the iterative development of theory and practice" (2010, p.16). Whereas Lantolf and Poehner draw heavily on the work of Vygotsky (2004) in conceptualising praxis, the term has resonance through other traditions of critical educational theory, particularly the work of Freire (1970) whose notion of praxis as *informed action* which is transformative (see Johnson 2006) has particular relevance for conceptualising the development of teachers' language assessment literacy within the context described above. In this sense, our view of praxis is broadly aligned with the Critical Tradition (see Introduction – this volume).

The notion of praxis as "informed action" has its roots in the Aristotelian concept of praxis as a disposition "to act truly and rightly". In other words, praxis has an ethical and moral dimension: combining knowledge of theory with pedagogical practice allows teachers to function as agents of change; forces for social good. This line of thought may also be interpreted in the work of Freire where, as Mulcahy, Mulcahy and Mulcahy et al. (2015) have argued, "Freire understood praxis and transformation as a force for good, not only for naming and changing the world but in transforming it for the better – for the betterment of self and others" (p.152). This moral and ethical dimension of praxis has clear connections with the ethical practice

required at the highest levels of expertise in various conceptualisations of language assessment literacy (see Fulcher 2012; Pill and Harding 2013). Within this view, enhancing language assessment literacy through developing a more research-based and theoretically-robust understanding of assessment practice, with the ultimate goal of effecting change within the broader structures of language education policy in Luxembourg, would embody the principle of informed action.

In contexts such as the current project, a key challenge remains how to evaluate the effectiveness of the teacher-researcher partnership in fostering the development of LAL expertise. A praxis perspective helps to articulate more specific questions: whether the collaboration process has enabled the iterative nature of theory informing practice, and practice informing theory, to take place, and whether a trajectory towards informed action can be observed throughout the course of the project. However, more traditional methodologies for investigating LAL – questionnaires/surveys and observations – may not be well-suited for locating instances of praxis as such methods lend themselves less to a developmental perspective. Although these traditional methods might be implemented longitudinally, they may not provide insight into meaning-making *as it is taking place*; to glimpse changes in thinking/understanding as and through their being brought into discourse. Narrative inquiry, however, is well-suited to investigating phenomena that are changing during the investigation. Further, narrative inquiry provides a method for exploring the positioning of the teacher (or researcher) in relation to other key agents in the project (both individuals and "bodies"), and to the knowledge base itself. Narrative inquiry, therefore, provides a promising alternative for evaluating LAL from a praxis perspective.

Narrative Inquiry

Narrative inquiry as a research method is concerned with stories told by people. The basis for the approach is the ubiquity of narrative as, according to Barthes (1977), "narrative is present in every age, in every place, in every society" (p.79). However investigating narrative provides not only a way of understanding the content of individuals' life stories, but also yields insights through *analysis* of how individuals present themselves through stories: what they include and exclude in their narratives, how they characterise themselves and others, and how they construct reality through their storytelling. In this sense, according to Webster and Mertova (2007), "[narrative] provides researchers with a rich framework through which they can investigate the ways humans experience the world depicted through their stories" (p.3). At the same time, narrative inquiry is not a single method. It spans a diverse range of approaches from more structuralist approaches to narrative analysis (e.g., Labov 1997) to highly interpretive accounts. For this reason, Chase (2008) describes narrative inquiry as, "an amalgam of interdisciplinary analytic lenses, diverse disciplinary approaches, and both traditional and innovative methods – all revolving

around an interest in biographical particulars as narrated by the one who lives them" (p.58).

Narrative inquiry is not a method commonly found in language testing and assessment research. The dearth of research utilising this method led Cumming, in 2004, to note, "I am at a loss to explain why I cannot think of a single application of narrative inquiry … to examine teachers' beliefs about language assessment" (2004, p. 9). A survey of available literature since that time suggests that narrative inquiry has still received only scant attention in language testing and assessment research, mostly in investigations of teachers' assessment practices (e.g., Xu and Liu 2009, though see also Baker and Germain, this volume). This is despite narrative inquiry holding great potential for exploring, for example, the biographies of test-takers, raters, score users and a range of other stakeholders.

Narrative inquiry has, however, become a more popular method within the field of TESOL and Applied Linguistics broadly (e.g., Barkhuizen 2008, 2010, 2011; Duff and Bell 2002; Pavlenko 2002). Barkhuizen, Benson and Chik (2014), for example, have pointed towards a "narrative turn" which is premised on several advantages of narrative approaches, one being that narrative research can "both involve and empower the people whose experiences are the subject of research" (Barkhuizen 2016, p. 29). In his own work, Barkhuizen has drawn on positioning theory (particularly the work of Bamberg 1997) to explore how language teacher and learner identities are constructed through narratives. Indeed, narrative inquiry has been particularly well-suited to exploring issues of identity in language pedagogy, allowing – as it does – for an emic research perspective, and for the researcher to analyse the narrator's construction of self, and of other characters, both in relation and over time. Within a praxis orientation, however, narrative inquiry could move beyond helping researchers to understand different perspectives temporally, and could be an important tool for helping to understand *how* changes are occurring and, perhaps, even function as part of what might bring about change.

The eclectic approach to conducting narrative inquiry has led to a wide diversity of analytical approaches. An in-depth overview of these analytic positions is provided in Pavlenko's (2007) paper, which argues for a more linguistic approach to narrative inquiry. Pavlenko, in observing the increasing number of narrative inquiry studies in TESOL research, notes that the largest group of studies utilising narrative inquiry in language education is focused on investigating "subject reality" in participants' stories through thematic or content analysis. Pavlenko concedes that the strengths of this approach include that "recurrent motifs" in stories are highlighted. However, she also argues that the most problematic aspect of the content/thematic approach is that the linguistic means by which "storytellers use language to interpret experiences and position themselves as particular kinds of people" (p.167) is not given sufficient attention, and should be of prime importance to applied linguists working with this methodology. Pavlenko argues that narrative inquiry analysts should instead focus on the interplay between content, context and form. These three levels are described in Table 4.1.

Table 4.1 Analysing autobiographical narratives

Content	Consideration of what is said/written as well as what is excluded from the narrative
Context	Macro-level: "Historic, political, economic, and cultural circumstances of narrative production" (p. 175)
	Micro-level: Local context of narrative production (e.g., modality, audience, language choice)
Form	Macro-level: Narrative structure / genre (e.g., private journals vs. elicited tasks)
	Micro-level: Multiple "tools" (p. 179) – Most prominently **positioning approach**: Reflexive positioning (positioning self) vs. interactive positioning (positioning others), through "morphosyntactic and lexical choice, image and metaphor" (p. 179)

From Pavlenko (2007)

Pavlenko (2007) is careful to note that all three levels are "interdependent" such that choices of form, for example, cannot be understood without consideration of macro- and micro-context.

As the aim of this chapter is to explore the utility of narrative inquiry as a method for tracing the development of language assessment literacy within a teacher-researcher partnership, we now turn to describing a dataset of narratives we collected, and then to analysing excerpts of data through a narrative inquiry approach, drawing on a range of elements from Pavlenko's (2007) framework as required.

Data

As part of ongoing data collection around the Luxembourg exam-reform project, and specifically the effectiveness of the teacher-researcher partnership, we asked teachers on the TDE team to write narrative accounts of their experiences on the project thus far (i.e., since 2011). We were specifically interested in the teachers' recollections of the beginning of the project, critical incidents or stories they were able to share, their perceptions of any transformation (or not) in their professional practice, and their views of the teacher-researcher partnership over time. The instructions we developed for writing the narratives are provided in Fig. 4.1. These were developed from guidance on "collecting stories" in Webster and Mertova (2007).

Although less structured instructions for participants might have been preferable, one of the reasons for providing this level of scaffolding was to avoid adding to the already very heavy workload of the team. Teachers in the TDE team were only granted 1 day of non-teaching time per fortnight to focus on the exam reform project, and this time was typically consumed with test development and communication activities (e.g., planning presentations). In fact, at the time of writing we were only able to gather two narratives from teachers, despite several group members contacting us to say they would like to be involved in the narrative research project. The team simply did not have sufficient time for individual reflection of this nature, a point which will be taken up in the Conflict & Challenge section below. In our report on the analyses, we use the pseudonyms Charlie and Dominique to refer to the two teachers who provided narratives.

Instructions – Writing your story

We would like to ask you to write the story of your experience so far working on the TDE team. Your story can be short (e.g., 2 pages) or as long as you like. You can also write in any style you like so you have free reign to be creative! We would like to emphasize that this is about *your* experience; there are no rights or wrongs in this and it is about *your* personal story.

The particular story you tell is entirely up to you, but you may wish to consider the following as possible areas to cover:

1. Thinking back to the beginning of the TDE project, what memories do you recall?
2. What memories do you have of the Lancaster-taught modules?
3. What memories do you have of the early stages of the *Épreuve Commune* development?
4. If there was one main memory of the project over the past 6 years, what would it be? What role did others play in this event?
5. Within the context of the project, do you remember a particularly stressful period? Do you remember a period when you felt particularly happy? What role did others play in this event?
6. How, if at all, do you think your involvement in this project has changed you?
7. How, if at all, do you think your involvement in this project has changed your views on or practices in language testing and assessment (and/or teaching)?
8. How have you experienced being part of researcher-teacher partnership?

Fig. 4.1 Narrative writing instructions

As well as asking for narratives from the Luxembourg teachers, we (the researchers) also committed to writing our own narratives – referred to as Chris and Sam in the sections below. We wrote our narratives before we read any other narratives which had been provided by the teachers, and without communicating the plans for our narratives with each other. We acknowledge the limitations of performing an analysis on our own narratives given that they were written for the purposes of telling the story of a project we were involved with. On the other hand, we hold a unique position in being able to interpret and analyse the content and context of the narratives. It is believed that the choice to use narrative analysis for both teachers and researchers carries additional weight in praxis-oriented endeavours, as they strive for research symmetry and challenge the traditional expert/practitioner positioning (Poehner and Inbar-Lourie, Chap. 1, this volume). We have attempted to be as transparent as possible in producing lengthy extracts of narrative so that the reader can judge the trustworthiness of our analysis.

The analysis of the four narratives will focus on content and form in these texts. Connections will be made to the macro-context (described in the section on the Luxembourg exam-reform context above). It is also worth noting that the micro-level context of the writing task will have influenced the style of narrative that is produced. For example, the written modality may have given rise to a story which

had a more formal and organised structure than would be the case with a spoken narrative. Second, the genre of the task encouraged narrators to provide a clear beginning, middle and end. As such, the narratives may have been constrained by conventions of typical story structures. Finally, the audience of the narratives needs to be taken into account. Charlie and Dominique were aware that their audience were the project researchers (as well as a wider readership, as explained in the ethics information provided to participants); as such, the teachers may have been less forthright concerning any problems with the research team than would be the case in an "off-the-record" narrative. As for the researchers, Chris and Sam, our main audience – other than the readers of future research publications – was each other. Again, this is likely to have constrained full revelation of any difficulties experienced on the project. It is important to note, however, that the researcher narratives were written before the researchers had conducted a broader literature review on methods of analysis in narrative inquiry. Thus the analytical framework adopted in this study was not known by the researchers at the time the narratives were produced.

Analyses

In the section below, we will provide three analyses to draw out some applications of narrative analysis for the purposes of locating elements of praxis and LAL development. We will draw data from the two teacher narratives (Charlie and Dominique) and the two researcher narratives (Chris and Sam), with one narrative – Charlie – analysed at each point to demonstrate continuity. The analyses will provide a combination of broader interpretive analysis, and micro-level analysis of discourse (specifically focusing on positioning achieved through speaker roles, characterisation and the use of metaphor, as these emerged as salient features of the discourse in our initial application of Pavlenko's analytical framework). In each case, teacher and researcher perspectives will be juxtaposed to reveal convergences and divergences. The three analyses will be framed around the broad themes of Beginnings, Teacher-Researcher Collaboration, and Conflicts & Challenge.

Beginnings

Extract 4.1 below is drawn from the very start of Charlie's narrative:

Extract 4.1: Charlie

I was first approached about joining the Test Design and Evaluation (TDE) project barely 2 years after completing my teacher training programme. In my first few years of teaching, it had already become clear to me that the existing assessment tools and procedures in the Luxembourg school system, and particularly the school-leaving exam in English, presented various limitations and problems, and thus needed substantial improvement. Nevertheless, I had certainly not expected to be asked to become part of such an ambitious and potentially

> far-reaching project so early on in my teaching career; in truth, I was very aware of the fact that I only had a limited amount of professional experience to draw on in comparison to numerous other teachers (who would ultimately be the ones to respond to – and potentially criticise – the outcomes and products of the TDE project). I was thus certainly honoured, yet also slightly apprehensive and overwhelmed, when asked to help set up the TDE project in the summer of 2011.
>
> There were several factors that ultimately convinced me to take on this challenge. First, there was the above-mentioned perception that the English exam needed to change, to provide the final-year students in Luxembourg with a more adequate opportunity to demonstrate their English language proficiency, and to ensure that the school-leaving diploma they received would continue to grant them access to the university courses and/or professional careers of their choice. Second, it was virtually impossible not to be inspired by the vision and enthusiasm of Ms A., who, in her role as president of the curricular board for English language teaching, had fought with admirable effort and unwavering determination for this project to take shape and to get off the ground. Finally, there was the reassuring knowledge that this would be a team effort, in which I would be able to join forces with several other local teachers from various schools and educational streams, as well as language assessment experts from Lancaster University; this certainly helped to make this venture less daunting than it had first appeared to be.

Charlie's narrative presents an extended description of the beginnings of the project from a teacher-participant perspective. Thematically, the beginning is quite typical of the beginning of any collaborative project: a team is formed in response to a specific need (in this case, the need for a reform to the existing final-year school-leaving exam). As discussed earlier in the chapter, the existing approach to the design and administration of the school-leaving exam raised issues concerning fairness and consistency of marking, assignation of topic, and procedures across schools. In a broad sense, this opening to Charlie's story sets the scene for the "action".

A micro-level analysis of Charlie's extract, however, reveals several interesting features at the level of discourse. In terms of self-positioning, it is clear that Charlie identified as a novice teacher at the beginning of the project. This is evident in several examples within the text in which Charlie emphasises a lack of experience: "*barely two years* after completing my training programme", "I had certainly not expected to be asked … *so early on* in my teaching career". At the same time, Charlie positions the Luxembourg language examinations as unsatisfactory, even from the perspective of a novice, as they presented "various limitations and problems, and [were] in need of substantial improvement". Within this context, Charlie is portrayed as a character who is effectively drawn in to the story, as indicated by the repeated use of agentless, passive constructions: "I was first approached", "when asked to help set up the TDE project".

Charlie did not enter the project unwillingly, however. There are three key figures/groups referred to at this early stage of the project: Ms A., Charlie's future fellow team members, and the Lancaster University collaborators. The first, Ms A. (a key project originator), is described in highly-positive terms, "it was virtually impossible not to be inspired by the vision and enthusiasm of Ms. A". Through the use of a battle metaphor, Ms A. is further characterised as a staunch combatant who "fought with admirable effort and unwavering determination". While the foe of Ms

A. is unnamed, this must be understood to implicitly refer to the status quo for secondary-school examinations (as described earlier in the chapter). Lancaster researchers, meanwhile, are described as "language assessment experts", suggesting an initial divide between Charlie's initial self-perception and the portrayal of the research team. The battle metaphor is extended with respect to "joining forces" with fellow team members who set out on a potentially perilous journey, "helped to make this venture less daunting".

These initial passages therefore provide us with a clear insight into Charlie's initiation into the project: a novice teacher who was selected to join a team aiming to transform the Luxembourg examination system. The initial characterisation of the team as potential agents of change here is clearly evoked, and the key "allies" are identified.

The entrance of the researchers into the story is not detailed by Charlie, but Chris (one of the researchers) provides a complementary perspective in a story of the beginnings of the project in Extract 4.2.

Extract 4.2: Chris

> The very first contacts on what later became the TDE project are some of the ones I often think back to when reflecting on this project. It was facilitated by a consultant (Person B) who had been working on a reform of ELT in Luxembourgish secondary schools with English secondary school teachers and the Ministry of Education. The e-mail I received was essentially a call from a team of teachers (some of them Ministry associates; also including the consultant) to develop professional expertise in language testing in the country. The message expressed concerns about the tests and testing approaches in operation at the time, and about the risk of teaching and testing going into separate, potentially conflicting directions in the reform. The message clearly hinted that it was teachers who were driving the request for collaboration, which sparked my interest. After an 'in principle' expression of interest to explore the idea, and some time passing by, I received a more formal and concrete invitation for collaboration from one of the teachers, Ms A. I remember her e-mail as expressing enthusiasm and eagerness, as well as seriousness, determination, and awareness of challenges and political sensitivities. Two things in particular made the message stand out, I thought. One was that she had already spoken to the Minister of Education (I remember thinking: "pretty amazing to get a one-on-one appointment with a Minister!") and to the teacher education faculty at the University of Luxembourg (one of the professors later visited Lancaster and enabled the planning of the project). Essentially, the waters had been tested with key players (and won over, it seemed). The other thing that rose my interest was that the teacher's message contained a concrete list of aims, with the main ones being a new approach to language test design and teacher capacity building in language testing. The Skype meetings that followed with Ms A. also made clear that the learners and the maximising of their language learning formed the driving force – not in a naïve sense, but in a genuine manner. All of this signalled to me that this was a bottom-up driven project with some careful thinking and strategizing behind it already.

In a similar way to Charlie, Chris provides a vivid account of being drawn into a project that already had momentum. Chris self-positions, initially, as an outsider to the context (receiving a "call" which sparked interest). Also, resembling Charlie's narrative, Chris positions Ms A. as a key player in this early period as enthusiastic, eager, serious, determined and aware of challenges and political sensitivities. Although there are no metaphorical allusions to battling for change, Chris empha-

sises that the ethical nature of the work was one of the most attractive elements in taking part in the project: "language learning formed the driving force". The team were characterised at this stage as genuine, careful and strategic. We therefore see echoes here of Charlie's depiction of the teachers as future agents of change. Yet while Charlie characterised the starting team as less experienced (presumably with respect to their level of language assessment literacy), Chris characterised the team as highly-competent in other aspects of project management.

The contrasting narratives help to establish a baseline understanding of each participants' emic conceptualisation of their initial character role: Charlie as the naïve practitioner who has joined a team which has a mission to change the exams; Chris as an interested outsider who has received a call to help the team. The different trajectories of LAL/praxis, from these starting points, can be observed in the next section.

Teacher-Researcher Collaboration

Extracts in this section are related, specifically, to the teacher-researcher collaboration as it shifted over the course of the project, and to the concomitant trajectories of LAL as the iterative nature of praxis emerged. Extract 4.3 shows Charlie's perceptions of the teacher-researcher partnership.

Extract 4.3: Charlie

> Another truly rewarding element of the TDE project has been the way in which it has allowed me to learn about fundamental theoretical concepts and practical measures relating to standardised testing and assessment. The corresponding modules that we were taught by Researcher A and Researcher B in the first 3 years of the project were both eye-opening and enjoyable; in fact, having to do the resulting "extra work" on top of an already busy schedule did not generally feel like a huge burden. On the contrary: to my mind, this research-based, academic work complemented my daily classroom routine in an interesting and inspiring way (at least, that is the overall impression I have kept over the years – had I been asked about it five or 6 years ago, sitting in front of my computer screen to complete an online assignment on a sunny Sunday afternoon, my opinion might have differed considerably!). Importantly, the concepts and procedures encountered in these modules did not just help to guide the larger-scale projects of the TDE team, such as the standardised "Épreuve Commune" that we subsequently developed for lower-level classes and that has been implemented on a national scale for several years now. Instead, I could also draw on what I had learnt in the modules to improve my daily teaching practice, by consciously trying to increase the validity and reliability of my tests, for instance, and even passing on that knowledge and know-how to student teachers, in workshops that I led together with another member of the TDE team.

At a general level, Charlie's narrative illustrates a perception that the training stage was highly useful. The modules are described as "eye-opening", suggesting a consciousness raising element to the input, and the influence of more theoretically-oriented material on Charlie's practices is described as "interesting and inspiring", suggesting that the use Charlie made of this knowledge led to changes in practice.

One unanticipated consequence was that the benefits Charlie drew from the modules were not only related to the immediate needs of the development of the EC (described above as the "testing ground" for the TDE team's burgeoning test development skills), but for classroom teaching as well. It is clear, therefore, that at a content level, Charlie held a positive evaluation of the input of the research team in the training phase.

With respect to story development, Charlie's narrative also signals a shift from the self-positioning observed at the beginning stage. In Extract 4.3, Charlie is self-positioned first as a learner "sitting in front of my computer screen to complete an online assignment on a sunny Sunday afternoon", then as a test developer – "the standardised 'Épreuve Commune' that we subsequently developed for lower-level classes", and towards the end of the passage, as a conduit of new knowledge and skills through cascaded assessment literacy training "even passing on that knowledge and know-how to student teachers". In this trajectory we can also see Charlie's increasing agency and expertise reflected in the more pronounced use of active voice with first-person and possessive pronouns: "*I* could also draw on what *I* had learnt in the modules to improve *my* daily teaching practice, by consciously trying to increase the validity and reliability of *my* tests" [emphasis ours].

It is useful to juxtapose Charlie's perceptions with those of Dominique. While similarly positive, Dominique's narrative is framed as a rebuttal to a pre-supposition that a teacher's work is too far removed from research for the latter to be useful (Extract 4.4).

Extract 4.4: Dominique

> My overall experience of our researcher-teacher partnership is extremely positive. It has made me realise that we can all gain from each other's knowledge and insights and that research is not an aloof, far removed realm that one cannot possibly hope to be involved with. It has also made me realise that one can gain important data from even seemingly trivial in-class experiments as long as one tries to follow and implement the appropriate procedures.

In this passage, Dominique uses metaphors of proximity to illustrate a stereotypical pre-conception of research as "an aloof, far removed realm" from teaching, but appears to indicate that these two realms can be brought closer together if the "appropriate procedures" are followed. For Dominique, the project seems to have demystified research and brought a research-based approach into line with the routine work of the classroom teacher. We can see in both Charlie's and Dominique's accounts shifts in personal trajectories towards greater expertise, as well as shifts in perceptions of the proximity of research-based knowledge to classroom practice.

An associated passage on the teacher-researcher partnership from Chris's researcher perspective is shown in Extract 4.5.

Extract 4.5: Chris

> A considerable part of my regular academic role involves language assessment literacy training. This typically involves the development of skills-based know-how in trainees, and often also of developing a more critical take on assessment – the so-called narrow and broad view of language assessment literacy training, respectively. This was also what we aimed

> for and implemented in the modules we ran with the TDE teachers. However, the TDE project gave us an opportunity to go beyond that; it constituted the first opportunity (for me at least) in which we were able to introduce theoretical concepts and principles as part of an actual test development project – something that has been advocated by Glenn Fulcher. We ran the modules in parallel with the development of the lower-year test (Épreuve Commune), and we chose the module content and adapted it according to the relevant test development stage. Reflecting on this, I feel this worked really well. It allowed the theory to be made more concrete and to put it into practice, thereby helping the teachers to develop a good understanding of theory, see its relevance, and be able to operationalize it directly and in a very tangible manner. So, capacity building of the teachers happened in a theoretical and practical manner at the same time, with the benefit that there was concrete output to share with key stakeholders such as the Ministry of Education and other language teachers. The latter, I have come to realize, is really significant to ensure continued support from stakeholders and gain and maintain credibility when a project is bottom-up driven.

Chris also provides a very positive evaluation of the process from the trainer/consultant's perspective as an example of theory being applied in practice as the basis for LAL development. Reference is made to Fulcher (2012) who has advocated for the development of practical skills in language assessment in tandem with theoretical training. This was, essentially, a praxis-driven approach – integrating theory and practice simultaneously, and allowing the teachers to draw meaning from application, and use application to further their understanding of theory.

At the same time, we can also see a clear "sub-plot" of the development of the researcher throughout this passage. Chris states that the project "gave us an opportunity" to go beyond the more routine LAL training and employ a scaffolded, practice-based approach. In a similar way to Charlie in Extract 4.1, Chris self-positions as a novice in this regard: "it constituted the first opportunity (for me at least)". Just as the EC represented a training ground for the TDE team, the project as a whole represented a training ground for the Lancaster team to implement a novel approach to LAL training. This sense of a trajectory of professional development is also evidenced in the statement about the importance of stakeholder engagement, which is framed as a discovery: "I have come to realize". The development of LAL through the praxis approach can therefore be characterised as a reciprocal process, with researcher/trainers contributing to the development of the skills and knowledge of the TDE team, and the researchers learning through working with the teachers in an authentic exam reform context.

Conflicts and Challenge

Despite the positive evaluation of the training and research-based approach which was employed throughout the project, one of the key points across narratives is the detailing of various challenges which have hindered or prevented the easy application of theory to practice (and implementation). Extract 4.6 below provides a detailed account of these challenges from the perspective of Charlie.

Extract 4.6: Charlie

> Unsurprisingly, however, the project itself has also presented several stern challenges over the years. If finding time for the completion of the module-related assignments in the early stages was still rather straightforward, combining the numerous demands and responsibilities of the TDE project with the substantial workload of a full teaching schedule has proved increasingly difficult. Unfortunately, due to various circumstances, the team has become smaller, and fewer hours of work relief have been granted to the group; thus, juggling the various tasks simultaneously has definitely not become any easier. In addition, constantly defending the necessity and legitimacy of the project, fighting for its acceptance and continued existence in negotiations with various stakeholders (such as ministry representatives and members of curricular boards) has, at times, been rather exhausting. Here again, valuable help has been provided by Researchers A and B, both in terms of their strategic advice as well as through various written documents they compiled that described, analysed, endorsed and underpinned the merits of the work done by the group. Similarly, the team repeatedly benefitted from insights these experts could offer about other, comparable projects they (or other researchers from their university) had taken part in or were familiar with. Still, overall, I must admit that it has sometimes felt as though too much time and effort has been spent on matters other than the actual research into – and work on – a new, better school-leaving exam. In that regard, it has been rather frustrating that even 6 years after the inception of the project, we can still not focus our entire attention on the design of a new exam, due to a variety of other obligations and constraints. While the knowledge gained through the taught modules has certainly been put to use in a variety of ways (e.g. "Épreuve Commune", teacher training courses, individual classroom tests designed by TDE members…), I must say that I had perhaps not quite anticipated how complex and time-consuming the more "political" aspects of the project would be, and that I would often have preferred to be able to purely focus on the original, research-based aims of the project more.

Charlie's narrative, at this point, transforms from a story of increasing expertise and mutual collaboration to one of frustration. In the passage, Charlie refers to "stern challenges", "numerous responsibilities", a "substantial workload", and "complex and time-consuming … 'political' aspects" which have prevented the development of the project in the manner that the team had originally envisaged. The heart of this frustration is encapsulated in Charlie's reference to the need to defend the "necessity and legitimacy of the project, fighting for its acceptance and continued existence in negotiations with various stakeholders". Charlie here returns to the metaphor of a battle, albeit with a connotation of weariness and fatigue in contrast to the journeying sense expressed at the beginning of the story. Charlie self-positions in the narrative as a 'juggler' of multiple competing demands, and positions the team as attempting to bring about change, but thwarted by forces beyond their control. The sense of frustration can be understood as rooted in the moral and ethical nature of praxis; a sense that the team is not able to act rightly. This frustration becomes more pronounced as the team develop their expertise and see that it cannot be easily applied. A similar story is told by Dominique in Extract 4.7, where "political decision makers" are characterised as antagonists in the TDE story. Once again frustration is borne, not of a lack of understanding, but of a tension between the desire to do what is right, and the more bureaucratic obstacles placed in the way.

Extract 4.7: Dominique

> My involvement in this project has made me realise that it can be very frustrating to have to compromise between what one knows to be best practice and very demanding time and personnel demands. For an exam reform to be sustainable and successful, one needs to invest considerable time and resources, so that the test cycle can run through its different stages and be re-evaluated and improved successfully. However, political decision makers usually want fast improvements and results, which can often lead to tensions, quick fixes and frustrated stakeholders.

The teacher-researcher collaboration is, to a large extent, peripheral to these problems. As Charlie notes, the Lancaster team provided assistance that aligned with their expertise as researchers: they drafted evidence-based reports, and were able to offer insights from similar projects (that is, the research team attempted to learn from existing accounts in countries like Austria and Hungary which had documented their own exam reform procedures – see e.g. http://www.lancaster.ac.uk/fass/projects/examreform/; Spöttl et al. 2016). But here we find the limits of what a purely research-driven collaboration might offer. In the final sentence of the passage, Charlie clearly differentiates between two dimensions of the project: the research-based LAL dimension, and the political dimension. The former is construed favourably as the preferred pathway; the latter as the *realpolitik* of exam reform (see also Pižorn and Nagy 2009).

A relevant passage from the researcher perspective – from Sam in this case – helps to underscore this sense of two different dimensions of expertise required to deal with the full spectrum of challenges presented by the exam reform project.

Extract 4.8: Sam

> I have learned how much I still need to know about working with groups of teachers. There is a lot of responsibility required for consultants in this position, and while Chris and I do our best I often have the feeling that I wish we had all the answers (we don't!). Finally, I have learned more about how little connection there is, at times, between the theoretical aspects of our work in language testing and the practical issues that those seeking to reform examination systems face. A lot of advice we provide is not research-based, it is more experiential. And as we stick with the project this experiential knowledge grows, and can be applied elsewhere …

Extract 4.8 demonstrates, once again, the symbiotic nature of the LAL trajectory. Sam was able to develop a broader understanding of how theory is applied in practice, and this led to a more comprehensive, experiential knowledge-base. However there is also a sense of frustration in the researcher narrative – "I wish we had all the answers" – which mirrors that of Charlie and Dominique. In a broader sense, both "sides" of the teacher-researcher story demonstrate the tension at the heart of attempts to shift towards a praxis of *informed action*. In order to be agents of change, teachers *and* researchers need to be politically-savvy, understanding the nature of communications, having the ability to frame issues to persuade key stakeholders, and being able to access those who hold power or who contribute to decision-making processes. However, there is no theory of political engagement in language assessment to draw on; rather, consultants need to develop experiential knowledge and develop a pragmatic approach through trial and error.

Reflections

Throughout this chapter we have demonstrated that narrative inquiry is able to draw out useful and interesting observations around praxis and the development of language assessment literacy at a thematic level and at a micro-level. Thematically, we observed that while there was a sense of shared enthusiasm at the beginning of the project, and a positive evaluation of the training part of the project from both teachers and researchers, there was also a shared sense of frustration with barriers – environmental, political, financial – that prevented a more transformative praxis within the exam reform project. At a more micro-level, we were also able to reveal positioning strategies within narratives which revealed interesting alignments and disparities between narrators, and evidence of trajectories in agency and expertise as the project developed. Participants' use of metaphor proved especially revealing as it helped to illuminate teachers' and researchers' stances towards each other, and towards the exam reform project in general. Here, the reform project was framed, at times, through a battle/fight metaphor – a common metaphor in other domains of communication (e.g., health contexts: Semsino et al. 2018). This metaphor suggests a fruitful line of enquiry for other research into educational reform projects: how do participants (teachers, researchers, and other agents) frame their endeavours? As collaborative or oppositional? And in the latter case, how might more negative framings impact on project work more tangibly? In the current project, awareness of these framings might, for example, lead to different strategies around engaging and communicating with key decision-makers which prioritise collaboration and shared goals wherever possible.

While it was possible to locate trajectories of language assessment literacy, it was, however, not easy to locate details of particular developments beyond broad notions of "concepts and procedures". We were able to locate instances of the iterative development of an LAL praxis, but not necessarily the specific nature of the research and theory underpinning these shifts. If data had been collected through interview methods, for example, there would have been a chance to follow-up with questions asking about specific elements of LAL which were improved during the course of the project, and which were not. On the other hand, collecting data through a more conventional oral interview runs the risk that narratives do not emerge naturally, with the result that we see much more of a disjointed picture of trajectories. In addition, linguistic features such as metaphor may be quite different, or absent, in an interview format. The written narrative, in particular, allows for a more structured approach to reflection which can be highly useful for such a long-term project because it allows more time to put events into perspective. In this way, an analysis of what is included and excluded in the narrative becomes more meaningful. Alternatively, in future, we might want to explore the use of narratives as objects in themselves for researcher-participant reflection and discussion.

Finally, this chapter itself presents an iterative link from practice back to theory. We have attempted to apply narrative inquiry in order to document and understand the nature of LAL development during the exam reform project, and to identify bar-

riers which might be theorised within a revised understanding of LAL needs within this context. It is clear, for example, that a research-based LAL concentrating on principles and practice is not enough in this context. What is further required is developing expertise in engaging with those political players who have emerged in the story: communicating results; telling stories; framing the project in a way that is persuasive. The project is ongoing, and these issues have not receded. Indeed, the project is currently (late-2017) facing some of the most difficult resistance within the broader context that it has faced thus far. On the one hand, the TDE team now represents a very strong and highly-skilled team of test developers, and we have trialled a scaffolded process of LAL and achieved successful results in terms of knowledge and skills development and awareness of ethical practice. On the other hand, however, the ultimate goal of the exam reform to have informed, sustainable impact on assessment practices in the wider educational sphere in Luxembourg – the key reason why the TDE team was set-up – is an ongoing site of struggle and compromise.

References

Baker, B. A., & Germain, J. (this volume). Narrative inquiry as praxis: Examining formative assessment practices in a nature-based indigenous language classroom. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of L2 classroom assessment: Praxis and researcher-teacher partnership*. Cham: Springer.

Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing, 35*(4), 557–581.

Bamberg, M. G. (1997). Positioning between structure and performance. *Journal of Narrative and Life History, 7*(1–4), 335–342.

Barkhuizen, G. (2008). A narrative approach to exploring context in language teaching. *English Language Teaching Journal, 62*(3), 231–239.

Barkhuizen, G. (2010). An extended positioning analysis of a pre-service teacher's better life small story. *Applied Linguistics, 31*(2), 282–300.

Barkhuizen, G. (2011). Narrative knowledging in TESOL. *TESOL Quarterly, 45*(3), 391–414.

Barkhuizen, G. (2016). A short story approach to analysing teacher (imagined) identities over time. *TESOL Quarterly, 50*(3), 655–683.

Barkhuizen, G., Benson, P., & Chik, A. (2014). *Narrative inquiry in language teaching and learning research*. New York: Routledge.

Barthes, R. (1977). *Image-music-text* (S. Heath, Trans.). Glasgow: William Collins.

Brunfaut, T., & Harding, L. (2015). *Examen de fin d'études secondaires, Anglais – Analysis of past exams* (Report, Test Design and Evaluation project). Lancaster: Lancaster University.

Brunfaut, T., & Harding, L. (2017). *Examen de fin d'études secondaires, Anglais: A baseline survey with teachers* (Report, Test Design and Evaluation project). Lancaster: Lancaster University.

Brunfaut, T., & Harding, L. (2018). Teachers setting the assessment (literacy) agenda: A case study of a teacher-led national test development project in Luxembourg. In D. Xerri & P. V. Briffa (Eds.), *Teacher involvement in high stakes language testing* (pp. 161–162). Cham: Springer.

Chase, S. E. (2008). Narrative inquiry: Multiple lenses, approaches, voices. In N. K. Denzon & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 57–94). Los Angeles: Sage. Publications.

Cumming, A. (2004). Broadening, deepening, and consolidating. *Language Assessment Quarterly, 1*(1), 5–18.

Duff, P. A., & Bell, J. S. (2002). Narrative research in TESOL: Narrative inquiry: More than just telling stories. *TESOL Quarterly, 36*(2), 207–213.

Freire, P. (1970). *Pedagogy of the oppressed*. New York: Herder and Herder.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113–132.

Hamp-Lyons, L. (2017). Language assessment literacy for language learning-oriented assessment. *Papers in Language Testing and Assessment, 6*(1), 88–111.

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413–427). Berlin/New York: Mouton De Gruyter.

Inbar-Lourie, O. (2012). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford: Wiley Blackwell.

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing, 30*(3), 301–307.

Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly, 40*(1), 235–257.

Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*. Published online: https://doi.org/10.1080/15434303.2019.1674855

Labov, W. (1997). Some further steps in narrative analysis. *Journal of Narrative and Life History, 7*(4), 395–415.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169–197.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education. Vygotskian praxis and the theory/practice divide*. New York: Routledge.

MEN (2011a). Dossier de presse. Réforme du lycée. Luxembourg: Ministère de l'Éducation nationale, et de la Formation professionnelle. Retrieved from https://www.gouvernement.lu/729597/dossier-presse.pdf

MEN (2011b). Proposition de texte d'une loi sur l'enseignement secondaire. Luxembourg: Ministère de l'Éducation nationale, et de la Formation professionnelle. Retrieved from https://www.gouvernement.lu/729578/dossier.pdf

Mulcahy, C. M., Mulcahy, D. E., & Mulcahy, D. G. (2015). *Pedagogy, praxis and purpose*. New York: Routledge.

Pavlenko, A. (2002). Narrative study: Whose story is it, anyway? *TESOL Quarterly, 36*(2), 213–218.

Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics, 28*(2), 163–188.

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381–402.

Pižorn, K., & Nagy, E. (2009). The politics of examination reform in Central Europe. In J. Charles Alderson (Ed.), *The politics of language education: Individuals and institutions* (pp. 185–202). Bristol: Multilingual Matters.

Poehner, M. E., & Lantolf, J. P. (2010). Vygotsky's teaching-assessment dialectic and L2 education: The case for dynamic assessment. *Mind, Culture, and Activity, 17*, 312–330.

Semino, E., Demjén, Z., & Demmen, J. (2018). An integrated approach to metaphor and framing in cognition, discourse, and practice, with an application to metaphors for cancer. *Applied Linguistics, 39*(5), 625–645.

Spöttl, C., Kremmel, B., Holzknecht, F., & Alderson, J. C. (2016). Evaluating the achievements and challenges in reforming a national language exam: The reform team's perspective. *Papers in Language Testing and Assessment, 5*(1), 1–22.

Villa Larenas, S. (2017, July). *Language assessment literacy of teacher educators in Chile*. Poster presented at the Language Testing Research Colloquium, Bogota (Colombia).

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402.

Vygotsky, L. S. (2004). The historical meaning of the crisis in psychology: A methodological investigation. In R. W. Rieber & D. K. Robinson (Eds.), *The essential Vygotsky* (pp. 227–344). New York: Kluwer/Plenum.

Webster, L., & Mertova, P. (2007). *Using narrative inquiry as a research method: an introduction to using critical event narrative analysis in research on learning and teaching*. Abingdon: Routledge.

Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment, 6*(1), 133–158.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly, 43*(3), 492–513.

Part II
Resisting Researcher-Teacher Hierarchies in Favor of Cooperative Partnerships

Chapter 5
Learning-Oriented Assessment: More Than the Chalkface

Constant Leung

Abstract The idea of using assessment to promote learning has now gained powerful traction in education. However, a good deal of the relevant theoretical and research work tends to focus on the teaching-learning interface, paying particular attention to efficacious teacher and peer interaction that can enhance student learning. Relatively little attention has been paid to the importance of curricular provision and institutional facilities (cf. Norris 2016). In this chapter I will draw on the work of a case study exploring (a) the ways in which experienced university teachers try to develop and implement an assessment approach that they believe can promote student learning, and (b) students' perceptions of and responses to assessment in relation to their own ideas of learning. Taking an interpretive phenomenological approach, I track the experiences and expressed views of the teachers and students of a Masters programme in English Language Teaching over a 24-month period. The overall aim for this chapter is to provide an 'insider' account of the 'hinterland' of assessment within an academic context, and to call for a broadening of perspective to take account of both conceptual and curricular issues situated within institutional contexts.

Introduction

In their introductory chapter to this volume, Poehner and Inbar-Lourie explicate the value of partnerships between researchers and practitioners as a way to develop both L2 education and the theories and conceptual models that may inform them. The focus of the present chapter is on the diverse ways in which learning and teaching are positioned conceptually in theorizing and researching learning-oriented assessment (LOA), and the possible 'after-wash' or consequences of the different positions for curriculum design and pedagogic practice. That is, in line with the

C. Leung (✉)
Centre for Language, Discourse & Communication, School of Education, Communication & Society, King's College London, London, UK
e-mail: constant.leung@kcl.ac.uk

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_5

thematic focus of this book on praxis, understood as the dynamic interrelation of theory/research and practice, the present chapter recognizes that differing conceptualizations of what occurs in classrooms must be accounted for as part of the theoretical and practical elaboration of LOA. Moreover, the arguments and ideas explored in this chapter emerged not from a study designed and conducted by researchers external to a context of practice and for whom the teachers and learners involved were simply participants. Rather, the ideas emerged from a piece of practitioner research on student responses to tutor feedback comments on their written assignments. By 'practitioner research' is meant the observations and interview data included in this discussion were collected during the course of teaching and programme running with a particular focus in mind (Menter et al. 2011). The emerging data are then explored collaboratively with colleagues concerned with a view to achieving a shared and, hopefully, deeper understanding of the on-going practice. No contrived data were involved. For reasons that will become clear presently, 'feedback' is effectively the clutch mechanism, metaphorically speaking, that links curriculum, teaching, learning and assessment in this discussion.

Perhaps a brief word on provenance and nomenclature is appropriate here. LOA, a relatively recent coinage, is used as an inclusive umbrella term in this discussion to cover a range of the research and theoretical work in classroom-based assessment that has appeared under different labels such as *assessment-for-learning, dynamic assessment, embedded assessment, formative assessment* and so on (for examples of this work, see Chaps. 6, 9 and 10 in this volume by Baker and Germain, Poehner and van Compernolle, and Davin and Herazo). The discursive use of LOA as a superordinate term is not intended to supplant all the different schools of thought, each with its rich conceptual and theoretical articulations and principles of practice (see Leung et al. 2018). LOA is used here totemically to foreground the 'for learning' dimension of assessment; indeed, at the heart of LOA is obviously learning, and by association, teaching. The main aim of this chapter is to demonstrate that LOA is context- and people-sensitive, and that this renders the concept particularly useful for orienting to situated local research involving teachers and students. Unlike certain chapters in this volume, the present discussion does not re-examine LOA in light of a particular researcher-teacher partnership but instead it emphasizes the importance of engaging with the individuals in a given educational context whose lived experiences, goals, needs, and values shape that context and influence the ways in which practices are realized and interpreted. Such engagement is an essential starting point for praxis as well as a concern that must continually be returned to throughout collaborative undertakings, a point illustrated by Harding and Brunfaut (Chap. 4, this volume).

In the first part of this chapter I will draw attention to two foci on LOA found in the research and professional literature: the teaching-learning interface and the relationship between assessment theories (often implicitly and routinely embedded in practice) and institutional and external affordances and constraints. After that I will explore the complex links between classroom level LOA practice, curricular infrastructure and teachers' intellectual/professional dispositions. This part of the

discussion will be facilitated by some data drawn from a study of student responses to teacher feedback. In the concluding remarks I will suggest that it is not enough for LOA to be based on a set of sound concepts and theories, rather the practice of LOA requires on-going situated local practitioner-led research.

Focus on Teaching-Learning Interface

Conceptually LOA can be said to be at the intersection between curriculum, teaching, learning and assessment. A good deal of theorizing of and research in LOA is, perhaps not unreasonably, focussed on the teaching-learning interface. The following is a selection of statements on representing a range of concerns and orientations at different levels of education. Black and Wiliam (1998: 2), for instance, open their discussion in the highly influential school education focussed pamphlet 'Inside the Black Box' with this statement:

> Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their work to meet their [pupils'] needs – needs which are often unpredictable and which vary from one pupil to another. Teachers can find out what they need in a variety of ways – from observation and discussion in the classroom, and from written work of pupils whether done as homework or in class.

Wiliam (2011:46) puts forward the following advice for teachers for embedded assessment, that is, classroom-based assessment of student performance by teachers carried out as part of their everyday teaching activities:

1. Clarifying, sharing, and understanding learning intentions and criteria for success
2. Engineering effective classroom discussions, activities, and learning tasks that elicit evidence of learning
3. Providing feedback that moves learning forward
4. Activating learners as instructional resources for one another
5. Activating learners as the owners of their own learning.

In a discussion on the kind of assessment that would likely promote student learning in university settings, Gibbs (2006: 29–30) sets out the following 11 conditions:

Assessed task capture sufficient study time and effort.
These tasks distribute student effort evenly across topics and weeks.
These tasks engage students in productive learning activity.
Assessment communicates clear and high expectations to students
Sufficient feedback is provided, both often enough and in enough detail.
The feedback is provided quickly enough to be useful to students.
Feedback focuses on learning rather than on marks or students themselves.
Feedback is linked to the purpose of the assignment and to criteria.
Feedback is understandable to students, given their sophistication.

Feedback is received by students and attended to.
Feedback is acted upon by students to improve their work or their learning.

In a discussion specifically focussed on science and technology in primary school education (Year 1 to Year 8) in New Zealand, Cowie, Moreland and Otrel-Cass (2013: 9) define LOA (the authors use the term 'assessment for learning') as follows:

> Assessment for learning encompasses those everyday classroom practices through which teachers, peers, and learners seek/notice, recognise and respond to student learning, throughout the learning, in ways that aim to enhance student learning and student learning capacity and autonomy.

In an elaboration on their approach to assessment for learning, Cowie et al. (op. cit.:11) state that their version of sociocultural orientation '… directs attention to the interaction between teachers, students and tasks in a particular setting as these evolve over time.'

Despite the clearly different concerns, contexts and orientations articulated in the above statements, a common perspective can be seen running through them. That is, their focus of attention is on the teaching-learning interface. The assumption is that learning is largely, if not exclusively, a consequence of teaching.

The Broader Context of Teaching, Learning and Assessing

While the teaching-learning interface is undoubted as a key issue for any discussion on teacher-led LOA, it should not be regarded as the only concern. Teaching-learning in schools and universities does not take place in a vacuum. The assessment conduct of teachers and students is influenced contingently by a whole host of other factors such as statutory assessment requirements (e.g. type/volume of evidence/performance) and standards/criteria, phases of education, institutional assessment arrangements, teacher values (individual and/or collective), and disciplinary beliefs and practices.

Institutional ethos and disciplinary practices can be influential. In a series of three studies of teacher assessment practices in England involving over 40 primary and secondary schools, Clarke and Gipps (2000: 45) found, for instance, that teachers working in the participant primary schools and English departments in secondary schools tended to use more 'informal, formative methods (e.g. pupil self-assessment, regular notetaking , use of pupil portfolios)' whereas teachers in Mathematics and Science departments in secondary schools tended to adopt 'rather formal approaches to ongoing assessment (e.g. end of module tests, regular classroom tests)'.

Crossouard and Pryor (2012) look closely at the practice of LOA (formative assessment in their terms) in a Scottish primary school context. Perhaps it should be noted that the school education system in Scotland prides itself on resisting using standardized testing for accountability purposes and favouring formative assessment to promote learning. This approach is built into the Scottish curriculum framework

(see http://www.gov.scot/Topics/Education/Schools/curriculum/assessment). In this case study the two researchers examine, *inter alia*, the ways in which teachers conceptualized formative assessment within the *collaborative challenges* – these were extended collaborative problem-solving group tasks, a feature of the curriculum work in this school. The topics of these challenges were quite complex, e.g. students engaging in democratic politics (forming a political party and developing a political manifesto). They were meant to afford students with opportunities to generate their own idea and to produce work that is not necessarily right or wrong. However, as noted by Crossouard and Pryor (op.cit.: 256), there seems to be 'a disjuncture between teachers' declared espousal of the "freedoms" of a 'challenge', and the framing of these freedoms within expectations that pupils were to learn particular curricular 'content'. This disjuncture is reflected in the following statement made by one the participant teachers:

'… there's a kind of freedom element which I think most children thrive upon. And when they're in a group, the generating of ideas within the group, some of the things they can come up with – it's quite amazing. So sometimes making it quite open ended the product, or how it's to be presented, it's great, because what they'll come up with to do is maybe nothing that I'd even considered. The road I was going down was completely different, [and] as long as they've tackled the challenge and learned what they're supposed to learn, that doesn't matter'. (Loc.cit.)The phrase 'as long as they've tackled the challenge and learned what they're supposed to learn' is telling. The 'openness' seemed to relate to the ways in which the students developed their challenge and presented their work. In terms of content knowledge, the students were reminded in the formative feedback that they should conform to certain expectations such as 'Put in accurate information' (loc.cit.). So 'openness' was a constrained concept; there is a curriculum structure that defines the meaning of freedom. Convergence with the school's agenda is expected.

All of this is consistent with Sadler's (1989:121) observations that for students to benefit from LOA, three conditions should be met:

'the learner has to (a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap.'It would seem learning, particularly learning in institutional settings with a curricular framework, involves external reference points. These external reference points can be formally stated standards and benchmarks and/or tacit expectations shared among key stakeholders, such as teachers, trainers and supervisors, who are charged with the responsibility to help others to achieve their best. Furthermore, these studies, like many others, locate assessment within an intricate web of inter-connected beliefs and values, activities and regulations. Together they unambiguously suggest that LOA, much like any other organized educational activity, must be understood as part of a curricular infra-structure with inter-linking components. In the next section I will begin to explore the idea that LOA is inescapably intertwined with a curricular infra-structure and professional-cum-intellectual beliefs and values. The discussion will be contextualized by a sample of student responses to tutor feedback on written assignments. Beyond the substantive teaching-learning

issues raised in relation to the individual students concerned, the data extracts in the next section also alert us to the need for a more fundamental conceptual sensibility. LOA approaches and principles, as we have seen earlier, can be transformed into a set of recommended practices from the standpoints of the teacher and/or the curriculum requirements. However, unless we take account of the dynamics, contingencies, and demands of local contexts through the students' perspectives, we might simply be enforcing a top-down learning agenda in the guise of LOA. Seen in this light, understanding students' perceptions and orientations through dialogue is a critical point of engagement for LOA practice.

Individual Responses to Feedback

The data reported here are drawn from a corpus of student responses to formative feedback conducted in a London university in 2013/4 (for a fuller account see Andon et al. 2018). The context and setting of the study were as follows: The programme was a Masters in Applied Linguistics and TESOL; the teaching team involved was concerned that students should make the best use of their highly pressurised time on a one-year programme; and promoting effective learning through feedback comments on written assignments was one of the formative strategies adopted by the teaching team. As part of this pedagogic approach, the first written assignment in the first (autumn) term was designed to be an early opportunity to provide formative comments to help students understand the teaching team's expectations in terms of content selection, textual organisation and academic register. The programme leader was at pains to explain and emphasize this aim to the students in all the meetings with them throughout the academic year. Students were encouraged to see their tutors to discuss the feedback comments that they had received and to share their feedback comments with one another.

The study involved eight students on this programme: four UK-based students, four international students – two of whom were English L1 speakers. The student participants were all volunteers. All the students on this MA TESOL programme were briefed on the main purpose of the study – to explore the students' perceptions of written feedback as a form of formative support for learning. The study took place in the summer term 2015. All the students had received marks and comments on the assignments they submitted in the autumn term (the previous September to December). The participant students were invited to attend a 30-min semi-structured interview to discuss the feedback comments, and their responses to them. The students were advised that they were free to choose any marked assignment/s that they had written as a basis for discussion.

A qualitative content analysis was conducted on the interview data. To draw out the main themes in the data, the analysis followed inductive coding procedures to identify manifest and latent content (Berg 2009, chapter 11). Manifest content refers to explicitly expressed points of view or statements by the student participants (in response to questions in the interview); latent content refers to statements that

require some degree of interpretation. The contingent and co-constructed nature of the interview utterances was taken into consideration in the analysis (Mann 2011, 2016). The data analysis and findings were conducted and cross-scrutinized by the three members of the research team, and there was consensus on the emerging themes.

A number of themes emerged from the study. For reasons of space and scope, I will report some of these themes as they relate to feedback. It is important to make clear that the participant students expressed a large number of views and opinions on a wide range of issues in the individual interviews, but there was a tendency for one or two of predominant themes to emerge in most, if not all, the interviews. For clarity of representation and sense of immediacy the themes will be illustrated through four mini-accounts of individual student responses. The students and tutors presented here are either given pseudonyms or anonymised.

Sadie – Happy Let-It-Pass

Sadie was a UK-based practising teacher. On returning to university to study, she found academic writing at university to be quite challenging. She did not do very well in her first Grammar and Phonology assignments in the autumn term, partly because she had to do them hurriedly due to time pressure.

The Grammar assignment involved evaluating the strengths and weaknesses in the sample extracts of vocabulary exercises in two textbooks, each representing a different approach. Figure 5.1 below was the opening section of her assignment:

This opening statement was commented upon by the tutor in his/her overall feedback as follows (Fig. 5.2).

The contrast between the orientation adopted by Sadie and the tutor's expectations for the work she would produce is clearly discernible in this opening section, and this in fact shaped much of the feedback Sadie received for this assignment as well as her overall marks. Figure 5.3 below shows another section from this assignment immediately after the opening. The numbers refer to the specific remarks made by the tutor:

The tutor's numbered comments were:

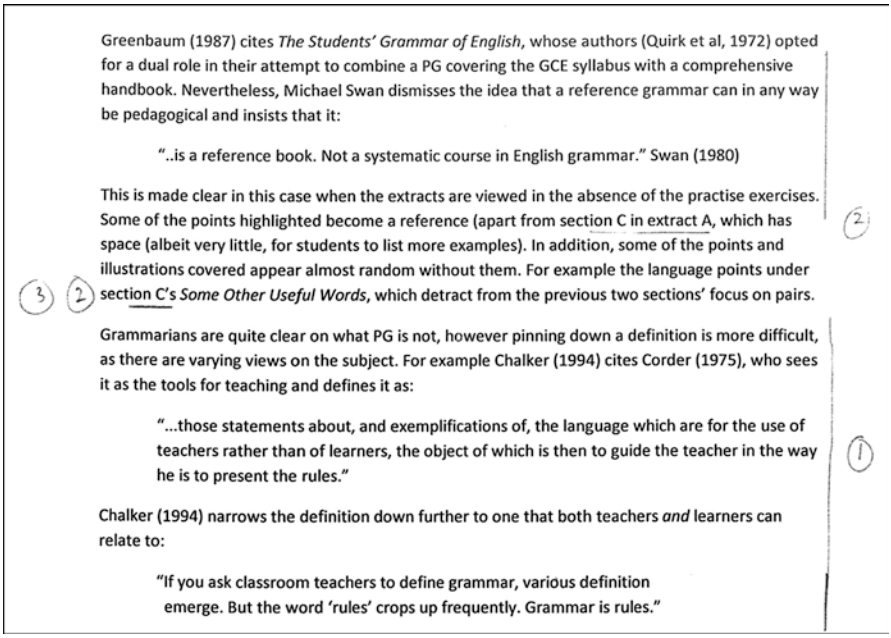
1. Most of this is irrelevant to the question asked … Just stick to the question asked.
2. Cross-refer more accurately. Section C of 26, 27 or 28?
3. 27C's.

> The term pedagogic grammar (PG) is used to describe the way in which the grammar of a language is used to communicate with other people and is aimed at people who specifically want to learn the target language. This is distinct from a reference grammar, for example, which teaches people about the language in question …

Fig. 5.1 Sadie's opening section- Grammar assignment

The beginning of this analysis suffers from a failure to stick to the question asked. You were not asked to discuss the difference between pedagogic grammar and reference grammar, or the nature of pedagogic grammar …

Fig. 5.2 Tutor's written comment on Sadie's opening statement (on separate comment sheet)



Greenbaum (1987) cites *The Students' Grammar of English*, whose authors (Quirk et al, 1972) opted for a dual role in their attempt to combine a PG covering the GCE syllabus with a comprehensive handbook. Nevertheless, Michael Swan dismisses the idea that a reference grammar can in any way be pedagogical and insists that it:

"..is a reference book. Not a systematic course in English grammar." Swan (1980)

This is made clear in this case when the extracts are viewed in the absence of the practise exercises. Some of the points highlighted become a reference (apart from section C in extract A, which has space (albeit very little, for students to list more examples). In addition, some of the points and illustrations covered appear almost random without them. For example the language points under section C's *Some Other Useful Words*, which detract from the previous two sections' focus on pairs.

Grammarians are quite clear on what PG is not, however pinning down a definition is more difficult, as there are varying views on the subject. For example Chalker (1994) cites Corder (1975), who sees it as the tools for teaching and defines it as:

"...those statements about, and exemplifications of, the language which are for the use of teachers rather than of learners, the object of which is then to guide the teacher in the way he is to present the rules."

Chalker (1994) narrows the definition down further to one that both teachers *and* learners can relate to:

"If you ask classroom teachers to define grammar, various definition emerge. But the word 'rules' crops up frequently. Grammar is rules."

Fig. 5.3 Sadie's grammar assignment page 1

This assignment was given 52 marks, a low Pass. Sadie expressed her disappointment with the tutor's comments because she thought they were not helpful and 'punitive' (her word).

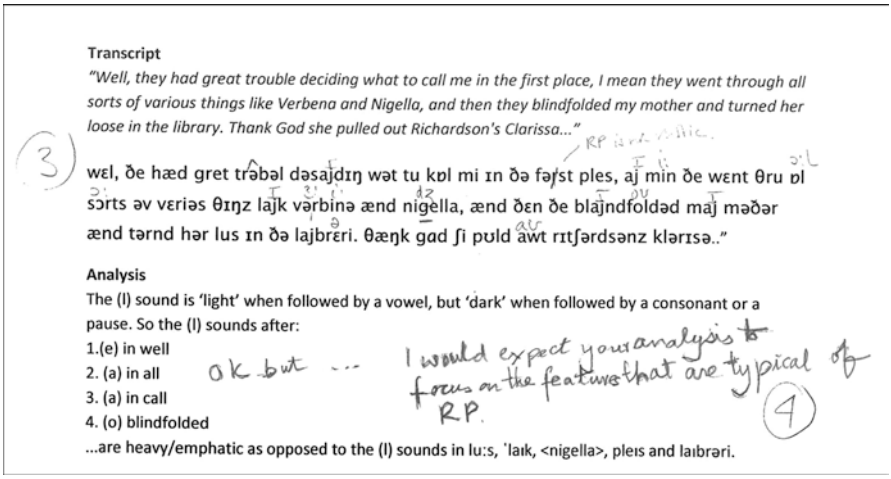
Figure 5.4 below shows a segment of Sadie's Phonology assignment with (hand-written) feedback remarks from the tutor:

The tutor's feedback corrections and comments, as shown above, suggested that her assignment did not pay attention to the expected issues, and that she should have focussed on Received Pronunciation (RP in the hand-written comment 1). Comments 3 and 4 advised Sadie as follows (on a separate feedback sheet) (Fig. 5.5).

This assignment was given a Fail (below 50). However, Sadie did not seem to be unhappy with the Fail mark at all. In fact, in her interview she made the following remarks in relation to the tutor's comments ('she' in lines 1 and 5 refers the tutor)[1]:*

[1] Transcription key:

RE – researcher
= – latching



Transcript

"Well, they had great trouble deciding what to call me in the first place, I mean they went through all sorts of various things like Verbena and Nigella, and then they blindfolded my mother and turned her loose in the library. Thank God she pulled out Richardson's Clarissa..."

wɛl, ðe hæd gret trəbəl dəsajdɪŋ wət tu kɒl mi ɪn ðə fɜ́ːst ples, aj min ðe wɛnt θru ɒl sɔ́ːrts əv vɛriəs θɪŋz lajk vərbinə ænd nigélla, ænd ðɛn ðe blajndfoldəd maj mədər ænd tərnd hər lus ɪn ðə lajbrɛ́ri. θæŋk gəd ʃi puld áwt rɪtʃərdsənz klərɪsə.."

Analysis

The (l) sound is 'light' when followed by a vowel, but 'dark' when followed by a consonant or a pause. So the (l) sounds after:

1.(e) in well
2. (a) in all
3. (a) in call
4. (o) blindfolded

...are heavy/emphatic as opposed to the (l) sounds in luːs, ˈlaɪk, <nigella>, pleɪs and laɪbrəri.

OK but ...

I would expect your analysis to focus on the features that are typical of RP.

Fig. 5.4 Extract from Sadie's phonology assignment

3. You have made a commendable attempt to do a very thorough transcription of the RP accent but there are several mistakes in your choice of symbols (see pencil notes).
4. For the purposes of this assignment, I would have expected your analysis to focus on the features that are typical of RP – maybe using as a check list descriptions of RP that can be found in Collins and Mees 2008 or the article on Accent Variation in your Reading Pack).

Fig. 5.5 Tutor's specific comments

1	Sadie	so she's saying you've done XYZ but maybe you should have done this (.) and I
2		don't take that too negatively (.) I take that as a suggestion (.) for me to do
3		something=
4	Re	=would you be able to make that amendment or addition
5	Sadie	yes if I had gone back to to the IPA (.) if I had gone back to that where she says
6		features that are of typical RP then I'd go back to the text and say OK I do know
7		what the features are but I omitted that because I thought this is probably would
8		have been probably be better now that I know I can go back to that pick that and

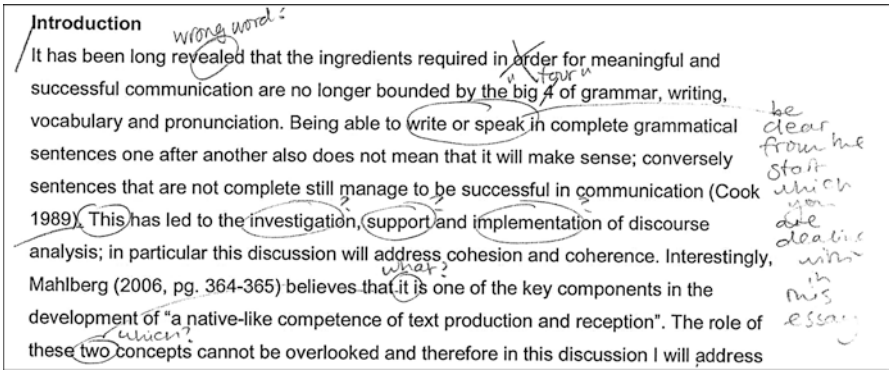
(.) – brief pause
… – text reduction
( ) – contextual information

9		add that to it …
10		(a few moments later)
11	Re	I think you said at one point a couple of minute ago that you should have gone
12		back to see [the teacher] … (.) so why didn't you
13	Sadie	I didn't because (.) uh I really enjoyed her lesson I thought she was one of the you
14		know uh she was a really good teacher (.) I thought I'd let her down somehow …

It seems quite clear that Sadie understood that her assignments did not orient the content in the ways that the tutors expected and she did not challenge their judgements. However, her responses to the two sets of comments were quite different. With the Grammar assignment she found the comments unhelpful and 'punitive'; with the Phonology assignment she appreciated the comments and the mark (a failure). She seemed to have responded to the tutors' comments in accordance with her appraisal of the quality of their teaching. While this was the disposition of one individual, it points to the need to understand the basis of students' responses to comments intended for formative purposes. In addition, the issue of affect was also involved.

Amy – Rejectionist

The topic of the assignment that Amy brought along to the interview for discussion was on Linguistic Analysis. Amy was from North America, and like the other students in the programme was a teacher. She was given 58 marks (a Pass mark). An extract of the assignment is shown in Fig. 5.6:



Introduction

It has been long revealed that the ingredients required in order for meaningful and successful communication are no longer bounded by the big 4 of grammar, writing, vocabulary and pronunciation. Being able to write or speak in complete grammatical sentences one after another also does not mean that it will make sense; conversely sentences that are not complete still manage to be successful in communication (Cook 1989). This has led to the investigation, support and implementation of discourse analysis; in particular this discussion will address cohesion and coherence. Interestingly, Mahlberg (2006, pg. 364-365) believes that it is one of the key components in the development of "a native-like competence of text production and reception". The role of these two concepts cannot be overlooked and therefore in this discussion I will address

Fig. 5.6 Extract from Amy's linguistic analysis essay

In the interview, Amy stressed the fact that she went to school and university in North America where the range of assessment marks awarded to discursive assignments is generally higher than that in the UK. From my own professional experience, I know that, for instance, in the fields of Applied Linguistics and TESOL a mark of 85 (out of 100) would be regarded as exceptionally high in the UK, whereas in North America this would not be regarded so. Throughout the interview Amy expressed her general disenchantment with her assignment grades. The Linguistic Analysis assignment under discussion was given 58 (out of 100) – a middle-of-the range Pass mark in UK terms; Amy's response to this mark was 'I have never been given a mark in the 50s before'. She thought the feedback comments were 'pretty harsh'. Furthermore, she did not see the relevance of the feedback for her next assignments; in fact she seemed to regard the tutor's comments as 'corrections'. The interview extract below captures Amy's view on the value and usefulness of the written comments on her assignment:

1	Amy	no I didn't actually really pay attention to it (.) to a lot of the feedback (.) I just read
2		the (.) I just saw the grade (.) and I saw the first page of the feedback comment=
3	RE	=right
4	Amy	I was not happy with it I didn't need to read the rest of it
5	RE	aha (.) right (.) so hmm (.) that's interesting (.) because the idea of (.) of the you know
6		feedback comment
7	Amy	I did look through it (.) but I didn't (.) study each and everything uh that was written
8		In the paper
9	Re	right (.) in fact (.) you had relatively(.) little idea of what they were saying to you
10	Amy	Yeah
11		(a few moments later)
12	RE	why do you think then that that (.) you know the tutors bothered to write all these
13		comments from your point of view (.) was it because they were trying to [show
14	Amy	[they were
15		trying to correct my work
16	RE	right (.) and (.) and you didn't get a sense that you could sort of pick up on some of
17		the comments and maybe (.) use them to develop your ideas (.) no you didn't
18	Amy	because this this linguistic analysis uh it wasn't a paper (.) it was eh (.) they told us
19		that it wasn't in a uh essay format (.) it was a different kind of assignment (.) it was

20 just an assignment (.) so it had nothing to do (.) with the following assignment (.) they
21 were in different styles (.) two different formats (.) so

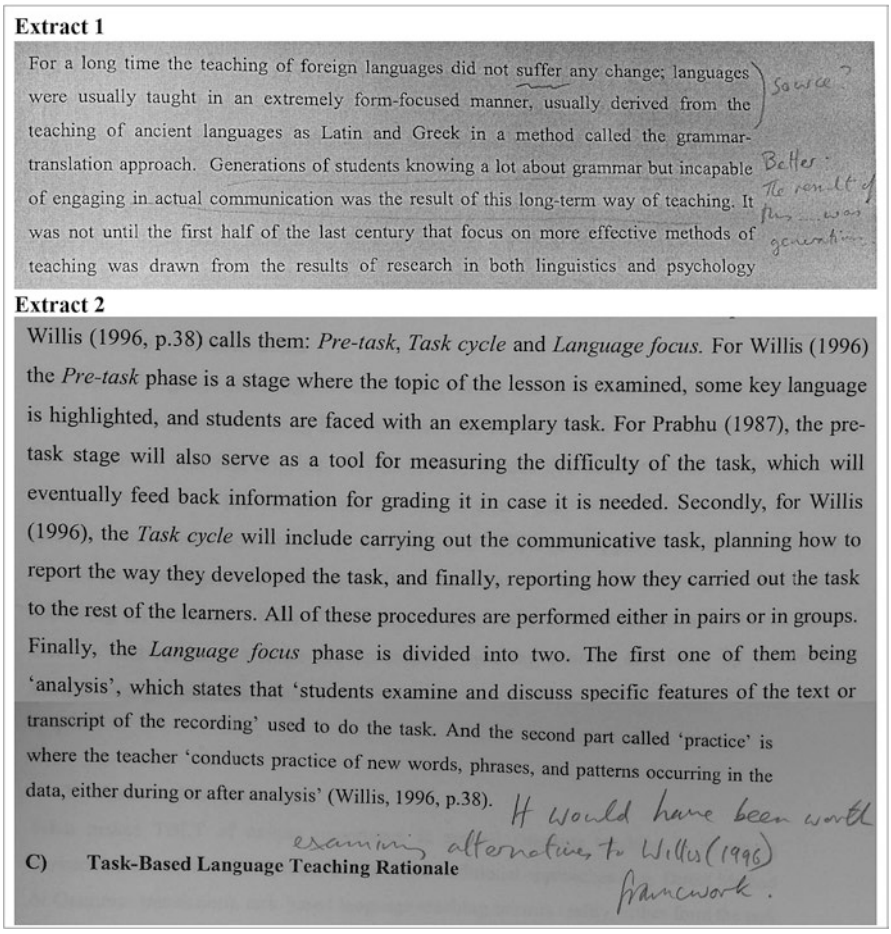
At the time of the interview, 5 months or so after the assignments were marked and returned, Amy was still expressing her unhappiness and dissatisfaction with both the tutor support for assignment writing and the marks that she received. She expressed the opinion that the tutors did not explicitly and adequately tell her what to do. While she was aware that there was a difference in the mark ranges typically used by North American and British university teachers, she was nevertheless unwilling to accept the marks that she was awarded. Furthermore, Amy did not seem to appreciate the intended formative purpose of the tutor comments, some of which could be helpful in the writing of other assignments. She seemed to regard each assignment as a stand-alone task for which a (numerical) higher mark should be awarded. This is an issue of both affect and intellectual orientation.

Miguel – Selective and Critical Acceptance

Miguel, an experienced teacher from South America, was highly motivated and doing well generally. He had expressly signalled a strong desire to improve his work and was keen to receive advice to do this. He was given 60 (at the low end of the good Pass range) for the assignment on Pedagogic Principles (see the two extracts in Fig. 5.7).

Miguel seemed to value positive comments and advice on subject content (extract 2), but not language and writing style issues. However, in the interview extract below Miguel expressed his critical response to his tutor's suggestion for an alternative phrasing (see extract 1 above: 'Better: The result of this was generations …').

1 Miguel (reading out teacher's comment) maybe it's just a linguistic thing but I thought
2 that was the way to emphasize my idea and
3 RE ah right so (.) this point was (.) not so much about (.) the content meaning as (.)
4 about the [style of
5 Miguel [yeah yeah
6 RE saying something
7 Miguel yes yes
8 RE ok [so
9 Miguel [that is very personal you [see
10 RE [aah



Extract 1

For a long time the teaching of foreign languages did not suffer any change; languages were usually taught in an extremely form-focused manner, usually derived from the teaching of ancient languages as Latin and Greek in a method called the grammar-translation approach. Generations of students knowing a lot about grammar but incapable of engaging in actual communication was the result of this long-term way of teaching. It was not until the first half of the last century that focus on more effective methods of teaching was drawn from the results of research in both linguistics and psychology

Extract 2

Willis (1996, p.38) calls them: *Pre-task*, *Task cycle* and *Language focus.* For Willis (1996) the *Pre-task* phase is a stage where the topic of the lesson is examined, some key language is highlighted, and students are faced with an exemplary task. For Prabhu (1987), the pre-task stage will also serve as a tool for measuring the difficulty of the task, which will eventually feed back information for grading it in case it is needed. Secondly, for Willis (1996), the *Task cycle* will include carrying out the communicative task, planning how to report the way they developed the task, and finally, reporting how they carried out the task to the rest of the learners. All of these procedures are performed either in pairs or in groups. Finally, the *Language focus* phase is divided into two. The first one of them being 'analysis', which states that 'students examine and discuss specific features of the text or transcript of the recording' used to do the task. And the second part called 'practice' is where the teacher 'conducts practice of new words, phrases, and patterns occurring in the data, either during or after analysis' (Willis, 1996, p.38).

C) Task-Based Language Teaching Rationale

Fig. 5.7 Extracts from Miguel's corrected assignment

11	Miguel	that's the way I expressed I would like to keep that (.) not to be not to be given
12		feedback on my English because I think that's not grammatically incorrect (.) but it's
13		my style and that (.) might be a bit face-threatening if (.) if they say say it this way it's
14		correcting someone's pronunciation … (talking about his English language learning
15		experience) I came here to do the masters (.) not to learn English …
16		(a few moments later)
17	Miguel	uh (.) what I'm a bit uh (.) not worried but uh I don't see anything positive here

18		other than the checks (pointing to the ticks written by the marker)
19	RE	Uh
20	Miguel	it's just like when there are comments it's alright that that they comment on my
21		weaknesses (.) I appreciate that yeah but I cannot work only on my
22		weaknesses I would like to maybe give more strength to my strong (.) uh points
23		uh that I may use in my future work …

As an experienced teacher, Miguel seemed to have a well-defined sense of what kind of advice and guidance he expected from the tutors. While he welcomed feedback on content matters, he was not so keen on comments on his writing style. The eliding of grammatical correctness and writing style seems to suggest an underlying issue of his status or identity as a professional user of English. This is quite a complex question as the claim to professional legitimacy of being an English Language teacher is often linked to the individual teacher's own English proficiency. This particular instance of response to tutor feedback signals that formative feedback is not just about what is to be learned, it is also about what students would like to learn.

Louisa – Secretive Marks

Louisa was a UK-based teacher. She was an energetic and highly motivated member of the MA class. She produced very good quality assignments; her Linguistic analysis assignment, which she brought along to the interview for discussion, was given 72 marks (in the Distinction range). In the interview Louisa provided an account of the ways in which she made use of the published marking criteria in her assignment preparations. She took tutors' feedback comments in her work seriously and understood the need to pay further attention to her use of language to enhance the clarity and overall quality of her writing. She was generally pleased with the efforts made by the teaching team to publish marking criteria at the beginning of the course and to provide formative feedback. When asked if the students actually shared and discussed with one another the feedback comments and marks, she gave the following response:

1	Louisa	… when we got our assignment back (.) there was no discussion at all we
2		were sort of quite (.) eh (.) we were sort of a bit cagey we didn't want to
3		talk about our marks and things like that (.) so it was very (.) you know
4		it was quite (.) a private secretive sort of thing when we got our first
5		assignments back.

Louisa's comments suggest that, at least amongst her peers in the MA programme, students saw their academic efforts and achievement as a highly individualized matter. While individual students are clearly entitled to choose what and how much information they would be prepared to share with others, the sentiments underlying the 'private and secretive sort of thing' can militate against the fostering of a community of mutually supporting learners. Furthermore, it raises the issue of whether LOA can be optimally developed where learning is construed by students as a private matter between themselves and the teacher or the curriculum content, or whether LOA requires a more open dialogic environment.

Connecting Post-Feedback Reactions to Programme Infrastructure

The sample of student responses to tutor feedback we have just seen are clearly highly idiosyncratic, reflecting a multitude of factors such as individual students' background experiences in education and in general; their current capacity for and disposition on the learning at hand; and their medium-to-long term investment in success. Therefore, any follow-up action by teachers and tutors aimed at providing further formative guidance for individual students has to be bespoke, tailored to their specific needs. However, for teachers to be able to provide such individually oriented follow-up support there has to be a proactive capacity and policy to deal with post-feedback student responses at the programme/course level. It would be fair to say that at the present time teaching programmes and courses, at all levels of education, tend not to have any provision for post-feedback follow-up by teachers. The additional costs may be an issue, but, I would argue, it is also a fundamental matter of a limited pedagogic vision of LOA hitherto. A good deal of the discussions on LOA, as seen in the opening section, tend to frame LOA in terms of a one-way and one-stage process of teacher offering formative guidance to students. However, for the pedagogic value of LOA to be more fully realized, it would be necessary to develop a open-ended iterative process that takes account of student post-feedback responses that, in turn, provides the basis for further teacher follow-up. And student responses to teacher comments are a point of connection to the infrastructure of a programme. The infrastructure of any teaching programme is a complex web of cultural, intellectual, financial, organisational, policy, physical (e.g. access to teaching rooms and materials) and social affordances and constraints. For the purpose of this discussion I will focus on two aspects of the MA programme infrastructure: teaching staff ethos and curricular provision. Teaching staff ethos refers to commitment to pedagogic innovation, curiosity as to students' ways and levels of learning, and willingness to collaborate with students in teaching-learning matters. Curricular provision includes syllabus specifications for different subjects and time-tabled allocations for different teaching and learning activities.

The responses from Sadie, Amy, Miguel and Louisa can be interpreted in a variety of ways. The focus of attention here is on their possible connections with the infrastructure of the MA programme involved. Sadie was unhappy with the feedback comments on her Grammar assignment. She was aware that she was under time pressure and did not spend sufficient time on the assignment. Nevertheless, she thought that the comments were 'harsh' and 'punitive'. This raises the issue of how far the tutor's sharp formative focus on the content of the assignment as expressed through the comments (Figs. 5.1, 5.2, and 5.3) was appreciated by the student. It seems clear that Sadie responded more favourably to personally tuned feedback comments (Fig. 5.5) on her Phonology assignment. The issue of managing student affect is potentially an important consideration in LOA. To be able to deal with student affect, it would require the teaching team to be interested in student reactions to feedback comments and find ways to collaborate with students to gather this information. This is connected to teaching staff ethos. Furthermore, there is a potential implication for time-tabling as talking with students in 'town hall' meetings and individual tutorials requires time provision.

Amy seemed to be upset by the mark she received and did not think the comments were helpful for her future assignments. In fact, she thought the comments were 'pretty harsh' and that they were 'corrections'. It was quite clear that Amy interpreted the meaning of her mark with reference to her North American experience. It might not have been so upsetting if she had transposed the mark to the corresponding North American range. That said, this particular student response points to at least two other issues. Firstly there are problems with giving marks and formative comments at the same time. Most teachers know that if a mark is awarded to an assignment, students' interest is drawn to the mark and they tend to pay little attention to discursive comments (Black et al. 2003). Secondly, in an increasingly internationalized higher education environment, students tend to come from a variety of backgrounds. In all likelihood there were other international students in the programme whose interpretations of the meaning of their marks reflected their background experiences and did not necessarily match the meaning intended by the tutor. All in all, to address the issues raised by Amy's experience, it would be necessary for the teaching team to gather the necessary information on student responses and to find ways of working with students, both individually and collectively, to share the formative purpose of feedback comments and the meaning of the marks that they have received.

Miguel was not pleased with the feedback suggestions for alternative formulations and expressions of his ideas. This signals that students, particularly experienced teachers returning to university to do an advanced degree, often have a strong sense of what they would like to learn and achieve. Some formative comments, however, well-intentioned, may seem insulting and offensive to the student concerned. The issue is not that the teaching team should only focus on what the students are interested in, as that would defeat the purpose of studying at masters level. It is more a case of finding out students' own learning agenda and creating a dialogue with students both collectively and individually to share the content and academic literacy objectives of the programme. All of this is connected to a more

general need for the teaching team to ascertain students' learning goals and priorities, and to develop a shared agenda for teaching and learning.

Louisa, as a committed teaching professional and a high-achieving student, seemed to have been able to use well the formative support already in place. However, her observation that the students on the programme regarded the marks awarded to their assignments as some sort of 'personal secret'. A possible explanation is that the students understood that their marks could be seen as an index of their personal achievement and ability in a competitive education system (likely to be informed by their previous educational experiences), therefore they were keen to guard their marks, lest they were judged inappropriately by others. This understanding of the meaning of marks reflects the still dominant influence of classical test theory that is concerned with differentiating and discriminating. From the point of view of this discussion though, the post-feedback 'non-disclosure' raises at least two related pedagogic issues. Firstly, the guarding of one's own mark, and by extension the feedback comments, precludes any peer discussions that can lead to productive understanding of the feedback comments; a potential for collective learning is negated. Secondly, the view that assignment marks are 'non-shareable' suggests that the LOA approach adopted by the teaching team was still seen by students in terms of summative evaluation. All of this raises a number of question such as: Should the teaching team actively seek to find out how students use the feedback comments? Should there be further time-tabled post-feedback provision to promote active discussions on feedback comments and use them to feed forward to inform future work?

Intellectual and Pedagogic Dispositions in Teacher Feedback

Teachers working in formal educational settings such as schools and universities tend to have little say in set-piece system-wide summative assessments, e.g. national school leaving examinations. The design, administration, rating and reporting of student performances tend to be externally organized. In contrast, in classroom-based LOA the teacher has a good deal more autonomy within the structural constraints of the curriculum framework. Whether it is conducted as part of an 'ordinary' teaching-learning activity (sometimes referred to as 'on the run teacher assessment') or as part of a stand-alone curricular event (e.g. an end-of-module assignment or test that can serve both summative and formative purposes), the teacher can have more say in the assessment focus in terms of disciplinary knowledge, criteria of judgement and the kind/s of feedback they provide.

Feedback in LOA is meant to be helpful in promoting student learning. It follows that teachers have in mind both the what (content) and the how (way/s of understanding and representing content) when they provide feedback comments. The 'what' and the 'how' are in turn influenced by teachers' own conceptualization of how students learn. Learning is a complex phenomenon that can be understood from a variety of perspectives. For the purpose of this discussion, I will draw on the

work of James (2006, 2008; Pritchard 2008) and characterize learning from three conceptual perspectives through the lens of assessment. These are briefly described as follows:

Assessing Learning as Receiving and Retaining Information by Individual Students

This perspective construes learning as:

- individualistic (focussing on an individual's ability, intellect, mind and so on)
- transmissionist (teachers transmitting knowledge and skills to students, some would liken this process as 'filling an empty vessel')
- passive (students receiving knowledge and skills from others, they do not act agentively to determine what to learn and how to learn)
- piecemeal (complex knowledge and skills can be learned on a decomposed basis)
- a consequence of responding to external stimulus
- repetition (repeating stimulus can lead to habitual response)

Many of these features are associated with behaviourist views of learning. This perspective will be referred to as Individual-Transmission.

Assessing Learning as Individual Sense-Making

This perspective construes learning as:

- an active cognitive process (involving thinking by the student)
- highly individualised (no two students are the same)
- building concepts/schemata (being able to link different pieces of information and understanding in a narrative)
- deploying known concepts & ideas to make sense of new information

Many of these features are consistent with constructivist views of learning. This perspective will be referred to as Individual-Construction.

Assessing Learning as Joint Activity with Others

This perspective construes learning as:

- a socio-cultural process (learning takes place through interactions with others)
- situated (context and contingent conditions can affect learning)
- involving thought and action (cognition is embodied)

- a shared activity (learning can be achieved through social participation – 'thinking' can be achieved collectively).

Many of these features broadly resonate with social interactionist and socio-cultural theory views of learning. This position will be referred to as Joint-Construction.

Given that students can and do respond to different kinds of pedagogic guidance in diverse ways, it is not the purpose of this discussion to endorse any particular perspective on learning. It is also important to acknowledge that teachers do not necessarily adopt any one of the perspectives discussed above exclusively in all aspects of their assessment work; indeed they may take an eclectic and hybridized approach because it reflects their pedagogic beliefs and/or it is judged to be strategically necessary. Feedback comments and guidance can be seen as a window into the teacher's intellectual and pedagogic disposition. By looking at feedback comments as a heuristic and student responses to them, we may begin to see the options for post-feedback follow-up. The benefits of examining one's own feedback with this in mind can facilitate teacher reflexivity (the capacity to think about one's own teaching from different points of views) and a sense of professional independence (the willingness to question established values and practices) (Leung 2009, 2013). Some examples from the feedback comments we have seen earlier will now be used to illustrate this point.

The opening section in Sadie's Grammar assignment attracted the following comment: 'The beginning of this analysis suffers from a failure to stick to the question asked. You were not asked to discuss the difference between pedagogic grammar and reference grammar, or the nature of pedagogic grammar …'. The deictic reference is clearly directed at Sadie as an individual. The phrase '… a failure to stick to the question asked' suggests that Sadie should have paid more attention to the task set (responding to external stimulus) and should not have taken upon herself to interpret it (i.e. acting agentively). This comment can be characterised as related to *Individual-Transmission*. Sadie did not seem to appreciate this kind of comment.

Sadie also seemed to have misunderstood or misinterpreted the main task in her Phonology assignment (Figs. 5.4 and 5.5). The feedback comment 'For the purposes of this assignment, I would have expected your analysis to focus on … RP – maybe using as a check list descriptions of RP that can be found in Collins and Mees …' is clearly directed at Sadie as an individual. The mention of reference materials though suggests that the response to the task set should be understood in terms of the course content and deploying known concepts and ideas to tackle the set task (i.e. building on schemata already in place). This comment can be characterised as *Individual-Construction*. Sadie seemed to appreciate this comment, saying '… I don't take that too negatively, I take that as a suggestion for me to do something'.

Amy was disappointed with her mark for the Linguistic Analysis assignment and did not paid careful attention to the feedback comments (Fig. 5.6). Had she taken the trouble to read and reflect on the comments, she would have found that some of them were concerned with her not articulating her ideas in the appropriate academic language that she was expected to have, e.g. the question marks over the words

'investigation, support and implementation' (i.e. making use of knowledge already in place). The comment 'be clear from the start which you are dealing with in this essay' on Amy's indeterminate choice of 'to write or speak' (line 3, Fig. 5.6) suggests that, in the tutor's view, Amy did not respond to the set task (i.e. not responding to external stimulus as expected). These comments can be characterised as related to *Individual-Transmission*.

Miguel did not welcome the feedback on his writing style: 'Better: the result of this … was generations' (Extract 1, Fig. 5.7), but appreciated the comment on a content matter: 'It would have been worth examining alternatives to Willis (1996) framework' (Extract 2). Miguel clearly felt that his writing style was a matter of individual preference; at the same time the tutor's comment also reflected the tutor's own preference. From Miguel's point of view the tutor's comment seemed arbitrary and unwarranted. It may well be that this impasse was due to the tutor and Miguel having different background experiences and these experiences were not shared. In so far as this tutor comment seemed to be premised on the idea that there are preferred ways of expressing meaning and these should be used as models, it is aligned with the assumptions underlying *Individual-Transmission*. The content-related comment signalled that there was a body of established work to consult and Miguel should actively engage with it; this comment can be characterized as related to *Individual-Construction*.

Louisa was clearly able to make use of all the feedback comments in productive way to improve her work. Her remarks on the 'cagey' way in which students received their marked assignments, and the assignment mark being 'a private secretive thing' represent an interesting observation. The students perceived the marking (of their assignments) was a part of a competitive system, and they did not have a sense that feedback comments could be shared and used for collaborative learning. In other words, there was little awareness of use of assessment for *Joint-Construction*. One possible reason is that the teaching team did not provide, or did not succeed to make explicit, guidance on the value and usefulness of sharing feedback comments.

The examples above show that by examining feedback comments can reveal the intellectual and pedagogic stances embedded within them.

Concluding Remarks

In this chapter I suggest that the teaching-learning interface is only but one of the constituent components of LOA. LOA is intimately connected to elements of the curriculum infrastructure. By focussing on feedback comments and students' responses to them, it is possible to obtain a better understanding of how students react and respond to feedback comments, what kind/s of intellectual and pedagogic positions are embedded in the formative support, and how the efforts made by teachers to provide feedback are or are not having the intended impact. In the process we have seen that LOA needs to be more than a one-stage and one-direction

process; a post-feedback follow-up by teachers should be considered. The iterative process of teacher feedback-student response-teacher follow-up can take the form of individual student consultation and/or teachers reviewing the processes, and the intellectual and pedagogic nature of the feedback they provide. Crucially though, for any of this discussion to happen at all, it is necessary to carry out situated local research that take account of the students' responses and the curricular provision in particular contexts. While it is certainly the case that situated research of this kind can be greatly enriched when undertaken collaboratively (see Chaps. 6 and 8 in this volume by Baker and Germain and Hill and Ducasse), it may also be realized through investigations, as reported here, in which an individual assumes the dual role of teacher and researcher. In this form of practitioner research, the teacher role is critically involved in sensing and articulating an issue or a question to be addressed; the researcher role provides (indeed demands) the intellectual space and capacity to put some distance between one's investment in one's own practice and how others perceive it. Examining multiple perspectives and exploring empirical data through different analytic and conceptual lenses are likely to be productive. To achieve this collaboration with teaching colleagues and students is indispensable.

References

Andon, N. J., Dewey, M., & Leung, C. (2018). Tasks in the pedagogic space: Using online discussion forum tasks and formative feedback to develop academic discourse skills at Master's level. In G. Samuda & M. Bygate (Eds.), *TBLT as a researched pedagogy* (pp. 236–263). Amsterdam: John Benjamins.

Berg, B. L. (2009). *Qualitative research methods for the social sciences* (7th ed.). Boston: Allyn & Bacon.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–73.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Maidenhead: Open University Press.

Clarke, S., & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996–1998. *Evaluation and Research in Education, 4*(1), 38–52.

Cowie, B., Moreland, J., & Otrel-Cass, K. (2013). *Expanding notions of assessment for learning: Inside science and technology primary classrooms*. Rotterdam: Sense Publishers.

Crossouard, B., & Pryor, J. (2012). How theory matters: Formative assessment theory and practices and their different relations to education. *Studies in Philosophy and Education, 31*(3), 251–263.

Gibbs, J. (2006). How assessment frames student learning. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23–36). London: Routledge.

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 47–60). London: Sage.

James, M. (2008). Assessment and learning. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application* (pp. 20–25). Abingdon/Oxon: Poutledge.

Leung, C. (2009). Second language teacher professionalism. In A. Burns & J. C. Richards (Eds.), *Cambridge guide to second language teacher education* (pp. 49–58). Cambridge: Cambridge University Press.

Leung, C. (2013). Second/additional language teacher professionalism – what is it? In M. Olofsson (Ed.), *Symposium 2012: Lärarrollen I svenska som andraspräk* (pp. 11–27). Stockholm: Stockholms universitets förlag.

Leung, C., Davison, C., Hamp-Lyons, L., East, M., Evans, M., Liu, Y.-C., & Purpura, J. E. (2018). Using assessment to promote learning: Clarifying constructs, theories, and practices. In J. M. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y.-A. Son (Eds.), *Useful assessment and evaluation in language education* (pp. 75–91). Washington, DC: Georgetown University.

Mann, S. (2011). A critical review of qualitative interviews in applied linguistics. *Applied Linguistics, 32*(1), 6–24.

Mann, S. (2016). *The research interview: Reflective practice and reflexivity in research processes*. Basingstoke: Palgrave Macmillan.

Menter, I., Elliot, D., Hulme, M., Lewin, J., & Lowden, K. (2011). *A guide to practitioner research in education*. London: Sage.

Norris, J. (2016). Language program evaluation. *The Modern Language Journal, 100*(S1), 169–189.

Pritchard, A. (2008). *Ways of learning*. London: Routledge.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomingdale: Solution Tree Press.

Chapter 6
Narrative Inquiry as Praxis: Examining Formative Assessment Practices in a Nature-Based Indigenous Language Classroom

Beverly Baker and Joyce Germain

Abstract This chapter focusses on the assessment of children in an Indigenous language nature-based immersion classroom in Listuguj, Canada. This is in fact a subproject, part of a decade-long university-community partnership dedicated to Mi'gmaq language revitalization across the lifespan—which in addition to early years immersion includes the establishment of an adult program using a teaching method developed by Mi'gmaq community-based researcher-instructors.

This study was collaboratively conducted by a language assessment researcher and an immersion teacher, using narrative inquiry—a method uniquely well-suited to Indigenous contexts and increasingly widespread in second language education research, but rare in language assessment research. Through this study, we explored critical moments of formal and informal assessment to reveal a number of guiding principles in the teacher's approach to integrating teaching and assessment. For example, summative assessment can happen either inside or outside, depending on where the best performance can be obtained from the student, and instruction and assessment are mediated by interaction with seasonally appropriate cultural activities and artifacts. Through this project, both members of the researcher-teacher team had the opportunity to reconsider and transform their understandings of assessment and educational theory.

B. Baker (✉)
University of Ottawa, Ottawa, ON, Canada
e-mail: bbake3@uottawa.ca

J. Germain
Alaqsite'w Gitpu School, Listuguj Mi'gmaq First Nation, Pointe-à-la-Croix, QC, Canada

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_6

Introduction

In their introduction to this volume, Poehner and Inbar-Lourie observe that research into classroom-based second language (L2) assessment has often sought to document teacher assessment practices, sometimes with an intent to build theory from the ground up, while others have regarded teachers in a more technical role, responsible merely for putting into practice ideas that have already been fully developed by researchers. As Poehner and Inbar-Lourie explain, the aim of this book is to explore possibilities that arise when theory/research and practice are understood as informing one another and emerging most effectively through researcher-teacher partnerships. In this chapter, we outline a researcher-teacher collaborative study of a nature-based Mi'gmaq language immersion classroom in Listuguj, Canada. We, the researcher (Beverly) and the teacher (Joyce), will first present the teaching context and details of our project, which took place in late 2015 and early 2016. We will then discuss the insights we gained from this collaboration from the praxis perspective that is the focus of this volume. We understand praxis as a dynamic relation "in which theory guides practice but at the same time practice influences, and if need be, changes theory" (Lantolf and Poehner 2014, p. 27). In other words, we will examine practice not as an application of theory or held up to judgment to the extent that it faithfully applies a theory. In this project, the validity of theory will be examined with reference to Joyce's classroom practice.

The context of our work is Listuguj First Nation[1] located in Eastern Canada, in the Canadian province of Quebec. The community is separated from the province of New Brunswick by the Restigouche River, and community members move freely between the two provinces for work, study, and shopping. Of its current population of about 2000, only 5% are fluent speakers of the Mi'gmaq language, and they are almost all over 50. However, there is more cause for optimism than these numbers might suggest: the community in recent years has established a thriving immersion program from nursery to grade four as well as adult classes at its Education Directorate that together have proven popular and effective in rekindling interest in the language at all ages (see Sarkar and Metallic 2009, for further details about the community and success of its recent language revitalization initiatives).

"Miss Joyce" has been teaching Mi'gmaq immersion in the community's elementary school (Alaqsite'w Gitpu School) for 20 years. She has developed programs for Kindergarten and Grade 1 immersion (4–6 years old) and 5 years ago incorporated an outdoors nature-based component, called Nipugtugewei Kindergarten. In this model, mornings are spent in the classroom and afternoons are spent on the land. Both the in-class and outdoors components address provincial curriculum concepts in science, social studies, math, and language arts, as well as

[1] Canada's Indigenous peoples belong to one of three general groups currently referred to as First Nations, Métis, or Inuit. These words are always capitalised. The word Indigenous is not consistently capitalised in the literature, but we have chosen to do so here.

physical education. In the immersion program, all these subject areas include language-related and cultural learning outcomes.

Nature-based or forest schools are rapidly growing in popularity in North America from their beginnings in Scandinavia in the 1980s. Joyce was moved to introduce this component to her teaching because she was convinced of the physical and pedagogical benefits of time spent outdoors, as well as the potential for cultural and spiritual enrichment offered by this model. Proponents of nature-based schools point to benefits such as improved physical fitness; increased motivation, attention, self-regulation and self-discipline; improved collaborative and problem-solving skills, and reduced stress (Children and Nature Network 2012). Most important, nature-based schools are well suited to Indigenous contexts. Perhaps it would be more accurate to state that proponents of nature-based schools are recognising that this supposedly innovative approach to education is simply what Indigenous peoples have always practiced: "Above all, Aboriginal pedagogy is centred on observation of nature and trying to learn the lessons that the plants, animals and natural systems can teach us" (Forest School Canada 2014, p. 14). Like Indigenous pedagogy, nature based schools follow principles of discovery-based, experiential learning.

Joyce has observed that the outdoors component has created naturally occurring contexts for meaningful language learning in addition to increasing her students' understanding of and respect for nature. This model has also allowed her to bring in cultural, spiritual, and traditional aspects of Mi'gmaq ways, such as traditional stories and teachings of the elders; arts and crafts; traditional dancing; ceremonies; and lessons from the plants and animals. From the perspective of Vygotskian theory (e.g., Lantolf and Poehner 2014), the outdoors environment is a powerful mediating force, one which has formed the traditional basis of all Indigenous learning. Indigenous knowledge is holistic, "incorporating spiritual, ecological, human and social experiences into one understanding of Native people's place in the universe" (Hoare et al. 1993, p. 48). By going back on the land, this connection is being re-established—allowing students to express their cultural heritage and reclaim their connection to the land through the language.

Our Project: Examining Assessment Practices Indoors and Outdoors

As part of a university-community research partnership established to support and document revitalization efforts in Listuguj, Beverly visited the community in the summer of 2015 to participate in a "Mi'gmaq language workshop," a day of round tables and talks designed to raise awareness of the language in the community. Beverly (a language assessment specialist) had collected information on the ways that Indigenous communities across Canada were approaching language assessment (See Baker and Wigglesworth 2017). She came to present what she had discovered

during the workshop. Joyce participated in this discussion and shared her recent successes obtaining quality assessment information during the outdoors component of her program. This was where the seeds for our collaboration were planted. Following Beverly's suggestion, we decided to spend time discussing and examining Joyce's assessment activities in the classroom and on the land, in order to share effective practices with other teachers—especially those who were interested in introducing a nature-based component into their immersion classroom.

Focus on Moments of Assessment

In beginning our project, we did not explicitly set out to take a theoretical framework as a lens through which to view our project. Our praxis-oriented project focusses on identifying and discussing moments of assessment in the classroom first, then investigating connections with current assessment theories in the light of these classroom-based insights. However, Beverly came to the project with background knowledge in classroom-based assessment (CBA), dynamic assessment (DA), and learning oriented assessment (LOA) from the field of language assessment. Briefly, CBA has been described by Turner (2013) as involving "strategies by teachers to plan and carry out the collection of multiple types of information concerning student language use, to analyze and interpret it, to provide feedback, and to use this information to help make decisions to enhance teaching and learning. Observable evidence of learning (or lack of learning) is collected through a variety of methods, and most often embedded in regular instructional activities" (p. 66).

DA has been preoccupied with the exploration of assessment integrated with instruction in obtaining information for teacher decision-making (see chapters in this volume by Davin & Herazo and Poehner & van Compernolle). Growing from work in formative assessment (Black and Wiliam 1998), LOA also represents a reorientation from a traditional emphasis in the domain of language assessment away from large-scale standardized tests and psychometric analyses to the examination of planned and unplanned assessment in the language classroom. LOA has an explicit focus on assessment in the service of learning through evidence elicited during language production (Turner and Purpura 2015, p. 260). Turner and Purpura (2015, p. 255) explain that

> [t]he LOA approach is not to be confused with nor is it in competition with other current L2 classroom assessment techniques (e.g., diagnostic, Alderson 2005; dynamic, Lantolf and Poehner 2011), but certainly shares common characteristics with them. Its premise is to begin with learning, that is, to prioritize learning when considering the interrelationships across instruction, assessment and learning.

In LOA, as in DA, assessment is viewed as a continuous and multi-dimensional process that includes interactive, cognitive, and sociopolitical components. In both traditions, there is a marked preoccupation with exploring the most appropriate feedback (by teachers or peers) to facilitate learning. The majority of DA research

in second language (L2) contexts has taken place in classroom settings, and LOA research, while still in its infancy, is also classroom-based. This project represents an innovative research setting in language assessment in that it examines assessment practices in indoor and outdoor settings combined, as well as in an Indigenous setting.

Joyce's assessment practice in the classroom exhibits many features of what Baker and Wigglesworth (2017) have identified as representing an Indigenous worldview applied to classroom assessment. These features include the following:

- *Content of assessment*: In addition to demonstration of knowledge, in Indigenous classrooms emphasis is generally placed on the process rather than the outcome, and the (primarily oral) expression of cultural knowledge with performative activities such as stories and jokes, rather than with more individualistic western academic forms such as presentations, or oppositional forms such as argumentation, persuasion, and debate.
- *Processes of assessment*: a near-absence of negative feedback is common in Indigenous educational settings, in order to decrease anxiety. In terms of the timing of assessment, evaluation activities do not follow a strict schedule but happen only when the student has achieved the desired level and/or is sufficiently prepared. Extensive feedback is provided on progress. The timing of assessment is often directed by the student, or is decided by a teacher or Elder only when they judge the student to be ready. Teamwork is valued to accomplish goals, with less emphasis on individual accomplishment. Collaborative processes are often assessed rather than individual processes.

The Collaborative Relationship in Research on Praxis: Anticipated Challenges

Winkler (2003) discusses challenges in conducting a narrative-based collaborative project with a number of primary school teachers in South Africa (similar to the current study, discussed in detail below). In her words, she had hoped that their collaboration "would build trust and provide opportunities for sharing successes, difficulties, questions and dilemmas, until powerful stories emerged" (p. 395). However, challenges emerged in terms of trust: Winkler shares that at least one teacher in the project felt that she was being judged professionally on her teaching, despite Winkler's assurance that she viewed the teachers as equals and was not evaluating their practice. While a praxis orientation recognises the value of teacher practice as informed by and informing theory, teachers may still feel like their work is being judged in relation to some abstract academic standard. In approaching Joyce, Beverly realised how essential it was to establish trust and create a relationship as colleagues and equals (see also Goldstein 2000 for a discussion of this). Beverly did not initiate this project because she had anything to *teach* Joyce; she did it because of what she had to *learn*.

Teacher-researcher collaboration in a First Nations setting requires an additional level of attention: like many other First Nations Communities, and indeed Indigenous peoples all over the world, Listuguj has had to contend with outside researchers who have arrived with their own agendas and may have conducted research which was not judged by community members to be useful or to have treated community members with sufficient respect. In their *Principles and guidelines for the protection of the heritage of Indigenous people* (1995), the United Nations website states, "To protect their heritage, Indigenous peoples must…exercise control over all research conducted within their territories, or which uses their people as subjects of study." Therefore, working together as co-researchers was not a methodological choice but an ethical imperative to ensure that any work done is in the interests of the community and its members first and foremost.

Research Objective

In this project, Joyce and Beverly were interested in capturing and reporting the essential elements of Joyce's approach to classroom assessment. By laying bare these elements, we hoped to achieve a twofold objective: (a) to stimulate our own reflections regarding practice-theory connections, to use as a springboard for our continuing professional development, and (b) to articulate these reconceptualised theories in a useful way for other teachers in similar situations. Our initial research question was therefore the following: How is assessment accomplished in a nature-based Indigenous language immersion setting?

Method: Narrative Inquiry

Narrative inquiry was our methodological choice for our project: Through telephone interviews, Joyce was encouraged to share anecdotes about her daily assessment practices in the classroom, and these anecdotes became the primary source of data for narrative analysis. Narrative inquiry generates data in the form of stories, as well as other field-based sources (like photographs), and is focused on the meaning that people make of what happened in the story (see Bond and Mifsud 2006; Clandinin and Connelly 2000; Chase 2005; Creswell 2008; Maple and Edwards 2009; Norton and Early 2011; Polkinghorne 1995, 2007; Riessman 2008; Swain et al. 2011; Winkler 2003). Stories can be defined in the most basic terms as any time-based chronological sequence of events that has the following three elements: characters in interaction, time, and place (see Barkhuizen 2008; Clandinin and Connelly 2000). From an ontological standpoint, if we accept that there is a moral imperative to better understand human experience, then human experience can be usefully understood through a form like narrative, which allows for the recounting of individual events placed within a wider social and relational context. Polkinghorne

(2007) asserts that "[s]toried texts serve as evidence for personal meaning, not for the factual occurrence of the events represented in the stories….[T]he storied descriptions people give about the meaning they attribute to life events is…the best evidence available to researchers about the realm of peoples' experience" (p. 479). This type of inquiry is well established in educational research (Barkhuizen 2008; Rosiek and Atkinson 2007), especially as teachers are increasingly included in the research process. Narrative methodologies are being seen more and more in L2 education research as well (e.g., Norton and Early 2011), but they are still rare in language assessment research, despite their clear potential: classroom assessment practices are sometimes not conscious or identified as assessment by teachers; they may be effectively revealed if teachers are given the opportunity to tell stories about their in-class practice. Two projects that illustrate the potential of narrative methodologies within language assessment research can be found in this volume. In the chapter by Scarino, teachers' narratives of learning and assessment are among the research tools used to trace and conceptualize the evolving process of learning and assessment in an interlingual and intercultural setting. In the chapter by Harding and Brunfaut, the authors employ narrative inquiry to gain an emic perspective into the effectiveness of a teacher-researcher partnership in the context of a larger project to design a national language test.

This methodological tool was also chosen because of how well-suited it appeared to be for research in Indigenous contexts. Canagarajah (1996) suggests that this methodology is ideal for groups that have been traditionally excluded from scholarly conversations. Allowing Indigenous practitioners to tell their own stories can help to decolonize the research process (Smith 1999; Steinhauer 2002). In addition, narrative research is consistent with Indigenous values and emphasis on the oral record. Storytelling plays a central role in Indigenous pedagogy and cultural practice, emphasized in curricula for Canadian Indigenous language and culture education programs (Baker and Wigglesworth 2017).

Procedures

We engaged in four telephone interviews of 60–90 min each over the course of 4 months, from December 2015 to March 2016. The interviews were spaced out partially because of Joyce's busy schedule, but also they allowed her some time for observation and reflection in between sessions. Generally, Joyce would make informal notes leading up to each meeting and decide what to speak about, and sometimes Beverly would ask her about the significant events that stood out since the last time they talked. Sometimes at the end of a session Beverly would ask for further details about certain topics that Joyce mentioned. For example, when Joyce mentioned about how she integrates the Seven Grandfather teachings, Beverly asked if they could devote more time in a subsequent interview to this.

These sessions did not follow the traditional question and answer format of an interview. Generally, Joyce had made notes and took the initiative to start telling stories about her recent teaching and assessment experiences without prompting. As

the interlocuter, Beverly strived to give up authority, aiming to cultivate "respectful curiosity" (Maple and Edwards 2009). Stories are not told without listeners, and Polkinghorne (2007) reminds us that "[f]ocussed listening and exploration can bring to the fore more of the intricate multiplicity of an experienced meaning" (p. 481) and that "[i]t is the interviewer's task to empower participants by acknowledging that they are the only ones who have access to their experienced meaning" (p. 482). Norton and Early (2011), in their narrative research on researcher identity in teacher-researcher collaboration, discuss how "storytellers use stories to reflect upon life and to explain themselves to others. The storyteller is thus seeking a human connection with the audience and is striving for an affective response" (p. 420).

In their discussion of commonalities in genre conventions in teacher narrative research, Rosiek and Atkinson (2007) identify "experiential narratives" (p. 513), which are characterised by teacher reflection on their own lived experience. This is the best description of the narratives that were produced here. In this case, Joyce was asked to identify assessment moments that occurred in her teaching and use them as a starting point for her own reflection, and then for our subsequent discussion.

Analysis of Interviews

Polkinghorne (1995) makes the distinction between narrative analysis and analysis of narratives. Narrative analysis utilizes "narrative reasoning" by transforming data into stories, while analysis of narratives identifies themes within narrative data. We engaged in the second activity: We separately made initial notes made on our transcribed interviews, keeping our general research objective in mind and naming the most important themes. As we worked on this separately, we asked ourselves, "What have we discovered through these talks that would be useful to other teachers in the same situation?" Our emphasis was on utility of our findings to further the cause of similar initiatives.

Two months after the completion of the interviews, Beverly visited Listuguj for an intensive 2-day face to face working session to interpret the transcripts. During this joint interpretation process, we re-read the transcripts together and reviewed our notes as well as hundreds of photographs that Joyce had taken during the outdoor component of the class from the beginning of the year. While reviewing the photographs, we were able to find examples of the assessment moments that had been discussed during the interviews. In some cases, the pictures prompted more stories. For example, a photograph of a child holding a caterpillar (Fig. 6.1) prompted Joyce to tell the story of how this child's discovery led to a conversation about life cycles, which made Joyce decide to introduce the science unit on life cycles in the classroom the same week, even though this unit had been planned for much later in the year (see Table 6.1).

From a DA perspective, social mediation (in the form of photographs) prompts students' memory from the day before, which is often sufficient for the students to

Fig. 6.1 Discovering a caterpillar—initiating an instruction-assessment cycle. (Photo: Joyce Germain)

Table 6.1 An instruction/assessment cycle as initiated by a student's discover

Activities in class and on the land	Language assessment moments
Initial event: A student (S) finds a caterpillar in the woods. The class gathers around and observes it. Discussion involves using the 5 senses. Teacher (T) takes a photograph.	*Formative assessment:* T determines what language is already known (*small*) and what needs to be introduced (*change/grow*).
The following weeks: The class watches a video about metamorphosis (caterpillar-butterfly). Video sound is turned off and T explains process in the language. A poster is put up in class (life cycles, from the year's science curriculum), provoking more discussion Life cycles reinforced daily indoors by drawing attention to the poster and the photograph from the initial discovery. Outdoors, connections made with other life cycles (tadpole-frog).	*Formative assessment:* T uses photo to stimulate memory and asks questions, *e.g*., "How does the caterpillar grow in the second stage?" Role play/TPR: T asks Ss to act out stages of metamorphosis in the language *Summative assessment:* T continues questioning until all students demonstrate that they can discuss all the stages.
Bridge to subsequent unit: Student asks if both butterfly wings are the exact same, initiating a discussion of symmetry—a math concept that initiates a new unit.	*Formative assessment:* Students paint on half a sheet of paper and fold to create a symmetrical image (with language introduced for math concepts—previous ones like "half" and new ones like "fold," etc.).

volunteer their understanding of the language and concepts that were introduced. If the students don't volunteer the concepts and language, Joyce uses comprehension checking through the wh-questions (*who what where when why*) (Fig. 6.1).

In deciding what final themes to include, Joyce was informed by her in-depth knowledge of the children, her cultural knowledge as part of her community, her pedagogical knowledge (including knowledge of nature-based education) and her professional intuition. For example, in discussing the connection between the outdoors component and language learning, Joyce discussed the language teaching method TPR (total physical response), demonstrating her knowledge of language teaching methodology. Beverly brought to bear her own language teaching experience in these discussions, as well as her knowledge of research in applied linguistics, language assessment, and education. We were essentially doing a *functional analysis* of the stories, which Bruner (1991) describes as making use of the stories to better make sense of Joyce's lived reality in the classroom. This was an occasion to systematically locate and map out the coherent elements related to assessment in Joyce's typical days with her students.

Presentation and Discussion of Themes

Here we present evidence in the form of themes as revealed by our joint interpretation process. In presenting these themes we aim to convince the reader of their credibleness and value, based on our systematic analysis, even though these interpretations may not be the only possible ones (Polkinghorne 2007). This analysis has to involve the person telling the story, because it is their reality. That being said, storytelling is inevitably a co-construction between the storyteller and the listener. Therefore, the final themes presented here are best described as a negotiated product of Joyce and Beverly together.

The major themes relate to the identification of evaluation techniques that align with Indigenous features of classroom assessment (Baker and Wigglesworth 2017) as well as with nature-based pedagogy. While the focus is on assessment, instructional and assessment activities are often discussed together because of their indivisible nature. As we present these themes, we make connections, as they are relevant, with principles of dynamic assessment/learning oriented assessment as well as with praxis—the interplay between theory and practice that emerged as we negotiated key meanings emerging from these stories.

Inquiry-Based, Discovery-Based Learning and Integrated Assessment on the Land

The first dominant theme concerns how inquiry-based and discovery-based learning characterised the nature-based portion of the school day. Joyce discussed how the students' discoveries outside often determined the next curriculum elements to be covered—showing how children had a certain amount of control over their own

learning. In Joyce's own words during our interviews, she says, "I know the whole year's concepts but I'm finding myself actually teaching when I see the opportunity and when it's ready for them to actually learn it, you know?" This is evidenced further by the following excerpt from our interviews:

> J: And so as we were climbing up, and of course the other kids are all excited, they were following so one behind the other and we came across a tree that was [laying across the path]….so I said "Yeah!" miti's nisiet [tree falling], nisiet, you know? And so, [the students said] "Miss Joyce," Talatigei, "What do I do?" I said, "Well," Angitasi, "you have to think." And I say these words often to them, you know? Trying to problem solve the things that they see or want to figure out, I try to encourage problem solving…

Joyce is in fact recounting an assessment moment, because she is using their production and questions to decide that the students are ready for the introduction of a new concept. She also realised that the students understood the present tense (falling) and the next day introduced the past tense to describe the picture of the fallen tree (Fig. 6.2).

The following extended excerpt is another illustration of the introduction of a concept—in this case, a science concept of the earth rotating around the sun—when she saw the opportunity:

> J: [W]e went trekking and it was sunny out….I took a piece of stick and I said, of course vocabulary is so important so I would say, "Oh! Look at "Aqate'n". So they know what aqate'n is already by now because we're going in the second term…. And I took a piece of branch and I stood it up in the snow, alright, it was about two feet high and I said "Ok, we're

Fig. 6.2 Over and under the fallen tree. (Photo: Joyce Germain)

> going to stick this in the snow" and I said it in Mi'gmaq… "I want you guys to look at the na'guset (which is the sun)." So they see it, it's nice and bright. I said, "You're going to see the shadow of the stick," and they would say, "oh, ok!" I put another piece of stick and I laid it down where the shadow was…. I said, "That shadow is going to move." I said, "we're going to go for a walk and when we get back we're going to see that that shadow moved away from the stick." And they would ask me, "Why?" I said, "the earth spins and the sun goes around the earth" … and I said it in Mi'gmaq….[S]o when we came back, you know, the kids were all running anxious to see it. "Ms. Joyce, aqate'n gumuj, look at the stick!"

Language and conceptual knowledge are often assessed at the same time, as students are asked to observe and repeat. Observation and emulation are key elements in nature-based learning, in Indigenous pedagogy, and in DA, through cultivation of the group zone of proximal development. Joyce recognizes the varied responses of her children to her prompts and models, and also observes that some students respond more effectively outdoors. Therefore, she has established a personal set of dynamic assessment procedures which align with Indigenous understandings of appropriate assessment. This practice is also reminiscent of LOA, which advises planned elicitation embedded in instruction—as the following excerpt illustrates:

> J: [In front of the log which had fallen across the path] And so the little kid said, "Well! I'm going to crawl under!" Of course they say it in English because they don't know the word but I had to teach him or her the concept of under. I said, "Well? Under (…)" I said, "you have to say it in the language," I said. And that's the…password. "In order for you to go under you need to say the word in Mi'gmaq." So I would say it in Mi'gmaq and he would repeat it! And by the time I got to ten or twelve kids that were behind and they're hearing it and they're actually doing it, they'd remembered much easier. The next morning when I reviewed—because I take pictures, and when I do review the next day, showing the pictures [of] each one of the children going under or over they were able to remember the word…. That's how I bring out the concepts and that's how I know how I can evaluate them also. To see if they remember the word, actually…or understand the concept."

As this excerpt demonstrates, Joyce observes her students, notices a link to curriculum and an opening for language and conceptual learning, then instantly creates an activity where students receive built-in repetition with feedback as necessary. She then reviews the following day with photographs, and collects further assessment information.

Assessing for Success

Joyce discusses how students are evaluated formally only when they were ready to succeed (to the extent possible, given the school calendar). She explains this process in the following comment:

> J: That's how I assess them. And then when I'm really, really sure that they've learnt all of the concepts then I can [stop teaching and give them their grades]. You understand? ….
>
> They understand now…. I know that I won't have to repeat, you know the ones that are a bit worse off or didn't know.

In terms of assigning final grades, Joyce describes her approach here:

> J: And I do assess them formally….I have a grid… an 'A' would be a student that is able to understand commands or questions and to be able to respond to the questions or to comply to a command… A 'B' would be a student understanding the command…but they would hesitate to respond to questions and at times they would need encouragement or assistance…. A 'C' would be a student that sometimes understands commands and questions, not all the time, you know? …. I would have them repeat or I would reinforce it by actually showing them, you know, physically showing them. Visualizing, you know, the things that they need to learn or work to be able to understand. A 'D' is if a kid just really doesn't understand the question or the command or whether [he] often needs help. But I'm never at the 'D' [Both Laugh]. I'm always 'A', 'B', and 'C' at times, you know…

Joyce describes each grade not in terms of success and failure, but in terms of the individual support each needs to be successful. In other words, feedback is automatically sensitive to the Zone of Proximal Development of each individual learner. She also discusses how she assesses students summatively both inside and outside—wherever she can get the best response from them.

The Use of Environmental and Cultural Artifacts

In addition to pictures, drawings and other classroom manipulatives, elements of nature (stones, sticks, feathers, etc.) are used to mediate the language learning outdoors and sometimes brought indoors to stimulate memories in the classroom. These natural artifacts are also called "loose parts" (Better Kid Care Program 2009; Play England 2016). For example, during the unit on *inuksuit* (stone towers/piles created by the Inuit in the north), students went to a beach to examine the stones and they chose the most appropriate sizes and shapes of stones to build their own inuksuit in the classroom and describe them for a subsequent assessment.

In our interviews, Joyce discussed how the visual and tactile nature of these manipulatives stimulate students' memory and create a tangible spiritual and cultural link to another Indigenous culture. Culturally based activities are usually closely related to their own community and ecosystem, and often linked to the seasons. For example, in the late summer and fall, Joyce sets up field trips for apple, hazelnut, and sweetgrass picking, and the children plant trees for the Elders. In the winter, students do animal tracking in the snow and go ice fishing. In this context, elements of nature are simultaneously cultural and spiritual artifacts. There is no artificial distinction made, as in western cultures, between man and nature, and the outdoor space that they are exploring every day is the land of their ancestors. Therefore, there is an explicit spiritual connection to all outdoor activities which serves as a mediating factor in the language being learned. For example, in an Indigenous nature-based group, students cannot remove environmental artifacts from the land without making an offering to give thanks (e.g., with tobacco). Therefore, one of the first language functions to be learned in Joyce's Kindergarten is how to offer thanks for the gifts of nature.

As previously mentioned, Joyce has internalized the curriculum for the grade level she is teaching and is able to recognize when student production provides evidence of meeting content or language learning objectives designed to be covered at other points in the school year. Joyce therefore links language learning and curricular learning naturally during cultural activities.

Unplanned Play as a Way into Assessment

Joyce recounted how the students sometimes had jobs to accomplish during their outdoors play—like finding a stick for the next day's ice fishing activity, for example—but they also had a great deal of unstructured play time. There is evidence of many of the different types of play recommended in nature based schools (see, e.g., quality assurance standards such as Play England 2016). The children developed their own games, did role plays, skipped rocks, and built structures, among other activities.

Proponents of nature-based pedagogy, as well as play-based learning, argue for the pedagogical possibilities of play. Joyce spoke often about how opportunities for cross-curricular assessment "on the fly" (Heritage 2007) arose during the students' unplanned play: students explained math concepts as they built snowmen; demonstrated sight word recognition as they drew in the sand with sticks; and showed evidence of phonological awareness while singing, chanting and clapping.

Enactment of a Spiral Curriculum

At several points during our interpretation sessions, Beverly became reminded of Bruner's theories (1960, 1961) regarding the spiral curriculum—briefly, that even complex subjects can be tackled with young children, but they are presented over and over again over a long period of time with increasing complexity. The best example of this is Joyce's discussion of the Seven Grandfather Teachings—part of a core curriculum which is covered through all grade levels. These are a series of guiding moral principles widely adopted in Indigenous communities across Canada yet adapted to each community. Joyce mentions the complexity of the teachings but says that students are introduced to them from the very beginning, starting with the names of the animals of the Teachings (though not necessarily in the correct order). Then, they learn the characteristics associated with each animal:

1. Wisdom = gu'gu'gwes (the owl)
2. Respect = gopit (the beaver)
3. Love = wapus (the rabbit)
4. Humility = migji'jg (the turtle)
5. Honesty = tiam (the moose)

Fig. 6.3 "Being a beaver" (showing respect). (Photo: Joyce Germain)

6. Courage = muin (the bear)
7. Truth = gitpu (the eagle)

Joyce recycles this foundational information, and later in the year begins to point out when the students are demonstrating the characteristics of each animal. For example, when a child gave her mittens to a friend with cold hands, Joyce pointed out that he was "being a rabbit," and when another child helped a friend climb up on a rock, she praised him for being "a beaver" (Fig. 6.3). In later grades, more complex language and conceptual knowledge related to the teaching is covered, such as the ritual elements of sharing the Teachings (*e.g.*, the set formula "I brought you this Teaching"). Joyce describes this process in the following excerpt:

> J: [In the early grades, students] are repeating the Seven Animals, they know how to say it and recognize it even by pictures, you know? Sometimes they know with the order, sometimes kids don't; they get mixed up, [but] it doesn't matter.
>
> B: Uh-huh.
>
> J: …Actually, they may not get everything about it but they know in grade one and two [the next teacher] will be teaching it too, you know? So at that grade level really that's all that I'm able to put across because they're at a very early age….[In] the older or other grades, grades three, four, and five other teachers have that responsibility [but] it's in the English language. It's just built up from there, you know you build up from those concepts.
>
> B: And they're getting them in, they're getting them in Mi'gmaq first which is cool.
>
> J: Yeah, and doing it outdoors which is even cooler because when like I said we've seen the [rabbit] tracks, of course the part about the rabbit came up [naturally]….and it just so happens that I had an opportunity to do something because of what one child did [gave her mittens to a friend]. And I reminded her that she was the wapus and that's love.

When I shared the concept of the spiral curriculum with Joyce during our interpretation sessions, we both were struck with how it seemed to capture what Joyce was intuitively doing. Joyce did not learn the concept of the spiral curriculum and attempt to apply it to her teaching. Instead, we are providing a certain validation to the theory: we judge that the theory is useful to the extent that it is borne out in Joyce's practice, given her context and worldview, and it can be useful as a framework for explaining Joyce's successful language teaching and assessment process to others, especially considering its appealing visual representation. Bruner (1960) discusses that the process engendered by a spiral curriculum leads to autonomy and discovery-based, inductive learning, all characteristics of Joyce's approach with her class as previously discussed—and typical of Indigenous views of learning. Consideration of this theory and its relationship to her practice allows Joyce to recognize and appreciate the deeply systematic nature of her own teaching.

Discussion and Conclusions

Summary of Findings

Joyce's stories about her teaching revealed a number of guiding principles in her approach to teaching and assessing her students, lending weight to certain claims of Dynamic Assessment and Learning Oriented Assessment. She makes judicious use of feedback in her formative assessment practice, and generally she integrates instruction and assessment rather than viewing them as separate activities happening at separate times, with different types of production. Joyce also discusses how both formal and informal assessment can happen inside or outside, depending on where she is able to obtain the best performance from the student. In both settings, she cultivates what may be referred to in DA terms as a Zone of Proximal Development by encouraging repetition and emulation until all students produce the language (enabling assessment on the fly). Joyce's in-depth knowledge of the curriculum and her teaching goals means that she can spontaneously seize on unplanned opportunities to elicit language production of interest for assessment.

In addition, Joyce enacts a spiral curriculum in her classroom, with key concepts for the grade level for the curriculum continuously brought back in to be re-verified and built upon. Through our collaborative work, Joyce was able to appreciate the systematic nature of her teaching and to provide validation for the usefulness of the spiral curriculum as an organising framework for her teaching.

Her decision-making is sometimes better described not in relation to assessment theory, but with reference to a First Nations perspective as it demonstrates features of Indigenous assessment. She concentrates on giving positive feedback and describes all students as successful. She concentrates on the instruction and assessment of oral production, mediated by interaction with seasonally appropriate cultural activities and artifacts. Cultural artifacts mediate children's learning as well as

their spiritual relationship with their ancestors along with the learning of the language. The language in turn mediates their development as fully realised Mi'gmaq community members.

Implications for Praxis: Reconceptualising Theoretical Understandings for Future Practice

Through this project both Beverly and Joyce have had the opportunity to reconsider and transform their understandings of assessment and educational theory. From a position of praxis, we were able to ascertain that Joyce's descriptions of her and her students' interactions lend a certain amount of validity to claims made in both dynamic assessment and learning oriented assessment regarding the Zone of Proximal Development as well as the interactive and fluid relationship between instruction and assessment. As the above excerpts show, she makes use of the students' own language to develop that of their peers whenever possible and makes use of their own discoveries and feedback. Also, she does not talk about or view her teaching and assessment activities as separate from each other. This finding aligns with statements of researchers such as Carless (2007), who identify student input and collaboration as well as the integration of assessment and instruction as key elements of LOA. Also, Joyce's success in assessment provides some support for the claim in DA that "Mediation should be only as explicit as it needs to be to prompt an appropriate response from the learner" (Lantolf and Poehner, p. 173). This is similar to descriptions of the provision of corrective feedback on an individual basis in learning oriented embedded assessments, during "talk in interaction" (Turner and Purpura 2015).

However, this project has done much more than allow us to consider the usefulness of educational theories in conceptualizing Joyce's practice. She has been able to reconceptualise these theories in combination with her own Indigenous approach to determine guiding principles to share with her colleagues. Joyce is increasingly being solicited to share her success with language and nature-based educators in both Indigenous and non-Indigenous settings across the region and even the country. For these professional development workshops, Joyce is making use of the notion of spiral curriculum as a useful way of explaining her practise to others. In addition, in presenting the importance of collaboration in nature-based activities, she has found value in viewing these practices as a cultivation of the group zone of proximal development. She enhances these ideas with her unique cultural perspective and her desire to incorporate the specific characteristics of Indigenous settings. For example, Joyce's observation that some students respond more effectively outdoors feeds back into her theoretical understanding and leads her to emphasize to her peers the importance of emphasizing assessment in the physical conditions where teachers can get the best possible response. In addition, she believes it is essential that an explicit spiritual connection to nature and the land must be cultivated

to achieve language, content, and cultural educational objectives. This particular mediating factor is not effectively expressed through Western educational and assessment theory.

The Benefits of Narrative Methodology

Experiential narratives produced during this project have allowed Joyce the luxury of reviewing what she was doing from a slight distance and to appreciate her own work and success more. It also had the effect of de-mystifying the research process so it is seen as being a systematic yet accessible way of examining her practice and explaining it to others. This exercise may also benefit Joyce's future practice: "When teachers articulate and interpret the stories of their practice, their own practice, they develop their personal practical knowledge to the extent that they act in the future with insight and foresight" (Barkhuizen 2008, p. 233).

Narrative methods seem simple on the surface yet open up possibilities for powerful conversations around praxis, across cultural and experiential boundaries. We believe that we would not have been able to obtain such rich information any other way than through our conversations and Joyce's stories. However, it must be acknowledged that when teachers see themselves in a new way it can be jarring at first. Joyce admitted to feeling a little uneasy when reviewing the transcript of our conversations. In addition, Joyce was surprised in Beverly's interest in hearing these stories because she didn't realise their value. This changed when we reviewed and unraveled them together—we both realised the extent to which the stories became a window into the systematic nature of her practice—which on the surface can seem messy and random.

Returning to Anticipated Challenges

Before beginning this project, Beverly had anticipated a challenge in developing a trusting working relationship with Joyce. In this, we were very successful. In our sessions, we did negotiate our expectations on several occasions, as seen from the exchange below:

J: So anyways…are you satisfied so far [with] what I've given you? As far as information?
B: Oh, satisfied? Oh God, I'm more than satisfied. This has been fantastic.
J: …Because I just wasn't sure of what you were expecting.
B: You know what? I wasn't all that sure of what I was expecting either except that I knew that once we got started talking then it would be really interesting; we'd be able to find some really interesting things to share with other people who are in this kind of situation.
J: Right. Right.

Joyce's choices of stories were not made in a vacuum: she recounted what was of interest to her but also what she thought Beverly was also interested in hearing. This is not negative: with a different listener, the story would have been different—the stories that emerged were of Joyce's experience but then interpreted as a team. We both were interested in putting our finger on Joyce's successes in assessment during her teaching, in order to make them more explicit for Joyce herself as well as to be able to transmit them to others. With this common goal, even though our experiences and perspectives are not the same, we were able to work together harmoniously and come to mutual decisions and understandings through respect and trust. What emerged was thus a fully dialogic, awareness-raising theory-practice partnership. That is, a joint partnership oriented by praxis.

One such example of the trust that characterized this partnership concerns our joint decision making regarding how to disseminate our findings. Our research output must be mobilised primarily through Indigenous channels, such as practitioners' conferences and resource banks for Indigenous educators. Beverly also believes an important aspect of this project is to influence academic researchers to engage with both current theoretical understandings as well as with similar inclusive research partnerships. Therefore, contributing to this volume may not have been Joyce's choice—but she trusts in Beverly's commitment to the community and her belief of the benefit of doing so. As Winkler (2003) reminds us, "The validity and ethical defensibility of collaborative research ultimately depends on the critical acknowledgment of multiple realities…" (Winkler, p. 400). Reconciliation between Indigenous and non-Indigenous peoples in Canada depends on not viewing these multiple realities as obstacles in working together, but instead on acknowledging them and moving forward, allied in the common cause of language revitalisation.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. New York: Continuum.

Baker, B., & Wigglesworth, G. (2017). Language assessment in indigenous contexts in Australia and Canada. In E. Shohamy & I. G. Or (Eds.), *Encyclopedia of language and education*. Springer.

Barkhuizen, G. (2008). A narrative approach to exploring context in language teaching. *ELT Journal, 62*(3), 231–239.

Better Kid Care Program. (2009). *Loose parts: What does this mean? Office of child development and early Learning*. Penn State University. Available at: http://extension.psu.edu/youth/betterkidcare/early-care/our-resources/tip-pages/tips/loose-parts-what-does-this-mean

Black, P. J., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.

Bond, T., & Mifsud, D. (2006). Narrative conversation between cultures: A novel approach to addressing an ethical concern. In S. Trahar (Ed.), *Narrative research on learning: Comparative and international perspectives* (pp. 239–251). Oxford: Symposium.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21–32.

Bruner, J. S. (1991). The narrative construction of reality. *Critical Inquiry., 18*(1), 1–21. https://doi.org/10.1086/448619.

Canagarajah, S. (1996). From critical research practice to critical research reporting. *TESOL Quarterly, 30*(2), 321–331. https://doi.org/10.2307/3588146.

Carless, D. (2007). Learning---oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66.

Chase, S. E. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In N. K. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 651–679). Thousand Oaks: Sage.

Children and Nature Network. (2012). *Children and nature worldwide: An exploration of children's experiences of the outdoors and nature with associated risks and benefits*. Retrieved from: https://www.childrenandnature.org/wp-content/uploads/2015/04/CECCNNWorldwideResearch.pdf

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. San Francisco: Jossey-Bass.

Creswell, J. W. (2008). Narrative research designs. In *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed., pp. 511–550). Upper River: Pearson Education Inc.

Forest School Canada. (2014). *Forest and nature schools in Canada: A head, heart, hands approach to outdoor learning*. Child and Nature Alliance of Canada. Retrieved from www.childnature.ca

Goldstein, L. S. (2000). Ethical dilemmas in designing collaborative research: Lessons learned the hard way. *Qualitative Studies in Education, 13*, 517–530.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140–145.

Hoare, T., Levy, C., & Robinson, M. P. (1993). Participatory action research in native communities: Cultural opportunities and legal implications. *The Canadian Journal of Native Studies, 13*(1), 43–68.

Lantolf, J. P., & Poehner, M. E. (2011). *Dynamic assessment in the foreign language classroom: A teacher's guide*. University Park: CALPER Publications.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. New York: Routledge.

Maple, M., & Edwards, H. (2009). Locating and understanding voices in narrative inquiry: A journey of discovery. In V. Minichiello & J. A. Kottler (Eds.), *Qualitative journeys: Student and mentor experiences with research* (pp. 33–48). Thousand Oaks: Sage.

Norton, B., & Early, M. (2011). Researcher identity, narrative inquiry, and language teaching research. *TESOL Quarterly, 45*(3), 415–439.

Play England. (2016). *Quality Play: Standards, assessment indicators, and evidence—a manual*. Retrieved from http://www.playengland.org.uk/wp-content/uploads/2015/11/QualityInPlay2017web.pdf

Polkinghorne, D. E. (1995). Narrative configuration in qualitative analysis. *Qualitative Studies in Education, 8*(2), 5–23.

Polkinghorne, D. E. (2007). Validity issues in narrative research. *Qualitative Inquiry, 13*(4), 471–486.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Los Angeles: SAGE Publications.

Rosiek, J., & Atkinson, B. (2007). The inevitability and importance of genres in narrative research on teaching practice. *Qualitative Inquiry, 13*(4), 499–521.

Sarkar, M., & Metallic, M. (2009). Indigenizing the structural syllabus: The challenge of revitalizing Mi'gmaq in Listuguj. *The Canadian Modern Language Review, 66*(1), 49–71.

Smith, L. T. (1999). *Decolonising methodologies: Research and indigenous peoples*. London: Zed Books.

Steinhauer, E. (2002). Thoughts on an indigenous research methodology. *Canadian Journal of Native Education, 26*(2), 69–81.

Swain, M., Kinnear, P., & Steinman, L. (2011). *Sociocultural theory in second language education: An introduction through narratives*. Bristol: Multilingual Matters.

Turner, C. E. (2013). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing*. London/New York: Routledge.

Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in the classroom. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment*. DeGruyter Mouton: Boston.

Winkler, G. (2003). Are we nice or are we real? Ethical issues emerging from collaborative narrative research. *Educational Action Research, 11*(3), 389–402.

Chapter 7
Learning from Each Other: School-University Collaborative Action Research as Praxis

Cheri Chan and Chris Davison

Abstract Hong Kong's education system has been undergoing major assessment reforms since 2000, exemplified by the introduction of a school-based assessment component into the secondary school English language curriculum in 2005–2007, and its extension to the final 3 years of secondary school in 2007–2010. The challenge many teachers faced at the time of the reform was understanding the new *assessment for learning* principles and practices that started to enter the English language teaching discourse. At the same time the researchers initiating the reform needed to theorise what it was that worked – and didn't – thus enabling continual refinement and improvement of the key assessment principles and protocols. Theorising school-university collaboration as praxis, this paper explores how participation in a collaborative action research project helped both teachers and researchers understand the new assessment discourses and practices. Textual data on teacher feedback were analysed to show how assessment theories were teased out, tested and taken up (or not) in practice. This study makes a contribution to the fields of assessment literacy and second language teacher development by highlighting some of the complexities the teachers and researchers experienced while engaging in Western theories of learning in a Confucian heritage culture. Findings showed how definitions of assessment, in this particular case, teacher feedback to learners, were socially organised and managed through socio-political and socio-cultural discourses and norms circulating in the Hong Kong context at the time of the project, and how this informed and shaped theory-building in ways which provided the assessment reform with long-term sustainability and legitimacy. Implications for school-university partnership as praxis for teacher development in the context of assessment reform will also be discussed.

C. Chan (✉)
University of Hong Kong, Hong Kong, China
e-mail: yycheri@hku.hk

C. Davison
University of New South Wales, Sydney, Australia

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_7

Introduction

This chapter explores the tensions and complexities of developing common understandings of assessment literacy, in particular feedback, through a collaborative action research project between researchers and teachers in Hong Kong. Hong Kong's education system has been undergoing major assessment reforms since 2000, exemplified by the introduction of a school-based assessment (SBA) component into the secondary school compulsory English language subject in 2005–2007, and its extension to the final 3 years of secondary school in 2007–2010. The challenge many teachers faced at the time of the reform was making sense of the new *assessment for learning* principles and practices that started to enter the English language education teaching discourse. The university researchers who initiated the SBA reform needed to theorize what it was that worked – and what didn't– thus enabling continual refinement and improvement of the key assessment principles and protocols. This chapter reports on the outcomes of a collaborative action research (CAR) project that was initiated at the time of the language assessment reform to build junior secondary (Grade 7 & 8) English language teachers' capacity to implement the principles of assessment for learning in their classrooms. To theorize how researchers and teachers negotiated understandings of assessment practices during the collaboration process, we examined the data through the lens of praxis. Praxis, as explained by Poehner and Inbar-Lourie (Chap. 1, this volume), considers theory as the orienting basis to practice and practice as a testing ground to determine the viability and usefulness of theory. Important from the perspective of praxis is that there is not a simple, unidirectional flow from theory to practice, but rather knowledge results from the activity of theoretically driven practice that informs theory.

The CAR model also drew on Vygotskian Sociocultural Theory, which itself emerges from and embraces praxis (Lantolf and Poehner 2014). Moreover, as Michell and Davison (Chap. 2, this volume) illustrate, Vygotskian theory brings a crucial developmental aspect to praxis, calling attention to the ways in which activity is both mediated by and allows for the creation of tools. Together, a praxis orientation and Vygotskian theory guided efforts to create formalized opportunities for dialogical learning about assessment, interwoven with systematic opportunities for teachers and researchers to try out ideas in the classroom and share their evolving understandings as well as unanticipated problems. Textual data were collected from one of the five major foci of the CAR study, a research group focusing on teacher feedback to learners. We analyzed the data (semi-structured interviews, CAR project meeting dialogues and field notes) using discourse analysis to identify how new assessment ideas and theories were teased out, tested and taken up (or not) in practice. We also explain how assessment principles and practices were socially organized and managed through socio-political and socio-cultural discourses and norms circulating in the Hong Kong context at the time of the study, and how this informed and shaped theory-building in ways which provided the assessment reform with long-term sustainability and legitimacy. The implications for school-university

collaborative learning as praxis for the professional development of teachers in the context of assessment reform will be discussed in the final section of the chapter. This study makes a contribution to the fields of assessment literacy and second language teacher development by highlighting some of the complexities the teachers and researchers experienced while engaging in Western theories of learning in a Confucian heritage culture.

Changing Conceptions of Assessment: The Case of Feedback

The concept of assessment *for* learning, widely promoted through the work of the Assessment Reform Group (1999), was introduced to ensure "a clear distinction be made between *assessment of learning* for the purposes of grading and reporting, which has its own well-established procedures, and *assessment for learning* which calls for different priorities, new procedures and a new commitment" (Assessment Reform Group 1999, p. 2). The defining characteristics of A*f*L include assessment strategies and tasks being embedded in learning and teaching, learning goals being explicitly shared with students, students being supported to know and to recognize the standards they are aiming for; students being involved in peer and self-assessment, feedback which leads to students recognizing their next steps and how to take them, multiple structured opportunities for both teacher and students to review and reflect on assessment data, and all underpinned by confidence that every student can improve (adapted from the Assessment Reform Group 1999, p. 7).

In the assessment for learning literature (Black and Wiliam 2009, 2010; Black et al. 2003) there is a strong consensus that ensuring students have the opportunity to access quality feedback following their participation in assessment tasks is the crucial link between assessment and learning (see also discussion by Leung, this volume, Chap. 5). Black et al. (2003) argue that formative feedback is essential in order to provide the learner with information about "their current achievement and to indicate what the next steps in their learning trajectory should be" (p. 42). Findings from their research strongly suggest that formative feedback should focus on what needs to be done because this encourages all students to believe that they can improve:

> Such feedback can enhance learning, both directly and indirectly by supporting the motivation to invest such effort. A culture of success should be promoted where every student can make achievements by building on their previous performance, rather than being compared with others. Such a culture is promoted by informing students about the strengths and weaknesses demonstrated in their work and by giving feedback about what their next steps should be (Black et al. 2003, p. 46).

Formative feedback not only provides students with information about their strengths, areas to work on and improvement strategies, but it also feeds into the teacher's ongoing planning. As teachers become more informed about the strengths and weaknesses of their students through interactive dialogue with their students, they are better placed to plan tasks and activities that address more closely the needs

of those students and to guide and facilitate their learning. Black and Wiliam (1998) argue that feedback can only be described as formative when it is used to change learning and teaching.

Early research into assessment for learning practices (Clarke 1998; Sutton 1998) found that to impact on learning, formative feedback needs to:

- be based on explicitly articulated learning goals which are fully understood by students and teachers prior to embarking on the task
- encourage and take account of student reflection and self-evaluation to maximize the potential for feedback being understood and implemented
- focus on evidence of success and where improvement could take place
- be delivered in a way which students can understand
- be specific and task focused
- be descriptive and questioning, rather than evaluative
- elicit or offer alternative ways of doing things
- be future orientated, looking forward to the specific next steps to improve performance rather than emphasising current mistakes
- provide concrete strategies to help students to improve
- be offered during and / or as soon as possible after the task
- be put into practice as soon as possible.

Black et al.'s (2003) classroom-based research reinforced the need for students to have a clear understanding of learning goals: "The criteria for evaluating any learning achievements must be made transparent to students to enable them to have a clear overview both of the aims of their work and of what it means to complete it successfully" (p. 52). If criteria are abstract or difficult to interpret, modeling needs to be provided to develop understanding. Later research supports this view, suggesting that a combination of goal setting and feedback effectively and significantly increases individual performance (Hattie and Timperly 2007; Van den Bergh et al. 2014). Sharing criteria with students also provides a framework for genuinely dialogic feedback (Harris et al. 2015), with student strengths, areas to work on and improvement strategies recorded and tracked by students so that they can see evidence of progress.

Black and Wiliam (1998) also highlight the significant impact assessment has on student motivation and self-esteem. Research (e.g., Dweck 2007) shows that students explain success and failure according to internal or external and stable or unstable attributes: "I spent a lot of time revising/I didn't spend a lot of time revising" would be internal attributions for success or failure. "The test was easy/difficult" would be external attributions for success or failure. "I'm clever, so I passed the test/I'm not clever so I didn't pass the test" are stable attributions for success or failure since cleverness or lack of it is perceived as an inherent quality and therefore stable while "I worked hard/I didn't work hard" are unstable attributions for success or failure since the amount of effort a student assigns to tasks can vary according to circumstances. Black et al. (2003) argue that what is crucial to effective learning is that students attribute *both* success and failure to internal, unstable causes. In other

words, they need to take responsibility for their learning (internal attributions) and see it as dynamic and changing due to factors such as their own effort and ability.

This research has important implications for the provision of feedback. Through their formative feedback, teachers should promote the view that all students, irrespective of their achievement, have the capacity to improve, that their 'ability' is not 'fixed' and immutable. According to Black et al. (2003) students who see ability as being fixed "see every task as a potential threat to their self-esteem and their goal becomes to preserve this self-esteem …students who are motivated in this way… have a 'performance orientation' to their work - the goal in every lesson is not to learn, but to perform well to maintain self-esteem" (pp. 75–76). They go on to explain that if such a student feels unable to do well on a task then that student is likely to disengage to protect their self-esteem. Conversely, students who see ability as being dynamic or incremental are more likely to be 'goal' or 'learning' orientated. Hence, feedback should not focus on personal praise, evaluation, judgment or grades as this just feeds into ego and leads to comparison with others and the preservation of self-esteem. Rather the focus should be on the task and improvement strategies to develop a 'learning goal orientation' so that every student develops a dynamic view of ability and a belief that they can progress.

The way in which teachers perceive assessment will affect the way in which feedback is provided. Torrance and Pryor (1998) described two approaches to classroom assessment which they called 'convergent assessment' and 'divergent assessment'. "Convergent assessment aims to discover *whether* the learner knows, understands or can do a pre-determined thing" (p. 153) whereas "divergent assessment aims to discover *what* the learner knows, understands or can do" (p. 153). Convergent assessment reflects a behaviorist view of learning and is characterized by, among other things, "closed questioning and tasks, judgmental or evaluative evaluation, involvement of the student only as recipient of the assessment and a view of assessment as accomplished by the teacher" (p. 153). Divergent assessment reflects a more constructivist view of learning and is characterized by "open questioning and tasks, descriptive rather than purely judgmental evaluation, involvement of the students as initiator of assessments as well as recipient and a view of assessment as accomplished jointly by the teacher and the pupil" (p. 153). Convergent assessment practices are closer to those of summative assessment while divergent assessment reflects practices which "attend more closely to contemporary theories of learning and accept the complexity of formative assessment" (p. 153). Formative feedback strategies therefore need to reflect a 'divergent' view of assessment in which students are encouraged to participate in feedback discussions through open, reflective questioning and description rather than evaluation of events in assessment tasks. Black and Wiliam (1998) supports the view that any dialogue between students and teachers should be an opportunity for thoughtful reflection in which all students are able to participate. The provision of grades or marks should be postponed as long as possible as it reinforces traditional hierarchical teacher–student

power relationships and closes down discussion, rather than opening it up (Wiliam 1999).[1]

The way in which teachers structure feedback episodes, therefore, has a powerful effect on students. Tunstall and Gipps (1996) developed a typology to describe teacher feedback in British schools. Their typology categorized feedback as either 'evaluative' and 'self' orientated, involving comments construed as rewarding (e.g., general praise) or punishing (e.g., negative generalizations), or as 'descriptive' and task orientated, in which case achievement and improvement feedback were emphasised. They noted that most of the feedback they had observed when conducting their research was evaluative in nature. Evaluative feedback, such as praise, focuses students on themselves and therefore promotes attention to 'ego' and self-esteem and leads to a performance orientation rather than a learning goal orientation, with more negative than positive effects on students (Hattie and Timperly 2007).

Stimpson et al. (2000) described the language of feedback as falling into four main categories: descriptive, questioning/reflective, advisory and evaluative. 'Questioning' involves asking students open ended questions about their understanding of the learning goals, their strengths and areas they need to work on. It also prompts students to consider alternative ways of doing things and steps for working on these areas. 'Describing' involves reviewing task events in a non-judgmental way to prompt students to elaborate further on their performance. 'Advising' involves offering suggestions to help students improve while 'evaluating' refers to judgmental assessment of the strengths and weaknesses of the performance. In light of the arguments above, 'questioning' and 'describing' strategies, which encourage student participation and reflection in feedback discussions and direct students towards focusing on aspects of the learning criteria and task completion, are more likely to contribute to effective feedback than 'evaluative' strategies, particularly those which focus on personal praise or criticism. From this perspective, advising is a strategy that is better employed after first attempting to elicit deeper reflection and elaboration from the students themselves to see if they can suggest alternatives, in this way empowering students to self-correct. Encouraging students to reflect, offer alternative strategies and self-correct creates a much deeper learning experience for the students, as Black et al. (2003) suggest "when the teacher asks 'higher order questions – questions that explore understanding and require thinking – the student is not just recalling knowledge but building it" (p. 60). In reflecting on their work in relation to learning goals students begin to think more deeply and concretely about their learning. Reflective feedback moves students beyond just being able to say *what* was good to being able to say *why* (Black et al. 2003). Reflective strategies for co-constructing feedback creates a deeper learning experience for the teacher too because, as they encourage students to be more explicit about their understandings

[1] Wiliam (1999) found that that assigning grades or marks was not conducive to learning, tended to undermine the motivation of weaker students and encouraged students to become more concerned with model answers, finding the right answers and/or trying to guess what the teacher wanted rather than focusing on the learning process, their own ideas and how they can progress towards learning goals.

and misunderstandings, teachers can fine-tune their feedback and planning to meet the real needs of the students as these are exposed through feedback discussions. Furthermore, Black et al. (2003) suggest that by encouraging students to reflect, by modeling ways of questioning and describing students' work, teachers 'apprentice' students into the process of assessing themselves and their peers. This skill constitutes one of the main aims of assessment for learning. Teachers in Black et al's research reported that when giving feedback to each other "students began to use the language they had heard the teacher using" and that "when teachers successfully developed effective feedback strategies with their students, self and peer assessment was enhanced" (Black et al. 2003, p. 67).

However, despite its pivotal role in the learning process, the nature of teacher feedback in classroom-based assessment of oral English as a second or additional language is under-theorized compared with studies of feedback in second language writing assessment (Hyland and Hyland 2001) and teacher-student interaction in classrooms more generally (Richards 2006). Even fewer studies deal with feedback on oral interaction in non-English speaking Western contexts. Hence, not surprisingly a number of crucial factors – socio-cultural, technical and practical - are never discussed in the literature, let alone built into models of feedback. For example, this very Western-oriented model of feedback seems to assume an individualistic, questioning, critical learner who is comfortable talking about their own strengths and weakness with a teacher as an equal (Carless 2011). It also appears to assume grades and marks can be jettisoned or "postponed" without any objections from students, school leaders or even parents (Hamp-Lyons 2007). From a technical point of view, this model of feedback naturalizes the language of feedback, with no attention to linguistic structures and features beyond discussion of the different function of feedback and the warning to favor substantive comments over practice. More significantly, little or no research addresses the issue of how to give effective formative feedback *about English* when the student is still learning *in and through English* (Hamp-Lyons and Tavares 2011). Finally, from a practical point of view this model of feedback appears to assume relatively small class sizes, a supportive school administration, a manageable teacher workload, and no competing priorities for teacher time (Davison 2007).

This chapter seeks to begin to fill this gap and address these issues by reporting on the construction of teacher feedback in a large–scale action research study in Hong Kong which sought to help junior secondary English teachers to improve their strategies for providing feedback in oral assessment tasks as part of an initiative to develop formative feedback practices. Drawing on pre and post-feedback interview data, teacher reflection, and the observation and analysis of the actual feedback episodes between teachers and students following assessed group interaction tasks, this chapter explores how Hong Kong English teachers interpreted 'feedback' and the contextual factors which inhibited or enhanced the provision of quality feedback.

The Hong Kong Collaborative Action Research Assessment Initiative

A school-based oral assessment component was introduced into the high stakes O-Level Hong Kong Certificate of Education Examination in English Language in 2005–2007, then extended to the final 3 year of secondary school from 2007 to 2010 onwards (Davison and Hamp-Lyons 2010).[2] The SBA (English Language) is based on assessment for learning principles (Black et al. 2003), and consists of an assessment of students' oral proficiency based on speaking assessment tasks developed from an independent extensive reading/viewing program over the course. Teachers are expected to design and implement appropriate assessment tasks, involve learners more actively in the assessment process, apply the standardized assessment criteria to making trustworthy assessment decisions and provide effective feedback and feed-forward to students to improve student learning. All these roles were new to teachers in Hong Kong as previously all oral language assessments had been externally set and assessed by the Hong Kong examination authority. More significantly, the SBA was initiated in an educational context very different to Western-oriented educational systems, so much so that when the SBA was first introduced, it was widely assumed it would not work. This pessimism was not just because of technical challenges (such as how to design assessment tasks that could be tailored to the needs of individual students and yet would still be comparable across schools) and practical concerns (heavy workloads and large class sizes of 40 or more students). There were also the socio-cultural challenges - it was assumed Hong Kong teachers could not understand and/or would not accept the "Western pedagogy" and the underlying philosophy of assessment for learning which assumed major changes in student and teacher roles for a whole variety of cultural and historical reasons (for a fuller discussion, see Cheng et al. 2010; Davison 2007; Davison and Hamp-Lyons 2010; Davison and Leung 2009; Hamp-Lyons 2016; Lee 2011; Qian 2010).

Since the SBA initiative marked such a significant shift in policy as well as practice for the Hong Kong educational community, the provision of professional development to junior secondary English teachers was supported by educational authorities as an important component in the implementation process. A large-scale capacity building collaborative action research (CAR) project was funded in 2006–2008 (SBA Consultancy Team 2008) which aimed to build English language teacher confidence in and knowledge of how to implement SBA for learning through integrating theory and practice. The project followed Burn's (1999) principles for collaborative action research for second language teacher education. The CAR project was presented to the teachers as a practice of professional development. It was also made explicit to the teachers that they would be co-researchers in the project. For example, university researchers and school teachers worked closely together

[2] For more information about the HKCEE school-based assessment system, see http://www.hkeaa.edu.hk/DocLibrary/SBA/CE-Eng-07IntroBooklet-0610.pdf, and for the HKDSE, see http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/ENG-Intro_to_SBA_Booklet-Mar10.pdf

throughout the project and followed a process of co-inquiry, with university researchers acting as facilitators of this process. Their key role was to provide support and guidance to the teachers so that they could implement their own school-based action research projects around the common goals of improving assessment practices. For example, the facilitators gave input and feedback to the teachers on the construction of their action research plans. They also visited the teachers at their schools during the action research cycles. In these face-to-face meetings, teachers and facilitators exchanged professional information and 'troubleshooted' any problems arising from the research. These regular conversations and meetings with teachers during the CAR project provided the university researchers with valuable insights into how teachers and schools were making sense of the SBA initiatives being implemented at the time.

Through this practical professional development initiative, English language students in the participating project schools would be exposed to many more speaking activities to provide them with the language and skills needed to interact in English with confidence at higher levels of schooling. At the same time, they would be apprenticed into the principles and practices of school-based assessment for learning beginning in Form 1 (Grade 7) through assessment of their participation in group interaction and individual presentation tasks. In place of traditional, tightly structured one-off annual oral exams, students were to be given several opportunities within their normal English classes to participate in more formal speaking assessment tasks, each opportunity allowing for recurrent and formative feedback to take place, thus encouraging assessment *for* learning practices.

When the CAR project was launched, small teams of teachers from 24 Hong Kong secondary schools volunteered to participate in the project. Following a praxis orientation, the researchers sought to avoid imposing a one-size-fits-all approach to professional development, opting instead to offer a set of options reflecting the differing needs and concerns schools might have about implementing SBA (see chapter by Harding and Brunfaut, Chap. 4, this volume, for a similar discussion of negotiating professional development opportunities with teachers). There were five key topics for schools and teachers to explore: interactive assessment, grouping, self and peer-assessment, task design and feedback. The topic of *feedback* was the most popular, with seven schools signing up to explore this issue. The high response was not a surprise as the project had sparked great teacher interest in formative feedback, in particular, the nature of constructive, dialogic quality feedback and how to use it in ways which would enable students to do better in subsequent assessment tasks. However, with the majority of teachers still working in classrooms orientated towards summative assessment, grading and comparing students, attempting to implement assessment *for* learning practices was inevitably going to be challenging. At the same time, the project had recruited a number of new researchers who had not previously been involved with the SBA initiatives. Thus, the challenge for both sides was to make sense of the new assessment for learning principles and practices that had already started to enter the English language education teaching discourse, and to find out what worked, and what didn't, to ensure continual refinement and improvement of the key assessment principles and protocols.

The collaborative action research model drew on Vygotsky's notion of praxis where understanding is conceptualized as dialectical in nature, combining consciousness (knowledge and theory) with action that results in the creation of an object (Lantolf and Poehner 2014). Hence, an iterative process was adopted in the project with more formalized opportunities for dialogical learning about assessment interwoven with systematic opportunities to try out ideas and approaches in the classroom and share the evolving understandings, resulting in teaching ideas and strategies as well as unanticipated problems emerging as part of efforts to implement theoretical principles and then adapt them in practice. For reasons of space, this chapter focuses only on the experiences of the feedback CAR group, led by two university researchers, and described in more detail in the next section.

The Study

Seven secondary schools and 20 teachers participated in the teacher feedback group. To enhance collaborative learning and co-inquiry, teachers were encouraged to participate in the project with one or two colleagues. Although data were collected from all teachers in the project in relation to their beliefs about feedback and feedback events, for reasons of space this chapter focuses on the evolving understandings and practices of ten teachers in three case study schools. All the participants were English language teachers, with 3–15 years of classroom teaching experience. One participant was a native English language teacher from the UK and the other seven were Hong Kong-Chinese. Some of the teacher volunteered to participate in the CAR project, others said they were nominated by the school. All three case study schools were government funded public schools. In each school, teacher participants were invited to discuss and formulate a two-cycle action research plan around a particular issue or challenge they wanted to explore as a team in relation to teacher feedback for oral tasks. The three schools shared a common interest in wanting to investigate various ways of involving students more in feedback discussions, presented through a series of university forums in the early stages of the project (Davison et al. 2009). Teachers in these three schools were also interested to see whether the kind of language offered by teachers in their feedback following oral work was reflected in the way in which students offered feedback to each other. In addition to exploring the ways in which the language of feedback could enhance students' participation in feedback events, the teachers also wanted to explore other strategies for making feedback events more memorable and to optimize the potential for students to implement advice and targets set.

Two university researchers enacted the role of action researcher facilitators by helping the teachers in each school identify an action research focus, formulate a research plan and suggest methods for data collection and analysis. The Feedback group's lead facilitator was Anna (pseudonym), who worked as a teacher educator at the university at the time of the study; the first author of this chapter, Cheri, was a co-facilitator. Anna and Cheri enacted the roles of 'critical friends' and input

providers. For example, they facilitated a series of workshops about feedback to help teachers plan the action research project and then visited the teachers at each school before, during and at the end of each action research cycle to help them analyse the data, reflect and review the interventions implemented. These 'professional conversations' during the school-university collaboration were then recorded and transcribed for textual analysis to examine the process and outcomes of collaboration for professional learning. Anna and Cheri were also supported by the CAR project's core research team, e.g. the principal investigator (second author) and Katy (pseudonym), the project manager. The core research team also assisted the teachers with the actual data collection process, for example the video recording of feedback sessions during lessons and the transcription of feedback discourse.

Drawing on the principles of discourse analysis, the authors analysed the 'everyday talk' collected during the collaborative activities to examine how key assessment concepts and principles were presented as a body of knowledge in the wider Hong Kong assessment reform discourse, and how they were then (re)negotiated and made sense of by the researchers and teachers in their particular institutional context. Researcher field notes and the recordings of the meetings/conversations and interviews with teachers were transcribed, read twice and analyzed for themes related to the two key research questions: (1) How were the new assessment practices tested and taken up (or not) in practice by the teachers and researchers? (2) How did the teachers help the researchers make sense of assessment for learning principles and theories? Textual properties of vocabulary, grammar and textual structure were analysed to identify how beliefs and practices about teacher feedback were instantiated in the collaborative talk by the teachers and researchers, enabling the problematization of complex relations between language use and societal practices (Phillips and Jorgensen 2002). The following questions were used to guide the analysis of the textual data discussed in this chapter:

- What word choices and key terms (e.g. attitudinal words and ideational metaphors) were given prominence/suppressed/backgrounded in the texts to represent the teachers'/researchers' understanding of feedback?
- What themes emerged to represent the teachers'/researchers' understanding of feedback?
- How did this compare with the particular constructions of feedback represented in the school-university collaborative action research project?

Developing New Assessment Constructs: The Case of Feedback

Building on Chan (2015, 2016), the findings presented in this chapter show how the researchers and ten English Language teachers from three secondary schools (see Appendix A) made sense of the assessment for learning principles and theories, discourses and practices during the school-university collaboration process, and

perhaps even more significantly, how the researchers were forced to engage with a number of new issues to build a more context-sensitive model of feedback for Hong Kong secondary schools, one which took into account their specific socio-cultural, technical and practical issues. In the data excerpts that follow, the schools are referred to as School 1, 2, and 4. All teacher names are pseudonyms.

How Were the New Assessment Practices Tested and Taken Up (Or Not) in Practice by the Teachers?

A recurring tension which emerged in the data was how the researchers and teachers dealt with the challenges of applying the Western assessment for learning principles in non-Western school contexts. For example, in the initial stages of the CAR project, it was found that the teachers and researchers had different understandings of feedback:

> Anna (researcher): OK …why are you interested in impact of feedback?
> Jennifer (teacher): We thought it's very important for the students to be aware of their weakness and strength in the group interaction. Because junior forms it is *basic skills for them to have training*, and after they go to the senior forms or *have the SBA training*, they will be more aware of themselves. So feedback is very important for both junior forms and senior forms. We want to have the pilot study first, so after we know about the curriculum and we can decide for the whole form, whole junior form ...
> (Extract 1: School 2)

Thus, in the initial stages of the reform it was found, on the one hand, that the teachers believed that feedback was important, but on the other hand, their understanding of feedback also reflected more traditional and grade-oriented assessment practices:

> Anna: Do you give them feedback so that they can improve for the next assessment?
> Mary: I (do) but some don't want it. Quite a lot on book reports. Comments in compositions are usual, special marks for compositions.
> Anna: Do you give them a grade, A, B, C, D or marks?
> Mary: I am doing totally grades … (but) the calculation will be twice a year. Like for example ERS (extended reading scheme), if they read 7 books they share, we pick up the top 3. It's all very fair. You say to students, "You do more work. You actually do better."
> (Extract 2: School 4)

However, the reasons the teachers gave to explain their traditional feedback practices were not cultural, but very practical, such as the lack of time and opportunity:

> Anna: What stops you from giving (more qualitative) feedback to students at the moment?
> Patsy: We don't have enough time to talk with individual students.
> Mary: Time.
> Patsy: Some of them (teachers) can't afford to give so much time for feedback.
> Mary: No opportunity. I have done it in lunch time with my Form 4 in chat room (giving oral feedback). And I said to them it was compulsory, they had to come to me for *training*. I told them please come to me for *training*. It's a valuable workshop. It actually means discussing, a discussion group. That really really worked well, making it compulsory. (seeing the teacher for feedback session).

Mary: We really cannot do it every lunch time because we give up lunch. I just try to do it once a week.
(Extract 3: School 4)

Another overwhelming problem was seen as the language of feedback:

Mary: Students don't understand what I say. They look very blank or they say something in Chinese. They ask their friends.
Patsy: Language is a problem. The duration of feedback is maybe another problem. I think students' attention span (listening in English) is quite short. I'd better keep it short (feedback).
(Extract 4: School 4)

Kerry: and sometimes, for example they are doing oral activities, individual presentations or group discussions, usually we will just give verbal feedback and we are not sure, you know, whether they understand all of the things that we have mentioned to them and then sometimes if we just, you know, maybe mark it down and then…well, it's really difficult because time is limited and there's only one teacher for 40 something students and all of them say "I don't understand, can you come to me?"
(Extract 5: School 1)

This led to teachers expressing concerns as to how to structure productive feedback:

Mary: I go to the library when they are doing their reading and they read individually, and I can give them some feedback during their reading. I sometimes find what I can say is *weak*, I just say, "Oh, that's good. Your reading is improving." But *I am not quite sure what to say.*
(Extract 6: School 4)

In School 1, teachers commented on the same kind of problem:

Kerry: well, in fact, usually in the past, we tended to give limited types of feedback, I mean, usually we would try to focus on identification of strengths and weaknesses and maybe and give them comments, but *there are something that we have ignored,* that is, for example, giving them chances to reflect on their own performances by asking them more reflective questions, so then we would really know whether students also think well, just like what we do.
(Extract 7: School 1)

As a result, teachers appeared more concerned about technical issues, such as how to structure effective feedback opportunities, what language to use and how to support students to interact and engage with the feedback, especially given their previous experiences with giving feedback on writing:

Fiona: well, in fact we try to design a feedback sheet in the writing task starting this year, so when I started using this method, I wrote many comments, but ah, next time they made the same mistakes again and so, every time when I give them back their feedback sheet, I'll just ask them to spend 5 minutes reading it together in class and ask anything they don't understand about my marking or my comments and I think it works.
Linda: for me, because, I don't give them 5 minutes because I teach the remedial class in form 2 and when I write the feedback for them, some of the students, they don't understand what I was writing, I need to translate for them and so…because I just only have 20 students in my class and I have more time to take care of them, maybe I will try to ask them one by one come out and then to show them the mistakes and then let them be aware, and then try to explain to them. Because if I just ask them to read by themselves, they don't understand … usually they won't read it, so that's my way.
(Extract 8: School 1)

However, with the input and support of the researchers, and the structured opportunities to test out new different strategies and techniques and learn from each other over the course of the project, teachers gained the knowledge and confidence to change their feedback practices, as described by teachers in School 1:

> Kerry: yeah, …the first cycle we just eh, describe what their strengths and weaknesses are, and the we give feedback, and then we give advice and after that we asked the students to write down, you know, what they have done well and what they have done badly and we try to see if they remember it or if they understand what the teacher, you know, have just mentioned and in the second cycle, we do the same, but differently in the feedback discussion. We deliberately focused on asking questions, you know, just get the students to say what they have done well or what they need to improve themselves and then just like what we did in the first cycle, we get them fill in the questionnaire about their own strengths and weaknesses and see how well they remember it, …and after each cycle, we teachers will sit together afterwards and then talk about it, how do we feel about it. I think there are two things that we've found out. First is that we thought before students (would) actually do better in the first cycle because we say it explicitly what they did well, and then we give suggestions, but what we found out you know, is not, they didn't remember much about, you know, what we have said, maybe because we (the teachers) have said a lot of things…and so in fact my students, I mean, my group actually did better in the second cycle when I got them to say it themselves.
>
> (Extract 9: School 1)

The teachers then extrapolated from this experience, proposing that students video-record and review their oral interaction so that they could take a more dominant role in leading the feedback event, commenting:

> Kerry: If we let them, you know, view their performance before we do the feedback section, they will remember it better. *It's very obvious*… and it's easier for us when we're giving feedback, you know, we just ask questions and then we just follow up our questions, we don't need to organize everything in a hurry or in a rush and then you know, say everything in an organized way … (and) they could actually, you know, say something about their friends.
>
> (Extract 10: School 1)

After trying this out, the teachers in School 1 were very positive:

> Wendy: there were four girls in my group and I think they were eager to you know, to try out to show off like how they improved in the second cycle, so, yeah, while they were watching the video, they really, noticed the details, the themes, the gestures like that, so I think the students, they improved a lot in the second cycle.
>
> (Extract 11: School 1)

As the teacher resolved some of the technical issues of how to engage students more actively in feedback interactions, some of the concerns about practical issues disappeared as the teachers realized the impact on learning of giving more time to quality feedback, although they still found some aspects of language (e.g., speed) easier to deal with than others (e.g., intonation):

> Kerry: well, we have one lesson for oral but I always spend two lessons, because one lesson is too short, you know, you can't explain what they should do, so I usually spend two lessons …(I) now realize it's important to let students engage in feedback e.g. watch the video and then discuss their performance … for example, if you say "you have to have more eye contact" or "you have to speak louder" it's just easier because you have the video and they can see how they did it, but sometimes when, for example, intonation

stress or content, you know, it is really impossible for teachers to really listen to them all and then you know, get them for example complete the sentence or phrase, you know, really give them some practice, extra practice, or give them some concrete suggestions and how to improve, this is difficult.

Linda: and they also, they can find what are the weakness of themselves and they can also find some strengths from watching others and then they can learn from them… not only feedback by teachers, but also feedback by students …peer assessment

(Extract 12: School 1)

The teachers in School 1 also discussed how to overcome some of the technical and practical issues of structuring the feedback:

Fiona: I think students have to be very sure of what the criteria they…will be assessed on, so that they know how they can improve or yes, what criteria they have to focus on, so they can prepare better for their SBA, this kind of video-taping can actually prepare them not to be so nervous, but time is really a problem, I think it is not very necessary to video tape the whole class for junior form, so I think videotaping needs extra time. Actually, but I think well, we can find some.

Kerry: maybe one or two groups

(Extract 13: School 1)

The final comments from the teachers suggest that as the specific technical and practical problems of structuring quality feedback in Hong Kong schools were collaboratively resolved, their cultural assumptions and attitudes about assessment for learning and its purposes also started to change, with many commenting explicitly that not only did they find feedback helped their students learn better, but that is also changed their teaching and, perhaps even more significantly, the nature of their relationships with students and the general classroom atmosphere and dynamics:

Patsy: I find …feedback from the students quite useful and when I see their feedback, I know how they feel about the lesson. My students are really young, Form 1 students, but … the AR project gave us more opportunities to talk about the lessons and they can give some feedback to me so that they are not afraid to express their ideas.

Mary: I think that it was fantastic, to actually talk to the students on a more one to one level, often the students are en masse and as a teacher, you don't get much individual time with your students, maybe apart from after school when they come to you. This … was actually sitting down and working with them individually so I found that really fantastic. I got to know them better, got to know their personalities better. In terms of what I actually did, the feedback, I felt I could really give them much better feedback (verbal) and I felt they really understood me. Whereas if they just read my comments (written feedback) just in writing maybe they would not have understood exactly what I meant, but because I spoke to them so they could question me and I could elaborate and it was much more successful … I think it also helps the weaker learners as well. The weaker learners and the group were helped by the stronger learners so maybe I didn't really understand what I'll say but they were able to translate and help the weaker learners, so it's like the pair helping session.

(Extract 14: School 4)

This suggests that English language teachers in the three Hong Kong secondary schools were able to adapt and modify their approach to feedback to more closely resemble assessment for learning, at the same time continuing to meet the expectations of their own schools and students. The teachers' problems were construed as more technical and practical than socio-cultural, with even young students apparently willing to take on the role of a questioning, critical learner who is comfortable

talking about their own strengths and weakness with a teacher as an equal, provided they were given appropriate scaffolding and support. Whether such practices can survive a longer-term implementation is more problematic, in that this was just an experiment over a school year with a small group of students, explicitly conceptualized by the teachers as a low stakes pilot study, so there were few objections from students, school leaders or even parents about the lack of traditional marking or grading. However, it does paint a more optimistic picture of teacher attitudes and practices and their willingness and capacity to change than that portrayed in some of the literature (cf. Lee 2008; Qian 2010; Carless 2011).

The next section explores how, through the collaborative dialectal nature of this project, the teachers' practical experiences and ideas helped the researchers make sense of the new assessment for learning principles and theories, particularly in relation to enhancing existing Western-oriented models of feedback.

How the Teachers Helped the Researchers Make Sense of Assessment for Learning Principles and Theories

The two researcher-facilitators in this collaborative project wanted to see to what extent assessment for learning principles could be applied in Hong Kong secondary schools, in an English as a second language teaching context in which English was both the object and the means of instruction. The researchers wanted to help teachers move from the traditional notion of feedback as grades or marks associated with high stakes summative tasks to formative assessment-while-learning, with much more dialogic and qualitative feedback opportunities involving self – reflection and, ideally, peer evaluation. This meant a shift in conceptualization from feed*back* to feed*forward* (Hattie and Timperly 2007), from teacher feedback to giving feedback to help students learn, from one-way hierarchical language episodes to co-constructed dialogic conversations. At the same time the researchers had to reinforce and respect the teachers as the experts in what might or might not work in their context. Hence, just as the teachers felt rather uncertain initially about what they were doing, so did the researchers:

> Anna: I have to say I was quite worried …I was concerned about the comments I gave them, even though it was *would you consider doing this or doing that*, I felt what they needed was someone to tell them these are all the things you are interested in looking at, but it's too many. So I didn't actually say you should do this or you should do that, I was just trying to untangle things for them.
> (Interview with CAR Facilitator)

After the first round of interviews with the seven schools in the feedback group, Anna raised concerns about the facilitation process:

> She felt uncomfortable because we were not giving teachers enough/adequate support in schools. We feel that we are not in control of the research process. I sense that she feels that 'AR' does not allow her to 'intervene' directly, but Anna feels intervention is not what she wants either, so we are struggling because she can see that teachers are not doing what is

> intended in their research plan. In the feedback group we are not seeing good examples of feedback in assessment for learning. (CAR facilitator, Author 1, Field notes)

The project manager assumed that this was because of a lack of knowledge and confidence in doing research:

> I would say that quite a lot of teachers have some ideas of what action research is, but it seems to me that some of them feel quite uneasy about the idea, they would say I don't know how to do research, you can come to do research on me, you come to the school and look at what we do and tell us how to improve, they resisted doing research themselves. I guess it is very important that the teachers do it themselves, so I would say we have to make it [AR] nice and easy for them, first of all to help them understand why they have to do action research in a very simple way and then be with them during this whole process. (Katy, project manager, pre-project interview)

Unlike the teachers, the researchers saw the introduction of new models of feedback as not just a technical problem, but socio-cultural, with the teachers demonstrating traditional assumptions about the nature and purpose of assessment (and research & teaching). For the researchers, change was not just a matter of solving technical or practical problems but changing teacher 'mindsets'. Early in the project Cheri commented:

> The way they (the teachers) gave feedback was mechanical and contrived in the sense that I felt that they had videoed the feedback for our benefit. Why did this happen? Too much guidance or too little guidance? …. We can see that teachers are still adopting a very traditional approach to assessment. We were a bit depressed after watching the video data, because we saw in reality, teachers were very traditional, exam-oriented and focused on training the students. Just because they are participating in this CAR project, it doesn't necessarily imply they are ready to adopt innovative assessment practices! Perhaps that's the point, that this is PD and they want to explore new ways of assessing students, but they know there are challenges (skills, pedagogical and assessment knowledge). For example, we identified teachers are not just having problems in the way they are giving feedback, but there are some problems in designing assessment tasks too. (CAR facilitator, Author 1, Field notes)

Even in the mid-stages of the project, these deep-seated concerns remained: "The video of the research cycle given to us by the teachers was very weak in the sense that they did not approach the assessment as assessment *for* learning, but did it as a task like an exam practice. Anna has been very concerned about how to guide the teachers, for example, not using timers to test students, but creating an environment so they want to use English (CAR facilitator, Author 1 Field notes).

However, through engagement in praxis, including the many conversations with the teachers, the researchers came to realize that it was actually the practical and technical issues which were getting in the way of changing the teachers' thinking about feedback, not entrenched assumptions about the nature and purpose of assessment. Developing a new model of feedback from practice up, not just theory down, the researchers incorporated much greater attention to the structure and language of feedback, developing a range of video exemplars and activities to help teachers and students understand what the new feedback practices were actually supposed to sound and feel like. They also systematically brainstormed how to address the various other technical issues raised by teachers, including how to ensure teachers were

as well supported as possible by their school leadership. The final evaluations showed that the teachers felt they gained from participating in the project both in terms of understanding of assessment for learning principles and the teaching and learning of English.

To sum up, the CAR project and the support of the facilitators offered the teachers a space to explore new feedback strategies. In this way, school-university collaboration was discursively constructed by the teachers to mean gaining practical support, professional knowledge and expertise from the university researchers to make sense of the assessment reform. The university facilitators helped the teachers make sense of the new assessment discourse by providing intellectual resources to help the teachers achieve their own school's teaching and learning agenda, and to resolve technical and practical challenges arising from the introduction and the implementation of school-based assessment practices in Hong Kong schools. However, in the process this dialectal relationship allowed the researchers to better understand and improve their own models of feedback and assessment for learning and gain a clear sense of what adaptions and modifications were needed for case study schools.

Conclusions

The analysis of the professional conversations illuminated how the teachers and researchers grappled with the theoretical, practical and technical challenges that they encountered when they collaborated to negotiate how to adapt Western theories of learning in a non-Western socio-cultural context before the SBA was formally introduced in Hong Kong schools. Three key insights emerged from this case study. One, it underscored how by engaging teachers to rethink their feedback practices helped the researchers understand some of the broader challenges of implementing assessment for learning practices, including what the teachers thought about teaching, assessing and learning more generally. This then informed and shaped theory-building in ways which provided the assessment reform in Hong Kong with long-term sustainability and legitimacy. Two, assessment for learning practices have to be examined, problematised and understood within the situated practices of teachers as well as within the socio-cultural and socio-political context of schools. Third, teachers need practical, technical and theoretical support to tease out new assessment for learning principles, but at the same time, conducting school-university collaboration in the context of government reform is highly problematic so it is important to adopt research practices that actively involve teachers as co-inquirers rather than as participants.

Acknowledgements We would like to thank the Hong Kong Quality Education Fund for funding this research, and the participating researchers, teachers and schools for their extensive assistance with this study. We are also grateful for the constant encouragement and support of our research team at the University of Hong Kong who helped so much with the data collection and analysis.

Appendix A: Schools Involved in the Study

School	No. of teachers	Level	No. of students	Location	EMI/ CMI	Banding
School 1[a]	4 teachers	Form 3	16 students (4 groups of 4)	New territories	CMI	2
School 2[a]	2 teachers	Form 2	8 students (2 groups of 4)	Kowloon	CMI	3
School 3	3 teachers	Form 1 Form 3 Form 2	9 students 3 groups of 3	Kowloon	EMI	2
School 4[a]	4 teachers	Form 3	16 students (4 groups of 4)	Kowloon	CMI	3
School 5	3 teachers	Form 2 From 2 Form 1	9 students (3 groups of 3)	New territories	EMI	1
School 6	1 teacher	Form 1	20 students	New territories	CMI	2
School 7	3 teachers	Form 1	100 students	New territories	EMI	1–2

[a]Case study schools included in this chapter

References

Assessment Reform Group. (1999). *Assessment for learning: Beyond the Black box*. Cambridge: University of Cambridge School of Education.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–75.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81–90.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning*. New York: Open University Press.

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge: Cambridge University Press.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.

Chan, C. (2015). Tensions and complexities in school-university collaboration. *Asia-Pacific Journal of Education, 35*(1), 111–125.

Chan, C. (2016). *School-University partnerships in English language teacher education: Tensions, complexities, and the politics of collaboration*. Norwell: Springer.

Cheng, L., Andrews, S., & Yu, Y. (2010). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing, 28*(3), 221–249.

Clarke, S. (1998). *Assessment in the primary classroom: Strategies for planning, assessment, pupil feedback and target setting*. London: Hodder and Stoughton.

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37–68.

Davison, C., & Mok, J. (Eds.). (2009). *Aligning assessment with curriculum reform in junior secondary English language teaching*. Quality Education Fund/Faculty of Education, The University of Hong Kong.

Davison, C., & Hamp-Lyons, L. (2010). The Hong Kong certificate of education: School-based assessment reform in Hong Kong English language education. In L. Y. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner*. New York: Routledge.

Davison, C., & Leung, C. (2009). Current issues in teacher-based English language assessment. *TESOL Quarterly, 43*(3), 393–415.

Dweck, C. S. (2007). The perils and promises of praise. *ASCD, 12*(65), 3439.

Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (Vol. 1, pp. 487–504). Norwell: Springer.

Hamp-Lyons, L. (2016). Implementing a learning-oriented approach within English language assessment in Hong Kong schools: Practices, issues and complexities. In G.-X. Yu & Y. Jin (Eds.), *Assessing Chinese learners of English* (pp. 17–38). London: Palgrave.

Hamp-Lyons, E. M., & Tavares, N. J. (2011). Interactive assessment: A dialogic and collaborative approach to assessing learners' oral language. In D. Tsagari & I. Csepes (Eds.), *Classroom-based language assessment. Language testing and evaluation series* (pp. 29–46). Frankfurt: Peter Lang.

Harris, L., Brown, G., & Harnett, J. (2015). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: Applying Hattie and Timperley's feedback model. *Assessment in Education: Principles, Policy and Practice, 22*(2), 265–281.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212.

Lantolf, J., & Poehner, M. (Eds.). (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York: Routledge.

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69–85.

Lee, I. (2011). Working smarter, not working harder: Revisiting teacher feedback in the L2 writing classroom. *Canadian Modern Language Review, 67*(3), 377–399.

Phillips, L., & Jørgensen, M. (2002). *Discourse analysis as theory and method*. London: Sage Publications.

Qian, D. D. (2010). Implementing school-based assessment in Hong Kong: Government policy and stake-holders' perceptions. In Z. S. Lü, W. X. Zhang, & P. Adams (Eds.), *ELT at tertiary level in Asian contexts: Issues and research* (pp. 108–119). Hong Kong: The Hong Kong Polytechnic University.

Richards, K. (2006). *Bridging discourses in the ESL classroom: Students, teachers and researchers*. London: Continuum International Publishing Group.

SBA Consultancy Team. (2008). *Aligning assessment with curriculum reform in junior secondary English language teaching*. Hong Kong: Hong Kong Examination and Assessment Authority/ Faculty of Education, The University of Hong Kong.

Stimpson, P., Lopez-Real, F., Bunton, D., Wai-Keung Chan, D., Sivan, A., & Williams, M. (2000). *Better supervision better teaching: A handbook for teaching practice supervisors*. Hong Kong: Hong Kong University Press.

Sutton, R. (1998). *School-wide assessment: Improving teaching and learning*. Wellington: New Zealand Council for Educational Research.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching and learning assessment in the classroom*. Buckingham: Open University Press.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–404.

Van den Bergh, L., Ros, A., & Beijaard, D. (2014). Improving teacher feedback during active learning: Effects of a professional development program. *American Educational Research Journal, 51*(4), 772–809.

Wiliam, D. (1999). Formative assessment in mathematics part 2: Feedback. *Equals: Mathematics and Special Educational Needs, 5*(3), 8–11.

Part III
Addressing Problems of Theory and Practice Through Researcher-Teacher Engagement

Chapter 8
Advancing Written Feedback Practice Through a Teacher-Researcher Collaboration in a University Spanish Program

Kathryn Hill and Ana Maria Ducasse

Abstract This chapter reports a study involving a collaborative dialogue (Scarino, Lang Test 30(3):309–327, 2013, Pap Lang Test Assess 6:18–40, 2017) between an expert in classroom-based language assessment (Kathryn) and an expert language teacher (Ana Maria). The study explores how *theory*, in the form of a teacher assessment literacy (TAL) resource developed by Kathryn (Hill, Pap Lang Test Assess 6:1–17, 2017), and *practice*, in the form of Ana Maria's written feedback practices, might inform one another. The context for the study was an Australian university-level L2 Spanish program, in which students submitted written assignments in the target language as part of their regular course requirements. Participants included 15 students across beginner (CEFR A1), intermediate (CEFR B1), and advanced (CEFR C1) levels of the Spanish course. Data comprised copies of written feedback, recordings and transcription of unstructured think-aloud protocols (by Ana Maria and students respectively) (Ericsson and Simon, Protocol analysis: verbal reports as data. MIT Press, Boston, 1993) and semi-structured interviews (by Kathryn with Ana Maria and students respectively), and notes from our reflective journals. Data were analysed using thematic content analysis. Using the questions framing the TAL resource, analysis started with a consideration of the type and focus of feedback and the theories, understandings, or beliefs informing the feedback practices, as well as learner responses and contextual influences. Discussion examines how the collaboration informed revisions to both the TAL resource and Ana Maria's feedback practices and highlights the importance of incorporating learner perspectives.

K. Hill (✉)
La Trobe University, Melbourne, Australia

A. M. Ducasse
School of Global, Urban and Social Studies, RMIT Melbourne Institute of Technology, Melbourne, Australia
e-mail: Anamaria.ducasse@rmit.edu.au

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_8

Introduction

There has been a tendency to think of the researcher-teacher relationship as asymmetrical and unidirectional. That is, teachers are often positioned as the recipients of 'expertise', as occurs in many professional development programs wherein teachers are informed of what the research community has determined to be best practice (for discussion, see introduction to this volume by Poehner and Inbar-Lourie). At the same time, it has been proposed that researchers might take greater account of teachers' experience, perspectives and practical knowledge (Freeman 2007; Golombeck 1998; Poehner 2014) as well as to examine what experienced teachers actually do when engaged in "the activity of language teaching and learning" (Freeman and Johnson 1998, p. 413). In the case of assessment, some have argued that classroom teachers might be especially well placed to determine appropriate methods and criteria for assessing their students, even citing evidence of the effectiveness of assessment practices among teachers with little or no formal theoretical training (Black et al. 2002). However, the move to allow teachers greater autonomy in classroom assessment has raised questions regarding teachers' theoretical knowledge of assessment and its connection to their assessment practice (see, for instance, Chap. 6, Baker and Germain, this volume). Instantiating again the asymmetrical and unidirectional relation between research and practice, studies of language teacher assessment literacy tend to proceed from a 'deficit' model whereby the aim is to identify gaps in teachers' knowledge and skills in order to determine professional development needs.

An alternative understanding of the relation between theory/research and practice has recently come to the attention of L2 researchers: praxis. According to this view, theory and practice are in a dialectic relation, with each informing and modifying the other (Lantolf and Poehner 2014). As Poehner and Inbar-Lourie (Chap. 1, this volume) explain, a praxis orientation in L2 assessment re-envisions the role of both assessment researchers and assessment practitioners, including classroom teachers. Praxis positions them as collaborators in the sense of differential expertise and therefore differential contributions to partnerships in which neither have full control or full authority and from which both may learn. A praxis orientation to L2 assessment approaches the classroom, inter alia, as a site for testing and refining theory, which then offers the teacher-as-collaborator a principled basis for improving their assessment practice. In this view, the teacher-researcher relationship may be characterised as a collaboration between experts, with both teacher and researcher seen as potential beneficiaries of the process.

This chapter reports on a collaboration between two experts: one, a researcher in classroom-based language assessment (CBA) and teacher assessment literacy (Kathryn), and the other, a highly experienced Spanish as a foreign language university lecturer and academic (Ana Maria). While their partnership is longstanding and ongoing, in this chapter we discuss a focused joint investigation into how Ana

Maria's written feedback practices and a research-based teacher assessment literacy (TAL) resource (Hill 2017) (designed by Kathryn) might each inform the other. In particular, Kathryn wanted to test the utility of the TAL resource (Appendix 8.1) and Ana Maria wanted to develop a better understanding of her feedback practices and how they might be improved. For the purposes of the study, the TAL resource represented the 'theory' and Ana Maria's written feedback practices represented the 'practice'. With this in mind, our project sought to explore two main questions: How does the TAL resource advance the teacher's feedback practices (practice)? How do the teacher's feedback practices inform the TAL resource (theory)?

In line with tenets of Vygotskian theory, as discussed in several chapters in this book (Chap. 10, Davin and Herazo; Chaps. 2, 1 and 9, Michell and Davison; Poehner and Inbar-Lourie; and Poehner and van Compernolle), we note that the TAL resource served as a tool or 'mediating artifact' in that it offered potential to enhance and even transform practice. Of course, given the praxis orientation we have described, this was not a one-way relation but a dialectical one; that is, the practical activity in which Ana Maria employed the TAL resource may lead to revisions to that tool. Given the ongoing nature of our collaboration, we consider primarily the process through which Ana Maria came to reflect upon her feedback practices by referencing the TAL resource as a tool. We conclude with provisional remarks concerning the implications of this process for the refinement of the TAL resource, a topic that will be further investigated in future work.

Approach

In an attempt to move away from a more hierarchical teacher-researcher relationship, the approach to the study involved an ongoing 'collaborative dialogue' (Scarino 2013, 2017, Chap. 3, this volume) involving two distinct participant roles; 'teacher' (Ana Maria) and 'researcher' (Kathryn). Kathryn has had extensive experience in the fields of language testing research and educational measurement with a particular focus on classroom-based language assessment and teacher assessment literacy. Most recently she has been involved in the redesign of curriculum and assessment practices across a range of university disciplines. Ana Maria is a university level lecturer with extensive experience in teaching Spanish at three different universities in Melbourne, Australia. She is committed to continuous improvement in her teaching practice, achieving high student satisfaction ratings, and has been awarded university citations for excellence in teaching. Other areas of expertise include oral language assessment, rating scales, and rater behaviour. Ana Maria found the theory encountered during her postgraduate studies helpful for reflecting on her assessment practice and promoting discussions with teaching colleagues. She continues to use published research to inform her practice, for example by trialing a validated writing assessment rubric for an integrated reading and listening task.

Table 8.1 Relationship between CBA framework & TAL resource

Study	Basis	Intended users
CBA framework (Hill 2012)	Description of classroom practice	Other researchers
TAL resource (Hill 2017)	CBA framework & TAL literature	Teachers

The teacher-researcher collaborative dialogue was guided by the TAL resource (Appendix 8.1), which was designed as a heuristic for promoting teacher reflection on CBA practices. The starting point for the resource was a framework derived from a detailed observational study that Kathryn had conducted documenting the classroom-based assessment (CBA) practices of expert (foreign) language teachers (Hill 2012; Hill and McNamara 2012). In response to research suggesting teachers experienced difficulty identifying their assessment literacy needs (e.g., Tsagari and Vogt 2017), Kathryn suggested teachers were in need of a structure to aid systematic reflection on their assessment practices (Hill 2017). She reasoned that, although originally designed for researchers, a framework which was grounded in observed CBA practices might usefully be reconfigured as a tool for helping classroom teachers to understand and evaluate their own assessment practices. While the original framework (Hill 2012) is based on a documentation of what teachers actually do, the TAL resource could be seen as more prescriptive, representing what the research suggests teachers *should* do. We hasten to add that the TAL resource is prescriptive in the sense of providing to teachers an orientation to assessment practices rather than rigidly directing their actions; that is, in line with praxis, our collaboration is contingent upon Ana Maria's negotiation of her practice in light of the TAL resource and of the resource in the context of her practice. The relationship between the original framework and the resource used in the study is summarised in Table 8.1.

Participants

While much has been written about what constitutes good feedback practices, it is axiomatic that feedback is only effective to the extent learners actually use it to improve their learning. Andon et al. (2017) for example, found significant variation in the extent to which postgraduate learners engaged with the feedback provided with a number of them choosing to disregard the feedback altogether. For these reasons we believed it was important to also engage the learners in dialogue to incorporate the learner perspective in any consideration of the effectiveness of Ana Maria's feedback practices.

Learners were recruited as participants via a group-email inviting expressions of interest. Each participant received an AUD20 gift voucher as honorarium. Participants comprised 15 student volunteers (aged 18+) enrolled in courses from

Table 8.2 Student participants

Year level	Course level	CEFR level	No. of participants	Cohort (% of class)
1	Level 1 (Beginner)	(A1)	5	28 (18%)
2	Level 4 (Intermediate)	(B1)	5	29 (17%)
3	Level 8 (Advanced)	(B2-C1)	5	13 (38%)

beginner (roughly equivalent to CEFR A1), intermediate (roughly equivalent to CEFR B1), and advanced (roughly equivalent to CEFR C1) level courses in the university Spanish program (Table 8.2). Note that the university offers Spanish courses over a 12-week term and organizes them across 8 levels from beginner to advanced. Students' entry level is officially determined by a diagnostic placement test, although the majority start at Spanish 1 upon enrolling at the university. The majority of students are pursuing their undergraduate degree in International Studies and are required to complete at least four semesters of language study as well as an international internship with placements in a non-English speaking country. Entry to the course is very competitive and successful candidates tend to be highly motivated. A small minority of students enroll in Spanish as an elective, often in the final semester of their degree, and typically do not progress beyond Level 1.

Data

At the time of data collection Ana Maria was teaching each of the targeted three levels. This chapter will focus on analysis of the data for one of the Level 8 assessment tasks. As part of their assessment at that level, students produce a 3000-word reflective journal ('Memoria') recording their personal responses to the material covered in the course. Students were required to submit reflections during at least 10 of the 12 weeks of the course. In week 3 students were asked to bring their first two reflections (i.e., for weeks 1 and 2) to class for a peer-assessment activity. Students were given a copy of the marking criteria and asked to use them to assess one another's reflections with the aim of helping students arrive at a clearer understanding of expectations for their written work (Appendix 8.2). While the students typically had previous experience with reflective writing in other subjects, they often still had questions concerning what was required, and these were used to elaborate the original criteria (Appendix 8.3). Students had the option of putting their reflections on Google docs for feedback at any time up to and including week 11 (the penultimate week of the semester). Eight of the 13 Level 8 students (including all five study participants) took advantage of this opportunity. Students were permitted to submit as many drafts as they liked with the proviso that Ana Maria would only provide additional feedback if the student in question had responded to feedback provided on the previous occasion.

Data were collected over the course of a 12-week semester and comprised the following: Ana Maria's unstructured think-aloud protocols (Ericsson and Simon 1993); semi-structured interviews conducted with Ana Maria and with individual students; ongoing collaborator discussions; and various documents by Ana Maria in her practice (e.g., assessment tasks directions and rubrics).

Teacher Think-Aloud 'Marking' Protocols

Ana Maria was asked to produce unstructured think-aloud protocols during the process of providing written feedback on each student's work.[1] Specifically, she received the following prompt: "Just say whatever you're thinking as you're marking each student's work." These remarks were recorded and transcribed.

Student 'Response' Protocols and Interviews

In accordance with the conditions of ethics approval, interviews with student volunteers were scheduled after the semester had ended and student marks had been finalised but before students received their final results. This ensured that the feedback, and what it meant to them, was still current and relevant to participants. As far as possible, appointments for each level were scheduled so that students at any given level were scheduled for the same day. This arrangement was chosen for pragmatic reasons (often students who had been classmates would arrive together or organise to meet up before or after their scheduled appointment time) as well as to give Ana Maria a better sense of what was going on at each of the three course levels.

Each session lasted approximately 15 min. After obtaining signed consent forms, Kathryn provided each student with a copy of their feedback and asked them to 'think aloud' as they read it: "Talk to yourself as you're reading through. Just your reactions to what you're reading." Participants occasionally needed prompting to remind them to speak. All participants were also asked a core set of questions to elicit their response to the specific feedback they had received as well their thoughts

[1] It is widely acknowledged that the relationship of think-aloud protocols to actual mental processes is problematic. However, this concern was mitigated to some extent through the provision of multiple opportunities to revisit the data and interrogate the ways in which it was being interpreted. This was achieved, for example, through a process of reviewing the original 'marking' protocols in relation to learner responses ('reflection' protocol) as well as through regular 'member checking' (Lincoln and Guba 1985).

concerning feedback in general.[2] These questions, and what they attempted to uncover, are as follows:

1. Understanding of the feedback
   In terms of the actual written feedback on the reflective writing task was there anything that was unclear to you?
2. Self-evaluation
   How do you expect you've gone on this? What about your mark sheet, is this more or less what you'd expect?
3. Affective responses to feedback
   When you're about to receive feedback how do you feel, what's your emotion around that?
4. Utility of feedback
   Have you ever had any unhelpful feedback? In Spanish and in general? Is there anything you would have liked in addition to what she's given you here?
5. Uptake of feedback
   Based on this feedback what will you do differently?

Additional questions followed up on issues arising from the protocol or interview phases (e.g., "Is that because of the nature of the task? I mean if you were doing an actual exercise on 'gender, number, agreement' would you get it right?"). The sessions with students were also audio-recorded and transcribed.

Teacher 'Reflection' Protocols

The reflection exercises were scheduled over three sessions, one for each course level, starting with Level 8. For this phase of the data collection Ana Maria was provided with a transcript of the protocol she had produced when marking a given student and asked to 'think aloud' again as she reviewed what she had said. She was then immediately given the transcript of the same student's interview and asked to 'think aloud' once again as she read what the student had said. This process was then repeated for each of the other students at the same level. In contrast to the original protocols, Kathryn was physically present and asked questions during the process (e.g., "Did reading that change anything about your perceptions?"). This was audio-recorded and transcribed.

[2] What the student participants said in interviews was inevitably influenced by factors including the hierarchical relationship of researcher to student and the line of questioning Kathryn adopted, along with any assumptions or biases she may have brought to the process (Mann 2011). This is also true of the ways in which data were interpreted.

Discussions

The final phase of formal data collection comprised a discussion between the two of us based on issues arising during a review of the transcripts from the previous phases (i.e., the teacher 'marking' protocols, student 'response' protocols and interviews, Ana Maria's reflections on her original protocols and students' responses). Topics for discussion included:

- the teacher's experience of recording the protocols
- the effect of revisiting her own protocols and reading the students' responses to her feedback
- the implications of the findings for her feedback practices
- the utility of the framework, and
- research articles (relating to issues arising during the process of data analysis).

Informal discussions between the two of us were (and are) ongoing.

Documents

Documents included copies of tasks and criteria as well as de-identified copies of written feedback on students' work at each level. We each kept a journal comprising notes from meetings, personal reactions to what was discussed, insights & reflections.

A summary of the different phases of data collection is provided in Table 8.3.

Table 8.3 Phases of data collection

Phase	Process	Stimulus	Data	Participants
1. 'Marking' protocols	Articulate thought process while providing feedback	Student work	Recordings & transcripts	T[a]
2. 'Response' protocols	Read & respond to feedback & discuss	Written feedback	Recording & transcripts	Sb, R[c]
3. 'Reflection' protocols	Reflect on original feedback & student responses to it	Transcripts: T marking protocol, S response protocols R/S discussions	Recordings & transcripts	T, R
4. Discussions	Explore emerging themes	All data research articles	Recording & transcripts	T, R
5. Ongoing reflection & discussion	Interrogate preliminary interpretations, identify negative cases, refine themes	All data research literature	Notes of meeting	T, R

[a]Teacher (Ana Maria) [b]Students [c] Researcher (Kathryn)

Analysis

Data analysis was guided by questions from the TAL resource. For the first question (*What does the teacher do?*), written feedback was coded according to type (tick, underline, comment, etc.) and level of explicitness in Excel. This latter coding was informed specifically by the Regulatory Scale developed by Aljaafreh and Lantolf (1994) and that has been highly influential in the area of Dynamic Assessment (for examples, see Chaps. 9 and 10 in this volume by Davin and Herazo and Poehner and van Compernolle). The Regulatory Scale arranges forms of support offered to students (e.g., leading questions, identifying the presence or location of an error, offering a choice, etc.) from most implicit to most explicit, thereby allowing for identification of the degree of support a learner needed to overcome a problem. For the remaining questions, data were analysed through a thematic content analysis approach using NVIVO (Version 11). This process involved repeated reviewing and discussion of transcripts (of protocols, interviews and discussions) and documents (journals and written feedback) in order to become familiar with the data and to identify broad themes. Coding was organised under each of the five questions of the TAL resource (specified in what follows). Data were constantly revisited to 'complicate', refine and combine codes, identify additional themes, and make connections between them as necessary.

Results and Discussion

For simplicity, we organize our presentation of findings according to the research questions that guided the collaboration as discussed earlier.

RQ 1. How does the TAL resource advance the teacher's feedback practices (practice)

The first research question was addressed through discussion of the insights gained into Ana Maria's assessment feedback practices and their value in supporting students' learning and is organised under five guiding questions posed by the TAL resource (Hill 2017). That is:

1. What do you do (types of feedback)?
2. What do you look for (features of writing, quality & standard)?
3. What beliefs or understandings inform your feedback practices?
4. What are your learners' understandings of feedback?
5. How does the context for teacher shape the nature and content of feedback?

With regard to 'types of feedback,' Ana Maria provided 'content ticks' to indicate that the reflection (Level 8) for any given week met the necessary content requirements; comments (provided in Spanish) were typically general in nature (e.g., '*buena memoria'*). Nevertheless, there were some instances where she provided specific comment on aspects that had been done well (e.g., "*I like the way that*

your reflection is so political"). With regard to identifying and correcting language errors, Ana Maria frequently provided the actual corrections herself, sometimes accompanied by an explanation (e.g., the relevant grammatical rule) in the target language. Errors which students were expected to be able to self-correct were flagged using the following schema: underlining ("there is a mistake here"), broken line (repetition of the same error), or circling (to make connections, e.g., repeated errors). A smiley face was used to indicate that the student had inadvertently written something funny. During the 'reflection' protocol phase (where Ana Maria reflected on the transcript of her original feedback and student responses to the feedback) she restated her policy of providing less corrective feedback in the sections where it was clear that feedback provided on earlier drafts had not been taken up by the student concerned.

In terms of what Ana Maria 'looked for' when providing feedback, analysis revealed discrepancies between the focus of her think aloud marking commentaries for each student and the written feedback provided on the paper itself. According to Ana Maria, the purpose of feedback provided on draft documents during the semester was to ensure students were critically reflecting on the course and how it was taught, rather than merely describing what had been covered. In contrast, while there was comment on the quality of the reflections in relation to the published criteria in her verbal protocols, there was still a significant focus on accuracy (e.g., grammar, punctuation and style). Furthermore, the *written* feedback provided on the final (graded) versions of students' work focused almost exclusively on accuracy. In other words, what Ana Maria recorded on the page did not match earlier feedback on drafts or her comments during the marking process nor did they reflect the published criteria.

Given their involvement in elaborating the marking criteria for the task, and that these did not include a focus on accuracy, students were understandably puzzled by the mixed messages. As one Level 8 student (M) put it:

> Student M: *I don't know if I registered that we weren't supposed to [write] colloquially in this task… I thought that if it was a pure reflection you would just spew it out and I think I've written it maybe how I speak.*

When Kathryn followed this up during the discussion phase, Ana Maria continued to insist on the importance of accuracy, criteria notwithstanding:

> *AM: I've already read these [on Google docs] and I've pointed them in the right direction as regards content. I've given them feedback whether they were actually doing the reflections in the right way. What I haven't done is gone over the surface errors and the nitty gritty.*

She later added that:

> *AM: At that level they're looking beyond, they're going to look at the more complicated things, … but I'm still going to mark them, they should still know they're not accurate even though they can communicate high-level thoughts [emphasis added].*

This apparent discrepancy between the formal criteria and the nature of the feedback provided is elaborated in the following section.

Turning to the question of 'beliefs and understandings' that inform Ana Maria's feedback practices, her preoccupation with accuracy appears consistent with other research that suggests this persists as a key focus for many language teachers, regardless of other beliefs they may hold concerning language and communication and their expressed pedagogical commitments (Hyland 2003). Nevertheless, upon continued probing Ana Maria remarked that she had been unconsciously applying an external standard, the Diploma in Spanish as a Foreign Language (DELE), an internationally recognised certificate of proficiency in Spanish. As this standard had been applied unconsciously it was essentially 'hidden' from students. Ana Maria's written reflections provided further insight into how she came to be referencing this particular benchmark:

> *AM: I had been teaching with the DELE levels in mind before training as an examiner 5 years earlier. I had experience of teaching with C1 textbooks and even this course was not a C1 course students from B2 though C1 could take part the highest benchmark would have meant performance at C1 level.*
>
> I had not even realized that I had done this until writing this. *One of the tests included an integrated task in the style of DELE Task 2. I was emulating the DELE even with the assessment without thinking, "Oh I had better copy the DELE C1 …". I just prepared a task that followed on from the work we had done in class and that they were used to doing [this in class] so the test mirrored class and was not something they were not prepared for [emphasis added].*

Analysis identified two additional assumptions or beliefs that appeared to underlie some of Ana Maria's feedback practices. The first of these can be described as a 'mastery' view of learning resulting in an intolerance of error. In her think-aloud protocols she frequently expressed frustration at the number of 'basic' errors (e.g., agreement in terms of number and gender) students were still making at this level as well as the inconsistency with which they made them.

Kathryn also suspected that Ana Maria may have been defining achievement exclusively in terms of independent, error-free, performance (i.e., rather than taking a more developmental view of error, i.e., where errors are perceived as a normal and inevitable part of the process of learning a language and where changes in the explicitness of feedback required by learners could provide evidence of development)). With this in mind Kathryn asked Ana Maria to read two articles, Aljaafreh and Lantolf (1994) and Lantolf and Poehner (2011), and to use the 'Regulatory Framework' to reflect on her feedback in terms of the degree of explicitness required and how this changed over time. In subsequent discussions Ana Maria remarked that she had never previously thought about feedback as something that was scalable or that the need to provide progressively less explicit feedback to a given student over time (e.g., moving from Level 12 to Level 9 on the Regulatory Framework) could be interpreted as evidence of learning.

A second belief concerned Ana Maria's assumptions about the motivation of some of her students. She frequently expressed exasperation with two of the Level 8 students in particular, describing them as 'lazy', 'careless', or 'too confident' in their approach to writing. Such remarks occurred when she recorded the original, 'marking', protocol and again when she reviewed the transcript of the protocol in the

'reflection' phase. It was evident that these attributions influenced how she oriented to errors in the students' work (e.g., "He's just saying the first thing that comes out of his mouth."). However, some of these attributions were directly contradicted by the testimony from the students in question:

> *Student M: I started second guessing my grammar… I often don't know where I'm going wrong as I suppose is the case when you're learning a language. You sort of try to the best of your ability because there's no reason to be half-arsed, you know. And I wasn't in any way embarrassed to try my best … It's really rare that I would hand something in that I wasn't happy with*

There is evidence that on comparing her original comments in the 'marking protocol' with the student's 'response protocol', Ana Maria considered that her initial assumptions may not have been correct:

> *AM: Ooh, that's interesting. I thought it was a bit more casual than trying his best. So he was actually a bit more committed to being better than [my] perception.*

In the 'discussion' phase, Ana Maria suggested that these assumptions may have stemmed at least partly from a perception that students tended to give Spanish less priority than their other subjects.

The fourth point included in the TAL resource concerns learner understanding of feedback. By virtue of their participation in this project, all participants were clearly receptive to feedback, at least in principle. They were also unanimous in expressing a belief that error and correction were fundamental to language learning. One learner, E, offered the following insight:

> *Student E: You can't progress unless you get feedback because you don't know what you're doing correctly. But in a lot of other subjects ... It's just "this is wrong". Well how can you make it better? It doesn't matter because it's not a developmental learning process. Whereas with language you always build on your previous knowledge.*

Beyond this general receptivity, the Level 8 students, while not reporting any particular difficulty understanding feedback, varied with regard to the extent to which they found it valuable. Here we are reminded of Leung's (2014) identification of four different types of responses (or dispositions) to written feedback in his research involving postgraduate students: 'rejectionist', 'happy, let it pass', 'critical acceptance', and 'fulsome reception' (for further discussion, see Chap. 5 by Leung, this volume). In our work, we found particular differences among learners in their responses to feedback on accuracy. M explained,

> *Student M: But grammar and subjunctives and preterits and all this stuff I forget, I know how to use them in Spanish [and] to some extent in English fairly well but I forget what they're called. So sometimes when I just get feedback that says "pronouns joined after infinitive" my honest reaction is like "whatever that is…. The blurb which says, "M, you did a great job of this. You should work on this", I pay fairly close attention to that.*

M was more concerned with communicating his thoughts and taking risks than with demonstrating his ability to producing error-free writing. Another Level 8 student reported that she had used the Memoria task as an opportunity to experiment with imitating the different literary styles she had been exposed to during the course (e.g., "I didn't write it to be super correct; I wrote it to try new things…"). She also expressed a belief that accuracy was something that could wait:

Student E: I think that a lot of the errors I make just generally from this reflection are things that would be fixed if I was in a context where I was exposed to Spanish more often.

Students also varied in their affective responses to feedback. Participants were asked how they felt about coming to get the feedback on their Memoria task as well as about their feelings around receiving feedback more generally. One of the students admitted to feelings of anxiety about receiving feedback:

Student P: I always get a bit nervous before I'm getting a result back whether I've done well or not… [If I haven't done well] I always leave it, push it to the side for a day or two... because I'd prefer to forget about it, get over it and then come back to it.

On the other hand, some participants told Kathryn that Ana Maria's feedback made them feel cared about. When she asked whether they ever felt overwhelmed by the volume of feedback, they responded:

Student M: No. It's definitely an indication that the teacher cares and I really respond to it when you can tell the teachers [care].

Student E: I always want to do well in Spanish because it's the subject that I care about the most I think maybe because it makes me [feel] most cared about.

The final point included in the TAL resource asks how the context for teaching shapes the nature and content of feedback. There is increasing recognition of the situated nature of classroom-based assessment (Inbar-Lourie 2008; Scarino 2013, Chap. 3, this volume). Turner and Purpura (2017) suggest that contextual influences can be divided into macro and micro-level factors. Examples of macro-level factors include the socio-political context of education as well as sociocultural norms influencing parental and student expectations of teaching and assessment. A significant influence on feedback practices in the Spanish course at the socio-political level is the CEFR. The main textbooks used throughout the course are benchmarked against CEFR and, as discussed earlier, Ana Maria's preoccupation with accuracy was inspired by the DELE (itself reviewed in response to the CEFR), which offers Diplomas at the six CEFR levels (A1 to C2).

Relevant influences at the micro-level include what Andon et al. (2017) have characterised as the 'pedagogic space' (comprising organisational, pedagogical and curriculum-related factors) as well as teacher and learner-related factors. According to Andon and colleagues, teacher factors include L2 proficiency and teaching experience as well as the type of classroom environment she creates (e.g., the degree of learner autonomy). Relevant student factors include L1, previous L2 learning experience (formal and 'in country'), interests and motivation.

In the present case, institutional factors that appeared to have impacted Ana Maria's feedback practices include a marking policy where any student who questions their grade can request that the work be re-marked. As a result, Ana Maria commented that she tends keep in mind both the second marker and the student when she provides feedback, that is, she endeavors to ensure that the prospective second marker can see the reasoning behind the marks that she gives. This effectively results in more detailed feedback on final tasks than she might otherwise provide.

At a more local, classroom, level, Ana Maria reflected during discussions that the focus on accuracy at the exclusion of other aspects of the writing may have been partly an artefact of the marking grid. That is, she provided feedback in relation to the 'official' criteria by ticking the relevant box on the marking grid and while feedback on the paper itself related to other, 'unofficial', criteria (i.e., accuracy). This is attested to by a comment Ana Maria made during the 'marking' protocols.

> *AM: I've really mucked up… When I tick on the column [in the marking grid] I'm not really ticking on the page so they can't see, I'm not marking on the page what's good actually.*

There is evidence that timing of feedback (draft vs final) also influenced the nature and focus of the feedback. When Kathryn asked if there were any differences between the feedback she provided on the drafts (on Google doc) with the feedback provided on the final versions of the Memoria task, Ana Maria wrote:

> *AM: I think it is interesting what is marked as feedback to the student when I am not thinking remotely about the mark or providing a grade. The feedback is not from a 'this is all wrong' perspective rather than the idea that a 'you might like to know this'. 'You can take it up'.*

In other words, as might be expected, feedback on drafts had a more formative orientation than feedback on the final version.

Due to the small class size, Ana Maria had a fairly detailed understanding of a number of student attributes. Discussion of specific instances of written feedback provided evidence that she instinctively adjusted the level of explicitness in her feedback according to her expectation of what students "should know". This expectation was based on students' first language, previous experience in learning Spanish (e.g., whether or not they'd spent time 'in country' and where) and whether they had studied other languages as well as on her knowledge of the Level 8 curriculum and what she had previously taught them – within Level 8 and/or in previous years (e.g., "She wouldn't know how to correct that if I underlined it").

When Kathryn asked if there were any differences in her feedback on students' drafts, Ana Maria noticed that they were "all marked up quite differently…" (Table 8.4).

> *AM: Looking across the five Level-8 pieces three are showered with content ticks and two have none. I find this very unnatural for me not to have ticked for ideas. It is almost as if I was busy looking at expression without even noticing the content!*

Table 8.4 Comparison of feedback type x Level 8 student

Participant	Content ticks	Underlining, etc.	Correction	Explanation	Circling	Comments
K_S8	Y	Y	–	–	–	–
L_S8	–	Y	Y	Y	Y	Y
P_S8	–	Y	Y	Y	Y	
E_S8	Y	Y	Y	Y	Y	Y
M_S8	Y	Y	Y	Y	–	Y

Upon further investigation, it emerged that the two students in question (Students L and P*)* were the ones known to place a premium on accuracy. In other words, Ana Maria had unconsciously tailored her feedback almost exclusively to what she knew to be the learners' orientation but only where this was consistent with her own (as discussed earlier). Nevertheless, there is some evidence that Ana Maria's knowledge about the other students' interests and motivation may have had some influence on the nature of the feedback she provided to them. For example, the following comment appears to reflect Ana Maria's awareness of the student's passion for literature: "I'm going to give her a bit of stylistics here because she would probably like that."

For Ana Maria, the findings underscore the importance of developing in her learners a shared understanding of curricular goals and quality and standard as well as of the nature and purpose of feedback. They also highlighted the need to find a way of paying "attention to student views and dispositions without losing sight of pedagogic purpose" (Leung 2014, slide 28). Leading on from this she recognised the need to identify and interrogate her own 'covert' criteria and standards as well as any assumptions about the motivation of individual students.

At the time of writing Ana Maria had specific plans to increase learner involvement in decisions around how feedback is approached including discussion of the philosophy of learning which underpins the Regulatory Framework. However, in general, the process aroused Ana Maria's curiosity about the reasoning behind specific instances of feedback:

> *AM: There are many, many errors left unmarked. So why do I choose the ones I do? Hopefully in the protocol that comes out! Is it just random or is there method in the madness?*

We now turn to the second research question that guided the project.

> *RQ 2. How do the teacher's feedback practices inform the TAL resource (theory)?*

As we mentioned at the outset, our primary concern in this chapter has been the process through which Ana Maria came to re-examine her feedback practices as she employed the TAL resource as a mediating tool. Indeed, the five framing questions used in the TAL resource appear to have provided a productive basis for collecting relevant data to analyse Ana Maria's written feedback practices. That said, and given our commitment to praxis, we have begun to consider how this collaboration compels us to revisit the TAL resource itself. While this work is ongoing, we wish to mention some points that came immediately to our attention.

It became clear that one shortcoming is that the resource needed to ascribe a more dynamic role to learners than simple 'understanding'. That is, just as Ana Maria's assumptions and beliefs have been found to impact on her assessment practices, learner motivations and beliefs about language learning seem to have influenced how they oriented to the feedback and whether, and to what extent, they chose to engage with it. In addition, the situation which prompted Kathryn to give Ana Maria the Aljaafreh and Lantolf (1994) article suggests the resource might usefully be supplemented with suggested readings and 'worked' examples. Another insight

into the utility of the resource was the manner in which it was used in the study. That is, rather than functioning as a static 'stand-alone' tool for Ana Maria to reflect on her CBA practices, the 'break-through' insights often emerged as a result of engagement in our ongoing dialogue about what the results appeared to be showing. This experience suggested that it might be more useful as a tool for generating discussion and reflection amongst colleagues (e.g. Maher and Ducasse 2019) rather than for individual reflection. For example, the resource was subsequently used in a one-day professional development workshop for primary and secondary level Spanish teachers, run by the Spanish Government-appointed language advisor who reported that it generated an extensive and highly productive discussion. It is likely that the TAL resource will continue to be refined and adapted in response to the experience of its implementation by different practitioners in different contexts.

Conclusion

This chapter has reported on a study involving a teacher-researcher partnership with the aim of understanding and advancing the teacher's written feedback practices while simultaneously testing the utility of a theory-based resource for this purpose. The collaborative approach used in this study helped both teacher and research to realise their goals. Specifically, the TAL resource enabled Ana Maria to reflect on her feedback practices in a more systematic way than in the past; the process of engaging with Ana Maria in relation to her practice provided insights into how the TAL resource might best be used. However, we both recognise that the success of this collaboration depended to a large degree upon a relationship of trust. We therefore offer some reflections on this important feature of the partnership.

At a 'process' level, trust between teacher and researcher was based on a prior history of successful collaboration. It was further nurtured over the course of this project through a collaborative approach to research design and goal-setting, joint coding and analysis of data, regular and ongoing dialogue about the nature of the findings, and collaboration in dissemination through presentations and publications. At an interpersonal level, teacher-researcher trust relied on a mutual recognition of our respective areas of expertise, a commitment to continual improvement, a readiness to engage in critical reflection on feedback practices, and receptiveness to different interpretations and ideas.

However, despite the original intention, there was a sense that Kathryn had retained primary control of the shape of the collaboration. While agreeing with this perception, Ana Maria suggested that this was probably inevitable given that Kathryn had initiated the project and developed the TAL resource that framed it. That being said, as ensuing research projects have been jointly conceived (Hill and Ducasse 2019), we have continued to strive for a more genuine equality in our partnership. We have come not only to view the classroom as a 'testing ground for theory' but to regard our collaboration as praxis jointly conceived and undertaken by equals with distinct but complementary expertise.

Appendices

Appendix 8.1: TAL Resource

1. What do you do?	
1.1 Planning assessment	How does planned assessment relate to teaching & the intended learnings (including relevant curriculum standards & frameworks)?
	How are learners' existing knowledge, language background, capabilities, and interests taken into account?
	How are learners' social, emotional & psychological attributes taken into account?
	What role do learners have in setting learning goals and making decisions about when, how & why they will be assessed?
1.2 Framing assessment	How do learners become aware of when, how & why they will be assessed?
1.3 Conducting assessment	Who carries out assessment (*teacher, student, peers, others*) & whose judgement 'counts' in grading decisions?
	What proportion of assessment is planned & formal and what proportion is unplanned & incidental (*e.g., observation*)?
	What evidence of learning is provided by routine classroom activities & interactions (*e.g., class discussions*)?
	Who is the main target of informal (incidental) assessment (*the whole class, groups/pairs, individual students*)?
	Does formal & informal assessment focus on *processes* and well as *products*, e.g., are learners encouraged to discuss the basis for their responses?
	Where do formal assessment activities come from (*e.g., textbook, self-designed, other teachers*) & how well do they fit the intended purpose (see 1.4) in terms of nature, scope & level?
	Do you use a range of assessment methods and is the method appropriate for the intended purpose (see 1.4)?
	How do you ensure the fairness, quality & reliability (trustworthiness) of assessment activities and processes?
	Is assessment conducted in an ethical manner (*e.g., preserving student confidentiality)?*
1.4 Using assessment	To document growth in learning
	To judge & grade students
	To report to stakeholders (*student, parents, school, external authorities*)
	To prepare students for exams
	To inform teaching
	How is assessment used to diagnose needs & plan teaching?
	How is assessment used to evaluate teaching?
	To enhance learning, motivation & self-regulation by providing quality feedback
	Does feedback focus on features of performance (*rather than on innate qualities e.g., 'intelligence'*)?
	Does feedback explain which aspects were done well? *e.g., 'You used a good variety of vocab & sentence structures'*
	Does feedback tell the student how to improve? *e.g., 'You need to review the work we did last week on use of the passive form'*
	Does the timing (*immediate/delayed*) & format of feedback (e.g., *comments only vs. marks*) encourage learner uptake?
	To manage teaching
	Is assessment used to discipline learners or to encourage them to work harder?
	Is assessment used to socialize learners into a new assessment culture *(e.g., using assessment rubrics, preparing for high stakes exams)*

(continued)

2. What do you look for?
What is the balance of skills (*listening, speaking, reading, writing*), knowledge (*vocab, grammar, cultural*) and abilities addressed in assessment?
Does this reflect the relevant curriculum priorities?
What are the valued qualities (*e.g., accuracy, fluency, variety*), behaviours (e.g*., effort, presentation, attendance*) & student-centred factors (*e.g., well-being*) communicated in ● written or verbal instructions, ● written or verbal feedback, and ● formal reporting?
Are these qualities consistent with the intended learnings (*including relevant curriculum standards & frameworks*)?
3. What beliefs or understandings do you use?
What are your beliefs and understandings about
the nature of the subject (*the nature of language; relationship of language & culture*)
how students learn a second language
how language should be taught, &
how language should be assessed (*e.g., learner agency, appropriate uses of assessment*)?
What is the basis for these beliefs & understandings?
How do these beliefs & understandings influence your assessment practices?
4. What are your learner's understandings?
How do you ensure students understand the focus & purpose of assessment?
How do students perceive their role in planning, conducting & judging assessment?
How do you ensure students have understood & engaged with feedback?
5. How does the context for teaching shape your assessment practices?
Who influences decisions about content & methods in your assessment (*school, supervisors, students, external authorities*)?
What other factors do you need to take into account when planning & conducting assessment (*e.g., class size, learner characteristics, external examinations, student & parental expectations*)?
What is the impact of testing and assessment practices both locally and in the broader context and what is your capacity to influence change?

Appendix 8.2: Marking Criteria

No interroga el material o si mismo. No explora lo que piensa. No explora lo que siente. No se distancia del evento. No considera información externa.	Interroga el material o a si mismo pero no forma o responde a las preguntas. Empieza a explorar sus reacciones Empieza a explorar lo que siente. Algo de distancia del evento. Se refiere a algo de información externa	Señales de que se auto cuestiona. Las emociones se contextualizan y se cuestionan. Ejemplos de distanciamiento y cuestionar en contexto. Se refiere a teoría o investigación sobre el tema	Análisis crítico y personal de su propio comportamiento Se reconoce el papel de las emociones en formar una perspectiva sobre acontecimientos. Examina la situación y la valora desde varios puntos de vista Se refiere a la bibliografía y se integran citas correctamente.

Appendix 8.3: Elaborated Marking Criteria

Criterios y consideraciones para la nota de la memoria Valor (total 30%)	Criteria and considerations for marking the reflection task (total 30%)
Largo 300 x 10 (calidad no cantidad), Número de entradas y o regularidad de entradas 0–5	Minimum 10 x 300-word entries. Number and regularity of entries
Relación entre las entradas en la memoria y la semana relevante del curso 0–10	Relevance to the course
Se evidencia enfoque en profundidad y detalles de las reflexiones al contenido de la memoria 0–3	Depth and thoroughness of reflection
Claridad y buenas observaciones al presentar cuestiones o eventos 0–3	Clarity, depth and detail of observations
Presentación y legibilidad 0–3	Presentation and legibility
El contenido refleja los objetivos del curso y la intención pedagógica de la memoria 0–3	Reflects on the objectives and pedagogical approach of the course
Hay auto preguntas sobre el proceso de aprendizaje que provocan más reflexión con honestidad y autoevaluación 0–3	Serious and honest self-reflection
Representación de diferentes destrezas cognitivas (síntesis, análisis, evaluación 0–3;	Represents a range of cognitive skills (synthesis, analysis, evaluation)
Se evidencia de especulación; 0-3	Evidence of speculation
Se evidencia disposición hacia a revisar ideas; 0–3	Evidence of willingness to change mind
Se evidencia pensamiento creativo 0–3	Evidence of creative thinking
Se evidencia pensamiento crítico 0–3/30	Evidence of critical thinking

References

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal, 78*(4), 465–483.

Andon, N. J., Dewey, M., & Leung, C. (2017). Tasks in the pedagogic space: Using online discussion forum tasks and formative feedback to develop academic discourse skills at Master's level. In *TBLT as a researched pedagogy* (Task-Based Language Teaching). Amsterdam: John Benjamins Publishing Company.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. London: GL Assessment.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press. ISBN: 0-262-05029-3.

Freeman, D. (2007). Research "fitting" practice: Firth and Wagner, classroom language teaching, and language teacher education. *The Modern Language Journal, 91*, 893–906. https://doi.org/10.1111/j.1540-4781.2007.00676.x.

Freeman, D., & Johnson, K. E. (1998). Reconceptualizing the knowledge-base of language teacher education. *TESOL Quarterly, 32*(3), 397–417.

Golombeck, P. (1998). A study of language teachers' personal practical knowledge. *TESOL Quarterly, 32*, 447–464.

Hill, K. (2012). *Classroom-based assessment in the school foreign language classroom* (Language Testing & Evaluation Series) (Vol. 27). Frankfurt am Main: Peter Lang.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment, 6*(1), 1–17.

Hill, K. & Ducasse, A. M. (2019). *Developing student educational feedback literacy using technology and the Reflective Feedback Conversation*. Practitioner Research in Higher Education.

Hill, K., & McNamara, T. (2012). Developing a comprehensive research framework for classroom-based assessment – An empirical study. *Language Testing, 29*(3), 395–420.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*(2), 217–230.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402.

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for L2 development. *Language Teaching Research, 15*, 11–33.

Leung, C. (2014, October). *Learning from feedback: conception, reception, and consequences*. Paper presented at the TCCRISLS Conference, New York.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Maher, B., & Ducasse, A. M. (2019). Teaching and assessing language and culture through translation. In J. Fornasiero (Ed.), *Papers of the Language & Culture Network of Australian Universities-LCNAU*. Adelaide: Adelaide University Press.

Mann, S. (2011). A critical review of qualitative interviews in applied linguistics. *Applied Linguistics, 32*(1), 6–24.

Poehner, M. (2014). *Dynamic assessment and mediated learning: Toward a coherent theoretical framework for developmental L2 education*. Paper presented at the TCCRISLS Conference, New York.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309–327.

Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. *Papers in Language Testing and Assessment, 6*(1), 18–40.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment, 6*(1), 41–63.

Turner, C. E., & Purpura, J. E. (2017). Learning-oriented assessment in the classroom. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment*. DeGruyter Mouton: Berlin/Boston, MA.

Chapter 9
Reconsidering Time and Process in L2 Dynamic Assessment

Matthew E. Poehner and Rémi A. van Compernolle

Abstract The pre-test – treatment – post-test design in experimental research has long been a hallmark of psychological and educational studies emulating natural science research practices. In Dynamic Assessment (DA), an analogue to this design emerged as a means of understanding learner responsiveness to mediation (Haywood, Lidz, Dynamic assessment in practice. Clinical and educational applications. Cambridge University Press, New York, 2007). DA derives from L. S. Vygotsky's (The collected works of L. S. Vygotsky. Volume 1: problems of general psychology, including the volume thinking and speech (Rieber RW, Carton AS, eds). Plenum, New York, 1987) argument that psychological functions that are not yet fully developed but are still ripening may be made visible through joint functioning on the dialogic plane of interpsychological functioning. Specifically, the quality of dialogic support learners require while cooperating with a teacher, or mediator, along with their responsiveness to such support is interpreted as indicative of how near they are to more independent functioning. In DA, the use of a pre-test serves to establish a baseline of learner abilities prior to the introduction of mediation while post-tests allow for determining gains made by learners following mediation. Sternberg and Grigorenko (Dynamic testing. The nature and measurement of learning potential. Cambridge University Press, Cambridge, 2002) dubbed this the *sandwich* format of DA, and it has been the most widely used in cognitive education and psychology.

In this chapter, we report the implementation of a sandwich DA project involving U.S. university learners of L2 French. The aim of the project was to help teachers mark student mastery of particular features of the language as they progress to the next level of study. In line with previous DA research (e.g., Davin and Donato 2013; Poehner et al, Lang Test 32(3):337–357, 2015), pre- and post-test comparisons were

M. E. Poehner (✉)
Department of Curriculum and Instruction, College of Education,
The Pennsylvania State University, University Park, PA, USA
e-mail: mep158@psu.edu

R. A. van Compernolle
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: vancomp@cmu.edu

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_9

found to provide insights into learner abilities. However, additional conceptual work regarding manifestations of learner abilities during the procedure was required to more fully capture development. Following the work of Ollman (2003) on dialectics, which forms the driving logic behind Vygotsky's theory, a reorientation to the sandwich format of DA is proposed. We argue that the entire procedure constitutes the assessment, a position that contrasts with the more conventional view of two distinct measures and an intervening treatment phase. Regarded as a unity, the procedure brings into focus not isolated snapshots of learner abilities at two points in time but rather a picture of the processes of learner development through time.

This shift in view offered qualitatively different insights into learner abilities. In particular, learner initial performance, quality of interaction, receptivity to instruction, and subsequent changes to performance afforded a profiling of abilities in the process of development and documentation of instructional investment required to provoke change. Given the practical value of such information to teachers and learners, we advocate increased attention to the importance of time in assessment, not as a discrete occasion for glimpsing learner abilities but as unfolding across activities.

Introduction

Sociocultural Theory (SCT), as developed in the writings of Russian psychologist L. S. Vygotsky, first came to the attention of L2 researchers with the appearance of Frawley and Lantolf (1985). Since that time, the theory has been referenced by researchers interested in processes of L2 development in instructional contexts, and since the early 2000s it has been the basis for a praxis orientation to L2 research and education (see Lantolf and Thorne 2006). Increased availability of Vygotsky's writings in English has also revealed his own commitment to using theory as a way to not only understand human psychological functioning but to actively intervene to guide its development. Indeed, when outlining his vision for a scientific psychology, Vygotsky (1997, p. 331) explained the task as needing "to learn from Marx's whole method…how to approach the investigation of the mind." As Lantolf and Poehner (2014) explain, Marx's (1978) famous Eleventh Thesis on Feuerbach, in which he argued that the task of philosophy was to change the world rather than to simply interpret it, strongly influenced Vygotsky's scientific enterprise. Specifically, this view compelled Vygotsky to unify theory with practice as a dialectic, that is, as *praxis* (for further discussion, see Michell and Davison, Chap. 2, this volume). When one considers that Vygotsky's professional efforts co-occurred with the post-revolutionary period of significant social upheaval in Russia, it is not surprising that many of his most well-known proposals emerged from his attempts to work out theoretically grounded recommendations that could both take account of pressing problems and work toward remediating them. In this regard, Vygotsky's efforts foreshadowed many of the concerns and undertakings reported in the chapters in this volume. That is, as Poehner and Inbar-Lourie argue in their introduction to this book, an appealing feature of praxis lies in its rejection of the entrenched dichotomy

between theory/research and practice. They explain that consequences of that dichotomy have included theory and research undertaken for their own sake as well as difficulties in determining any relevance of the resultant knowledge to practice. Praxis enables us to practice as fundamental to the elaboration and refinement of theory and theory as providing practice with a necessary orientation and framework.

An illustration of the centrality of praxis in Vygotsky's work, which connects directly to the topic of our chapter, Dynamic Assessment, is the notion of the Zone of Proximal Development, or ZPD, often described as the difference between what one is capable of doing alone and what becomes possible with intervention (e.g., assistance, teaching). According to van der Veer and Valsiner (1991), Vygotsky's first known mention of the ZPD occurred in a 1933 lecture concerned with understanding why some children, but not others, experience gains in measures of general abilities, such as IQ, over the first year of formal schooling. As van der Veer and Valsiner explain, Vygotsky hypothesized that the first-year curriculum was sufficiently challenging for some children that it promoted development of their abilities while other children were already close to mastering the requisite forms of reasoning and problem-solving even at the start of schooling. For this latter group of children, Vygotsky reasoned, the curriculum provided little opportunity for development beyond their current abilities. To test his hypothesis, Vygotsky supplemented the existing beginning- and end-of-year ability measures with a procedure in which the examiner interacted with children when they experienced difficulties. Vygotsky (1987) referred to the use of hints, prompts, feedback, models, and leading questions to characterize these examiner-child interactions. Several interesting findings emerged from the procedure. First, van der Veer and Valsiner (1991, p. 337) note that Vygotsky distinguished between learners with 'large' and 'small' ZPDs as some learners evidenced much more substantial gains when offered assistance than did others. Second, learner responsiveness to support did not correlate with the measure of their independent functioning; thus, high IQ learners could have either a large or small ZPD, as could learners with a lower IQ. Most important, following the learners over time revealed that it was their ZPD that was a better predictor of their success in school. Put another way, what was revealed during interaction was learners' potential future, which they realized through additional instruction. Thus, Vygotsky's application of his theoretical understanding of psychological development guided his approach to helping educators identify learners who might need additional instructional support as they entered school. At the same time, this work simultaneously spurred his discovery of the ZPD, the significance of which he described as follows:

> The zone of proximal development furnishes *psychologists and educators* with a tool through which the internal course of development can be understood. By using this method we can take account of not only the cycles and maturation processes that have already been completed but also those processes that are currently in a state of formation, that are just beginning to mature and develop. Thus, the zone of proximal development permits us to delineate the child's *immediate future and his dynamic developmental state*, allowing not only for what already has been achieved developmentally but also for what is in the course of maturing (Vygotsky 1978, p. 87, emphasis added).

Not only did the ZPD emerge from Vygotsky's use of theory to examine practical problems, but as this statement makes clear, he regarded it as important for both research and education. That its importance further pertains to diagnosing the full range of learner development is what has inspired researchers around the world working in the area of Dynamic Assessment, or DA. To date, DA has been employed with populations of learners ranging from individuals with special needs to young children, the elderly, and learners studying specific academic subjects, including languages (for an example of DA with L2 teachers and learners, see Davin and Herazo, Chap. 10, this volume).

In their chapter, Michell and Davison (Chap. 2, this volume) report the creation of an online assessment system for learners of English as an additional language. The authors describe this system as a mediating tool that emerged from researcher-teacher praxis and that provides a resource for continued partnerships. In a similar manner, the present chapter is concerned with collaborative efforts to design a tool according to DA principles for use with classroom learners of L2 French. Specifically, we report data from a pilot study of a larger project that aims to create a DA instrument targeting a particular grammatical feature of French, the syntax of negation and adverbs in simple and compound tenses. This is an area where French syntax differs considerably from English, and it is therefore a common source of difficulty for L1 English learners of French (Prévost 2009). Indeed, while it is introduced relatively early in pedagogy (i.e., by the end of the first semester of university study), learners often struggle to control it even as they matriculate to more advanced-level coursework in French and Francophone studies. For the small-scale pilot study we describe here, four university learners enrolled in advanced coursework were recruited to participate in one-to-one sessions with an assessor, or mediator. The choice to include advanced, and not lower-level, students was motivated in part by a desire in the French program at the university where the study was conducted to develop diagnostic tools for students who may not have gone through the elementary and intermediate course sequence (e.g., they placed into an advanced course after having studied the language in secondary school) that may be used by instructors to determine what, if any, remediation students require in order to succeed.[1] As in Vygotsky's research on school-aged children with different ZPDs cited above, there is considerable variation in proficiency levels starting at the third year of French study because of students' varied language learning backgrounds: some students have strong control over academic French while others are still struggling with basic communicative language. This heterogeneity can lead to difficulties in the classroom where academic content is inaccessible to some students. Indeed, in addition to attempts to develop a diagnostic test using DA principles, the program has also begun to offer a 'mini course' in French grammar and stylistics that upper-

[1] In many US universities, advanced-level coursework in Modern Languages tends to focus on 'content' as opposed to 'language' in that students read, discuss, and write about cultural and literary topics, with little if any overt language instruction. The current study was part of a larger initiative to explore ways in which the teaching of language and culture can be integrated into the curriculum at all instructional levels and in a formal, systematic way.

level students can enroll in if they feel they need additional language help in order to succeed in their advanced course work.

As explained below, diagnosis of learner knowledge of French was approached through a three-phase assessment procedure: (1) a pretest; (2) a mediation session in which the pretest was reviewed and test-takers cooperatively revised their errors with the mediator; and (3) a retest. One of the goals of the pilot study was to devise procedures that can eventually be computerized so that teachers can assign their students to complete the test and gain insights into learner control over syntax that reflects independent abilities (i.e., pretest and retest) as well as the extent to which they benefit from an instructional intervention, which can be determined through comparison of their score before and after the intervention. As we explain, however, the study underscored for us the valuable insights into processes of learner development that can guide teacher and learner work and that are difficult to represent through scores alone. We argue that while comparing scores preceding and following instruction is certainly a step beyond a single conventional test, the interests of teachers and learners may be further served by the creation of learner profiles that include not only scores but also how individuals responded to instructional support during intervention and what this reveals about their developing knowledege of, and ability to use, the target grammatical features. Integrating such information in a tool for teacher use ultimately will allow for a better representation of learner developmental processes as they unfold in real time. This re-orientation to both the object of assessment and assessment administration procedures follows from Vygotsky's dialectical way of thinking, which we propose is at the heart of his commitment to praxis but is also a feature of his writings that researchers are only beginning to appreciate.

Background: Toward a Dialectical DA

Since Vygotsky's death, researchers around the world working with a wide range of populations have drawn on the ZPD concept to devise a host of procedures that have come to be known as DA (Haywood and Lidz 2007). Discussion of these approaches is beyond the scope of this chapter, and we refer readers to overviews by Lidz and Elliott (2000), Sternberg and Grigorenko (2002), and Poehner (2008). Generally speaking, these approaches accept Vygotsky's (1978) premise, discussed earlier, concerning the developmental significance of individuals' independent performance and their engagement in cooperation with an assessor or teacher, often referred to as a mediator. The former reveals abilities that have already developed, which Vygotsky termed the *zone of actual development,* while the latter reveals abilities in the process of forming, that is, the *zone of proximal development*. A common point of reference then for DA researchers is Vygotsky's (1998) argument that assessments that do not take account of the ZPD and instead focus exclusively on learners' independent performances can only provide a partial view of their abilities. Aside

from this theoretical commitment, however, the extent to which Vygotsky's ideas have influenced the design of particular DA procedures varies considerably.

With few exceptions (e.g., Kozulin 1998; Minick 1987), DA research in general education and psychology has appropriated the ZPD as a technique of providing support to learners during assessments and has not engaged with the broader orientation to science behind Vygotsky's thinking (Chaiklin 2003; Holzman 2009). As Davydov and Radzikhovskii (1985) explain, appreciating Vygotsky's enduring impact on so many disciplines as well as how quickly he was able to elaborate such complex proposals requires understanding him as both an *empirical researcher* and a *methodologist*. The former appears relatively straightforward: the greater portion of Vygotsky's published studies reports work with young children or learners with special needs. While it was indeed in those contexts that Vygotsky worked out features of his theory such as the ZPD, mediation, and internalization, it would be inaccurate to classify him as a developmental or child psychologist (John-Steiner and Souberman 1978). The reason for this becomes clear when Vygotsky is approached as a methodologist, that is, as concerned with elaborating a vision of a general psychology that unifies all sub-disciplines and specalized areas according to coherent scientific principles. Limiting one's reading of Vygotsky to his empirical work and to discussions of his theory in secondary sources can result in an impoverished appreciation of concepts such as the ZPD. In our view, careful reading of Vygotsky's work allows an understanding of the ZPD as much more than a technique for assisting test-takers because it situates the ZPD concept in the orientation to praxis that was central to his enterprise.

Vygotsky's Method

'Vygotsky the methodologist' (e.g., Vygotsky 1997) recognized that psychology had not developed a framework for conceptualizing its object of study, units of analysis, or explanatory principles and that very different orientations to building a scientific psychology were in competition. One of the dominant views held that psychologists needed to adopt an experimental approach that emulated practices in the natural sciences. Vygotsky, however, argued forcefully that psychology had to devise its own approaches to research, and the method he eventually proposed is referred to as the *genetic method*. As its name implies, the method sought to understand consciousness through its genesis, that is, through the processes of its formation. For this reason, Vygotsky focused much of his work on children and introduced auxiliary forms of mediation, which he reasoned "artificially provokes or creates a process of psychological development" (Vygotsky 1978, p. 61).

Van der Veer and Valsiner (1991) describe a classic Vygotskian experiment conducted by Vygotsky's colleague, A. N. Leont'ev, that illustrates this process. Children and adults were recruited to complete a task during which they were instructed that they were not allowed to say the name of a given color (e.g., red, green, blue). To assist them, participants were offered cards that corresponded to the

forbidden colors, but they were not directed how to use them. In this way, the cards were an external resource that had the *potential* to serve as mediation, and indeed they did function in this manner for some participants. According to Van der Veer and Valsiner (1991), children aged 5–6 years found the task difficult, and they were unable to use the cards in a helpful manner; in fact, the presence of the cards appeared to increase their confusion. Older children, aged 8–13, in contrast, were also challenged by the task but managed to succeed by employing the cards in a meaningful way, usually by holding or separating the card that corresponded to the color they were not supposed to say and glancing at the cards before verbally responding to questions. Adults, interestingly, did not require the cards. As Van der Veer and Valsiner (1991) explain, their psychological activity was still mediated, but the entire process occurred internally as they were able to employ verbal thinking to regulate their functioning. It is worth noting that had only adults participated in the study, the researchers would have been limited to noting observable behaviors; this can be interesting, as one might record reaction time or track eye gaze. However, the researcher's task becomes inferring underlying psychological processes on the basis of what can be observed. Genetic method offers the possibility of tracing the development of psychological functions as they are formed first on an external plane and later move inward. This insight also relates to Vygotsky's discussions of the ZPD. The provision of such external resources that come to be appropriated as forms of mediation allows the development of an individual to be traced over time, following how they come to use resources in more intentional ways and sometimes eventually reaching a point where they are able to function without external assistance. It is with this process in mind that Vygotsky and Luria (1994) maintained that the most interesting insights are gained prior to the start of most experiments, when the subjects are learning how they will be expected to perform.

The preceding example also points to the importance of dialectics for appreciating Vygotsky's genetic method and indeed his understanding of mediation and development. We will not go into a full discussion of dialectics here (for detailed discussion, see Ilyenkov 1977; Novack 1971), but we do wish to underscore that Vygotsky's commitment to dialectics as a meta-theory for understanding the world had implications for his own theoretical proposals as well as for current DA practices. In brief, Ollman (2003, p. 13) characterizes dialectics as "expanding our notion of anything to include, as aspects of what it is, both the process by which it has become that and the broader interactive context in which it is found." In this way, dialectics challenges the concepts that cultures make available to us for thinking as these tend to break reality into pieces or elements that, while offering the advantage of allowing us to contemplate manageable 'chunks' of reality, have a tendency to distort our resultant understanding of them. As Ollman (ibid.) explains,

> dialectics restructures our thinking about reality by replacing the commonsense notion of 'thing' (as something that has a history and has external connections with other things) with notions of 'process' (which contains as part of what it is its ties with other relations).

Perceiving the world in terms of isolated entities rather than processes and relations makes sense in terms of formal logic but is not necessarily helpful for thinking

about reality. As an example, consider the everyday phenomenon of 'day.' Commonsense tells us that day is day and night and night, and indeed the influence of formal logic (and nondialectical thinking) is so pervasive that many of us think of day and night as opposites; it is seemingly obvious that day cannot be night and night can never be day. Nonetheless, in reality what we think of as day, marked by the presence of sunlight, and night, characterized by the absence of light, exist on a continuum. Together they form a unity, which in English is often referred to by the same word 'day' to denote a 24-h period. Science defines day as the measure of time required for the earth to make one full rotation on its axis. Thus, the appearance, duration, and absence of sunlight are processes related to the Earth's rotation on its axis as well as its orbit around the Sun (i.e., the process of year). Our language of day and night obscures this process and deceives us into perceiving distinct, opposing states. Dialectics allows us to see that day has no meaning unless it is understood in relation to night.

Building a psychology according to dialectics, as Vygotsky did, means rejecting dualisms such as 'social' versus 'cognitive', 'internal' versus 'external', 'learning' versus 'development', 'past' versus 'future', and 'teaching' versus 'assessment', among many others. Instead, a psychology rooted in dialectics regards our psychological functioning to be mediated through external means and through internal means, as illustrated in the color experiment we described earlier. Similarly, the importance of such mediation for our functioning changes through time, as functioning that was at one time occurring on the external plane of engagement with others and reliance on external artifacts comes to occur more fully or even exclusively on an internal plane. Indeed, this insight is central to the ZPD concept and to its importance for education as Vygotsky envisioned. As he explained, provision of external resources, including through dialogic interaction, allows for identification of abilities that are emerging and that are most amenable to instructional intervention; those same processes of mediation as learners reach the limits of their independent functioning constitute a powerful pedagogical approach to provoke learner development (Vygotsky 1998). It is in this way that assessing and teaching exist in dialectic relation and that a learner's history of development, through independent performance, is brought into contact with a potential future development (revealed by responsiveness to mediation) in the present.

L2 DA and Praxis

Although DA was only introduced to L2 researchers in the early 2000s (Lantolf and Poehner 2004), it has been pursued in the fields of general education and psychology since the 1960s. However, because this work has generally not engaged explicitly with Vygotsky the methodologist but only with Vygotsky the empirical researcher, this has led to a preponderence of approaches and studies that reference concepts such as mediation and the ZPD but are decidedly non-dialectical in their embrace of controlled experimentation and measurement of abilities that are often

operationalized in a manner analogous to variables in natural science research. Interestingly, in their survey of DA research, Sternberg and Grigorenko (2002) conclude that the most widely used format for integrating mediation during a procedure is through a three-stage process that they dub a *sandwich*. In the sandwich format, learners independently complete a series of tasks that comprise a pre-test to establish a baseline of their abilities. This is followed by the introduction of mediation that targets problem areas identified by the pre-test. A follow-up post-test in which learners independently carry out tasks parallel to those employed in the pre-test allow for determining gains made by learners following mediation. This organization bears a striking similarity to a classic experimental design, which may partly account for the popularity of the sandwich format with educational and psychological researchers. The pre-test – treatment – post-test design in experimental research has long been a hallmark of psychological and educational studies emulating natural science research practices (Philips and Burbules 2000). The basic logic behind the design is that it allows for comparison between two points in time and plausible linking of changes to the intervening treatment.

In addition to the sandwich format, applications of DA outside the L2 field bear another similarity with experimental research in that they are often carried out by professional psychologists rather than educators and occur in clinical settings instead of instructional settings (e.g., classrooms). As Haywood and Lidz (2007) observe, in such situations DA is only conducted with learners upon referral from a school psychologist, and when the results of DA are shared with teachers and other stakeholders, they often lack the knowledge necessary to meaningfully integrate the information into subsequent practice. Following Haywood and Lidz, a disconnect thus remains between DA and the everyday practices of teaching and learning. Some DA researchers have worked to propose models for more closely tying DA tasks and goals to classroom learning experiences (see Lidz and Elliott 2000). Among the most successful applications of DA have been those linked with cognitive education programs, wherein the abilities targeted by DA procedures are also the focus of ongoing intervention programs that can be tailored to the needs and strengths of individual learners (Tzuriel 2011). As Feuerstein et al. (2010) observe, however, these programs tend to occur outside of the confines of mainstream educational settings, be they for young children or adults. Thus, while much can be learned from such work, its capacity to reach a large number of individuals has been limited. A praxis that brings theoretical insights to bear on problems that affect people's lives requires joint undertaking with those whose lives are affected. In our work, this means engaging with L2 teachers and learners.

L2 DA has been marked by an understanding of dialectics all along. We suspect that this is because the L2 field has a strong history of serious engagement with Vygotskian theory and a commitment to improving language instruction, especially in the classroom. Whatever the reason, L2 researchers have typically approached DA as a form of practice developed from Vygotsky's ideas, that is, as a form of praxis. The first in-depth study of DA with L2 learners, undertaken by Poehner (2008), identified processes of mediation that can occur during dialogic interaction with learners and how these may simultaneously function to offer assessment

insights into learner abilities as well as instruction to support continued development. This work has been extended to applications carried out cooperatively with L2 teachers in classroom settings involving learners from primary school through university studying a variety of languages and at different proficiency levels (Davin and Donato 2013; Poehner 2009; Siekmann and Charles 2011; van Compernolle and Williams 2012). As understanding of different forms of mediation that might be pursued with learners has advanced, L2 DA researchers have leveraged these insights for the creation of DA instruments that can be used for more formal assessment purposes (e.g., Levi 2012). Computerized approaches to L2 DA offer the possibility of providing mediation to large numbers of learners simultaneously and thus broadening the base of knowledge about learner abilities that can be used for making assessment decisions such as acceptance into a program of study or placement at an appropriate level of instruction (Leontjev 2016; Poehner et al. 2015; van Compernolle 2016; Zhang and van Compernolle 2016).

As we detail in the rest of this chapter, our efforts to design a computerized DA tool to help teachers diagnose and trace learner developing control over French syntax led to our adoption of a sandwich approach to DA. While we conceived of this format in dialectical rather than experimental terms, meaning that we regarded the entire three-step procedure as an activity of provoking and interpreting learner abilities, we nonetheless came to recognize limitations in our initial conceptualization of how such a tool might function. Specifically, through implementation of the theoretical principles behind DA with actual learners, it became clear that exclusive reliance on learner scores at different points in time did not adequately capture the dynamics of development. For that, attention must be given to learner struggle and responsiveness when mediation is made available. Consequently, while we examine test scores as one index of development, we also analyzed videorecordings of the mediation sessions in order to identify pivotoval moments in which we identify development ‘in flight’, as Vygotsky (1978, p. 67) said; that is, in microgensis.

The Study

Methods

As noted, the current assessment procedure focuses on two related features of French grammar with which learners whose L1 is English often demonstrate considerable difficulty: the placement of negation and of adverbs of manner and frequency in compound tenses (Prévost 2009; and see van Compernolle 2016 for a related analysis of interactional data from the study). Table 9.1 shows example sentences with common learner errors.

The assessment comprises three phases: a pretest of independent abilities, an interactive mediation phase, and a retest of independent abilities. Altogether, the assessment takes approximately 40 min–1 h to complete, depending on the amount

Table 9.1 Illustrative examples of grammatical features with common learner errors

Feature	French example	Common learner error
Negation in compound tense	*Je **n'**ai **pas** vu Jean* 'I have not seen Jean'	Placing second negative after the past participle, as in *Je **n'**ai vu ***pas** Jean*
Adverbs of manner and frequency	*Il a **rapidement** mangé son dîner* 'he quickly ate his dinner'	Placing adverb in preverbal position, as is possible in English, *Il ***rapidement** a mangé son dîner*
Adverbs of manner and frequency in a negative sentence	*Elle **n'**a **pas toujours** aimé le vin blanc* 'she hasn't always liked white wine'	One or both of the errors cited above, e.g., *Elle ***toujours n'**a aimé ***pas** le vin blanc*

of mediation required during the teaching phase. We hasten to add, as highlighted by Poehner and van Compernolle (2016), van Compernolle (2016), and van Compernolle and Zhang (2014), that all three phases of the procedure are considered together to be the assessment. This follows from the dialectical point of view that understands teaching and assessing to exist in relation to one another. As van Compernolle (2016, p. 178) explains, from this perspective "assessment is about responsiveness to teaching, not about a test-taker's manifest level of competence in one task or another." In the following sections, brief explanations of the instruments, the approach to mediation, evaluation procedures, and participants are provided.

Pretest and Retest Instruments

Two reconstructive elicited imitation tests (one each for the pretest and the retest) were designed for the assessment. Briefly put, elicited imitation involves test-takers listening to and repeating stimuli sentences, and it is often used as a means of tapping into language learners' implicit (i.e., nonconsciously controlled) linguistic competence (Vinther 2002). Reconstructive elicited imitation (Erlam 2006, 2009) pushes learners to focus on the meaning of the stimulus before repeating it. The idea is to build in a secondary task so that learners' attention to form, and therefore their ability to consciously monitor their performance, is minimized. Consequently, test-takers must reconstruct the sentence based on their comprehension of its meaning rather than relying on rote memorization of a string of sounds. Reconstructive elicited imitation is therefore a likely reflection of the current state of a learner's developing linguistic system.

For the present study, items were designed to be long enough to minimize the possibility of rote memorization (between 12 and 15 syllables), and a comprehension task was included in which the test-taker was prompted to recall the meaning of the sentence in English prior to repeating it in French. Thus, for the stimulus *Ma sœur Sophie n'a pas aimé le film hier soir*, the test-taker would be expected to recall the meaning in English, *My sister Sophie didn't like the movie last night*, and then to reconstruct the sentence in French. Each test included 18 items: 6 focused on the placement of negation; 6 focused on the placement of adverbs; and 6 focused on the

placement of both negation and adverbs. Half of the stimuli were ungrammatical, reflecting common learner errors, in order to evaluate test-takers' abilities not only in reconstructing well-formed sentences but in (automatically) correcting mistakes as well (see Erlam 2009). For instance, the test-taker would be expected to correct an ungrammatical stimulus such as *Sa tante Sandra vraiment a aimé son repas ce soir* 'Her aunt Sandra really liked her meal this evening' (error = placement of the adverb *vraiment* 'really' in preverbal position) as *Sa tante Sandra a vraiment aimé son repas ce soir* (correction = moving *vraiment* to post-auxiliary verb position).

Approach to Mediation

Mediation was integrated into the test during a "teaching phase" between the pretest and the retest. During the pretest, the mediator tracked the test-taker's performance using a score chart that included space for notes about problems in comprehension as well as in repetition of the pretest stimulus sentences. The scorecard and the pretest items were then used to frame the teaching phase. In other words, the mediator used the test-taker's own pretest performance as the basis for intervening with appropriate forms of assistance during the teaching phase.

The teaching phase was conducted as follows. The mediator prompted the test-taker to attempt an item that was missed or partially missed (e.g., comprehension without successful repetition, partial repetition) during the pretest. This was done as an initial attempt to confirm or reject the working diagnosis from the pretest that the item was in fact a source of difficulty for the learner. If the learner successfully recalled the item in English and repeated it correctly in French, the mediator continued to the next item that was missed on the pretest. If, however, the learner continued to experience difficulty, the mediator intervened to support the recall and/or repetition of the sentence. Although the mediator was free to pursue any and all relevant means of supporting the test-taker (cf. the concept of interactionist DA; Lantolf and Poehner 2004), the following two principles guided the approach: assistance was *graduated* (i.e., from implicit to explicit) and *contingent* (i.e., withdrawn when no longer needed) (see Aljafreeh and Lantolf 1994). In other words, the mediator did not engage in explicit or corrective feedback from the outset, but instead provided low-level hints at first (e.g., a questioning *hmm?*, or saying *that's not quite right…*), moving to more explicit means of assistance (e.g., narrowing the focus of attention to a particular part of the sentence, providing a metalinguistic hint) only if the test-taker needed it, and withdrawing support as the test-taker began to control his or her performance more autonomously. The objective of the teaching phase, following Zhang and van Compernolle (2016), was twofold: on the one hand, it aimed to progressively draw the learner's attention to the appropriate forms and, on the other, to forge a link between the learner's metalinguistic knowledge and controlled performance.

Scoring of Pretests, Posttests, and Learning Potential

Pretest and posttest scores were calculated as follows. Each item was worth 2 points: 1 point for an accurate recall of the meaning in English, and 1 point for a correct repetition in French. Because the assessment focused on negation and adverbs, test-takers received full credit if the structure were correct but they had replaced a particular word with a synonym or near-synonym in the French repetition. Partial credit (a half point) was given for comprehension if the sentence was recalled more or less accurately but it was missing a key piece of information (e.g., rendered in present tense instead of past tense) and for repetition if the sentence was repeated more or less accurately but was missing a key structure (e.g., a simple tense was used instead of a compound tense, an adverb was missing from a sentence that included both negation and an adverb). To capture gains learners made as a result of the teaching phase, a *learning potential score* (LPS) was calculated. LPS was originally proposed by Kozulin and Garb (2002) as an attempt to yield a single score that represents the degree to which a test-taker improves relative to the maximum score possible on the test. In this, Kozulin and Garb argued, LPS is more informative that a simple 'difference' score arrived at by comparing an individual's pretest and retest performance (i.e. such a comparison might yield a gain of 5 or 20 but this is difficult to interpret in itself as learners might remain at the very low or high end of the possible scale). The formula Kozulin and Garb proposed for calculating LPS is as follows (Fig. 9.1).

LPS was employed by Poehner and Lantolf (2013) in their research involving a computerized approach to DA with L2 learners. As they observe, the term 'learning potential' might be interpreted in a number of ways, including as a fixed ability along the lines of IQ. Poehner and Lantolf suggest that a more Vygotskian way of conceptualizing learning potential is as an openness or receptivity to mediation. It is that understanding that informs the present study. As we explain, our procedure generated LPSs for the assessment as a whole (comprehension and repetition) as well as for comprehension and repetition subscores.

Participants

Participants were four advanced-level US undergraduate students in French who volunteered to take part in an initial phase of test development. The mediator during the DA sessions (R. A. van Compernolle) is an experienced teacher of French, a member of the French and Francophone Studies faculty at the university where the research was carried out, and a researcher whose work focuses on the use of Vygotskian pedagogical innovations, such as DA, to support language teachers and

$$\text{LPS} = \frac{(2*\text{retest}) - \text{pretest}}{\text{Max retest}}$$

Fig. 9.1 LPS formula. (Adapted from Kozulin and Garb 2002, p. 121)

learners. Thus, his instructional responsibilities and approach to scholarship cohere around the same theoretical principles, which emphasize a commitment to praxis (see Poehner and Inbar, Chap. 1, this volume).

Among the student participants in this project, three—Chris, Nicole, and Stephanie—were enrolled at the time of the study in a third-year course in French sociolinguistics, and a fourth—Julie—was enrolled in a fourth-year guided research course under the guidance of van Compernolle. Chris and Nicole were native speakers of American English. Stephanie and Julie were native speakers of Arabic and Korean, respectively, although English was their dominant language (i.e., the language they used most frequently in daily life). None of the participants reported having spent a significant time living, studying, or visiting a French-speaking region of the world, nor did any of them report using French frequently outside of their French classes.

Findings

We report our findings in three parts. In the first, we focus on the pretest and posttest scores for all four of the participants. The second and third parts focus on up-close analyses of Nicole's and Chris's test performances, including data relevant to learning potential from the teaching phase of the test.

General Findings

Overall pretest, retest, and learning potential scores are provided in Table 9.2. Subscores for comprehension and repetition are given in Tables 9.3 and 9.4, respectively. Stephanie and Nicole performed relatively well on the pretest, whereas Chris and Julie experienced considerable difficulty, particularly in repeating stimulus sentences correctly. Chi-square tests revealed that, across the four learners, the differences in overall pretest scores were statistically significant ($\chi^2 = 14.814$, $df = 3$, $p = 0.002$) as were differences in pretest repetition subscores ($\chi^2 = 11.822$, $df = 3$, $p = 0.008$). Particularly noteworthy is the gap between the high scorers, Stephanie and Nicole, and the low scorers, Chris and Julie. Differences in pretest comprehension subscores were not statistically significant, however.

Table 9.2 Overall pretest, retest, and learning potential scores

Participant	Pretest	Retest	Change	LPS
Stephanie	28	34	6	1.11
Nicole	25	34.5	9.5	1.22
Chris	16	32	16	1.33
Julie	13.5	31	17.5	1.35

NB: Pretest/retest max. score = 36

Table 9.3 Pretest, retest, and learning potential subscores for comprehension

Participant	Pretest	Retest	Change	LPS
Stephanie	14.5	17.5	3	1.14
Nicole	14	17	3	1.11
Chris	9	17	8	1.39
Julie	9.5	15.5	6	1.19

NB: Pretest/retest max. subscore = 18

Table 9.4 Pretest, retest, and learning potential subscores for repetition

Participant	Pretest	Retest	Change	LPS
Stephanie	13.5	16.5	3	1.08
Nicole	11	17.5	6.5	1.33
Chris	7	15	8	1.28
Julie	4	15.5	11.5	1.50

NB: Pretest/retest max. subscore = 18

On the retest, all four students improved their overall scores as well as their comprehension and repetition subscores. It is especially important to note that Chris and Julie succeeded in closing the performance gap between them and Stephanie and Nicole on the retest: no statistically significant differences were found between the four learners with regard to overall scores or comprehension and repetition subscores. In other words, the differences found in the pretest essentially disappeared following the teaching phase in which the learners were provided support from a mediator. This is also apparent in the testtaker's LPSs, which reflect responsiveness to mediation. Although all four learners had high LPSs, indicating that they all responded well to mediation (i.e., they were able to benefit from it during the teaching phase), Chris's and Julie's LPSs are particularly high because they made the most improvement from pretest to retest.

In what follows, the performance of two of the learners, Nicole and Chris, is examined in greater detail, including analysis of interaction during the teaching phase of the assessment, during which greater insights were gained into specific sources of difficulty and learner emerging abilities. The two learners were chosen for closer analysis because they are representative of the kind of 'gap closing' made possible through DA. Recall that Nicole was a high scorer in the pretest, whereas Chris was a low scorer, yet their retest scores were virtually identical.

Case 1: Nicole

Comprehension and repetition scores for Nicole's pretest and retest are provided in Table 9.5. She experienced difficulty primarily with stimulus sentences that included both an adverb and negation, especially the ungrammatical items. On the retest,

Table 9.5 Nicole's pretest and retest scores

Pretest				Retest			
Grammatical		Ungrammatical		Grammatical		Ungrammatical	
Comp	Rep	Comp	Rep	Comp	Rep	Comp	Rep
7.5	6	6.5	5	8.5	9	8.5	8.5

however, Nicole improved substantially, earning the highest retest score of the four students (see Table 9.2, above).

The teaching phase revealed important information regarding the amount and qualities of mediation that Nicole needed in order to make progress. Namely, Nicole only required a small hint for item #4, *Ta soeur Aimée *lentement a mangé son diner* 'Your sister Aimée slowly ate her dinner' (ungrammatical, adverb): that *Aimée* was the name of 'your sister', not a form of the verb *aimer* (a homonym, e.g., the past participle *aimé* or the imperfective *aimait*). After this hint, she immediately recognized that the adverb *lentement* was in an ungrammatical place, and she corrected it. Subsequently, the mediator moved on the item #6, which Nicole had missed during the pretest: *Ma soeur Jeanne *lentement n'a bu *pas son café aujourd'hui* 'My sister Jeanne didn't drink her coffee slowly today' (ungrammatical, adverb + negation). The exchange is shown in Excerpt 9.1.

Excerpt 9.1

1.	Mediator:	alright, + um:: let's look a number six then. ((plays audio))
2.	Nicole:	okay. so: + my sister + Jeanne + did not drink her coffee slowly today
3.	Mediator:	Mhm
4.	Nicole:	um: but they put they put it in the wrong spot again.
5.	Mediator:	okay,
6.	Nicole:	so + it should be ++ *n'a:: + lente::ment-* ++ mm
7.	Mediator:	if you have a negation and an adverb, what do you do.
8.		(3.0)
9.	Nicole:	um (2.5) *n'a* + *pas* + *len::tement*? + mm *bu*?
10.	Mediator:	mhm, + yeah. *n'a pas lentement bu.*
11.	Nicole:	okay.
12.	Mediator:	so listen again, + and tell what's wrong here. ((plays audio))
13.	Nicole:	it said *lentement n'a:: pas bu*? + instead of *n'a pas lentement bu.*
14.	Mediator:	okay.

Nicole needed only minimal support here. She recognized without any assistance from the mediator that the adverb was in the wrong syntactic slot (turn 4), which she attempted to correct independently (turn 6). However, she failed to retain the negative word *pas* 'not' here. Nonetheless, a low-level metalinguistic prompt from the tutor (turn 7) sufficed in supporting Nicole's correct repetition of the sentence (turns 9, 13). Throughout the remainder of the teaching phase of the assessment, Nicole

continued to comprehend and correct such sentence with little to no assistance, and she was able to maintain her control over this feature on the retest of independent performance, as noted.

Case 2: Chris

Chris's pretest and retest scores are given in Table 9.6. In comparison to Nicole, he appeared to be a rather low performer on the pretest. However, upon inspection of his pretest score sheet, an interesting pattern was found. Chris's performance on grammatical items was flawless for about the first two-thirds of the test, until his performance broke down for whatever reasons (e.g., fatigue, frustration) and he missed all points for the final four items. In addition, he was generally incapable of comprehending, and therefore repeating, ungrammatical items throughout the pretest. As the figures in Table 9.6 show, Chris successfully recalled the meaning of ungrammatical sentences only three times (out of nine items), and he was only able to correctly repeat one such sentence. Chris demonstrated marked improvement on the retest. In fact, he matched Nicole's retest scores for grammatical items, and missed very few points for ungrammatical items. Indeed, his score sheet shows that he did not comprehend, and therefore was unable to repeat, the first ungrammatical sentence, but subsequently performed nearly flawlessly, missing only one additional ungrammatical repetition, and half-points for two others in which he corrected a negation but omitted an adverb.

During the teaching phase of the test, it emerged that Chris was in fact capable of understanding the individual words he heard in the ungrammatical sentences, but that the incorrect syntax blocked his comprehension of the meaning of such sentences. For instance, in a lengthy interaction dealing with item #4, *Ta soeur Aimée *lentement a mangé son diner* 'Your sister Aimée slowly ate her dinner' (ungrammatical, adverb), Chris reported hearing *lentement mangait* 'slowly was eating' (imperfective aspect), but that he was confused because he would normally put the adverb *lentement* after the verb. The mediator confirmed that this would be correct and explained to Chris that the sentence he had heard was ungrammatical. Subsequently, Chris was able to make progress, as shown in Excerpt 9.2.

Table 9.6 Chris's pretest and retest scores

Pretest				Retest			
Grammatical		Ungrammatical		Grammatical		Ungrammatical	
Comp	Rep	Comp	Rep	Comp	Rep	Comp	Rep
6	6	3	1	9	8.5	8	6.5

Excerpt 9.2

1.	Mediator:	um:: so let's try:: + six. + m'kay. ((plays audio))
2.	Chris:	my:: s:: brother + Zane (2.0) slowly:: drank his:: ++ coffee. +++ is that *café*? *café au lait*?
3.	Mediator:	yeah. *café* is coffee.=
4.	Chris:	=yeah yeah yeah. but I mean + did I hear that.
5.	Mediator:	yeah.
6.	Chris:	I don't even remember. + what was the (xxx) ((taps on table)) at the very end.
7.	Mediator:	well, + let's try it again.
8.	Chris:	okay.
9.	Mediator:	because it wasn't quite right. ((Plays audio))
10.	Chris:	oh. + my sister + Jane?
11.	Mediator:	mhm,
12.	Chris:	um: + *len::te:men:::* + again, + it should be *a lentement* + *bu sa café*? + *ce matin*?
13.	Mediator:	mhm
14.	Chris:	oh so my sister Jane + uh slowly drank her coffee this morning.
15.	Mediator:	okay. is it- is it + my sister Jeanne slowly drank her coffee?
16.	Chris:	what's the name?
17.	Mediator:	Jeanne.
18.	Chris:	Jeanne?
19.	Mediator:	Jeanne. Like Jeanne d'Arc? + like the feminine form of Jean,=
20.	Chris:	=oh oh oh oh. okay.
21.	Mediator:	listen to it one more time. + and see if it's + um + affirmative or negative. ((plays audio))
22.	Chris:	oh. it's negative. it's (xxx)
23.	Mediator:	so:
24.	Chris:	*n'a bu pas*? + *sa café.* + *aujourd'hui*. + right so::
25.	Mediator:	where does the:: + where does *pas* go.
26.	Chris:	*n'a pas*? i- it- I think it came after + *lentement*. + but it should go:: + *n'a pas lentement*. right? or di- drank her coffee slowly today.
27.	Mediator:	so:: how would it be in French?
28.	Chris:	*Ma soeur Jeanne,* + uh *n'a pas lentement bu* + *sa café* + *aujourd'hui*.
29.	Mediator:	there you go. okay.

Chris clearly experienced some difficulty initially, yet in turn 12 he realized that the adverb was in the wrong place in the stimulus sentence, and he offered a correction. However, he had not comprehended the negative meaning of the utterance, so the mediator moved to support his performance, first by repeating Chris's recall with questioning intonation (turn 15), and then by prompting Chris to listen again to see if the sentence were affirmative or negative (turn 15). This was successful in mediating Chris's comprehension (turn 22), but he still required some assistance in

putting the negative word *pas* 'not' in the correct syntactic slot (turns 24–26), but after some addition assistance he was able to produce a grammatically correct repetition of the sentence (turn 28). During the remainder of the teaching phase, the mediator withdrew assistance as it became less necessary in supporting Chris's comprehension and repetition of stimulus sentences. In other words, following the exchange shown above, Chris made progress toward greater independent control over his performance. As his retest performance shows, Chris was able to effectively close the gap with the pretest high performers, Stephanie and Nicole (see Tables 9.2, 9.3 and 9.4, above), following the teaching phase of the assessment.

Discussion and Conclusion

As we explained earlier, our work is informed by a view of Vygotsky as a dialectical thinker interested in understanding the processes and sets of relations that constitute reality. Stemming from this orientation to science is the genetic method whereby Vygotsky sought to understand psychological abilities not simply through observation of individuals' responses to stimuli but through procedures intended to actively provoke the development of abilities through the introduction of auxiliary – mediating – means. Following this tradition, our project was to devise an assessment tool that included mediation for L2 French teachers to trace learners' developing control over important features of grammar as they progress beyond basic language study. Rather than a conventional assessment administered to learners at a single point in time, we employed the sandwich format of DA in order to document learner responsiveness to mediation according to the changes in their performance evidenced by measures at two points. Following the dialectical logic of Vygotsky's genetic method, the entire three stage procedure, including the teaching phase, is properly understood as the assessment.

Creation of a computerized three-step procedure, in which mediation will eventually be programmed according to insights emerging from analysis of mediator-learner interactions during the teaching phase, will enable teachers to differentiate learners with 'small' as opposed to 'large' ZPDs, following Vygotsky's work with school-age children. A 'small' ZPD, as in the case of a learner whose performance is relatively successful to begin with and who requires little or even no support during the assessment procedure, indicates that the learner is linguistically prepared to succeed. Learners who evidence dramatic improvements between the pretest and the retest may benefit from continued linguistic support as they complete their upper-level coursework (e.g., the 'mini course' mentioned in the introduction to this chapter). A third group of learners, those who experience difficulties with the assessment tasks and who make little or no gains through the procedure, might require significant remediation before they are in a position to succeed. To be sure, no single test can predict future success with 100% accuracy, but we believe such a tool offers a viable response to the challenge faced by many Modern Languages programs at US universities where learners are frequently not fully prepared to make the leap to the study of advanced content (e.g., literature, film, history) in the L2. At the very

least, the kind of DA procedure described above could be used as one part of a broader effort to assess readiness, diagnose challenges, and remediate underperforming students in order to make them ready to participate in and further develop through advanced L2 coursework.

Given the praxis orientation that informs the work reported throughout this book, we recall Haywood and Lidz's (2007) admonition that DA will not realize its full potential to support learner development so long as disconnects remain between the psychologists conducting DA and the classroom teachers tasked with acting on insights obtained through the procedures. In dialectical fashion, the initial project reported here to design the assessment tool actually reoriented our understanding of how teacher and learner needs may be met. Our project led us to understand that, while useful for distinguishing between students who are ready to succeed and those who require remediation at a general level, an assessment tool that exclusively reports sets of scores, such as pretest, retest, and LPS, does not provide the kind of detailed diagnostic information about learner needs that will be easily translated into instruction tailored to individuals. Comparison of pretest and retest scores, and the calculation of an LPS, is indeed valuable for determining how much additional instruction learners may require to advance (i.e., a quantitative measure of readiness to learn). However, in order to glean insight into the underlying sources of difficulty that individuals encounter, information that is likely to be extremely relevant as teachers plan subsequent instruction, a more qualitative approach that focuses on the specifics of DA interactions is required.[2]

For instance, in the two cases we highlighted in this paper, Nicole and Chris, examination of mediator-learner interactions, originally intended as a step toward designing mediation procedures for the assessment tool, actually proved important for revealing the extent of learner understanding of the relevant features of French and the problems they encountered that could become a target for instruction and continued practice. In Nicole's case, limiting analysis to her scores alone reveals the considerable improvement she made following the teaching phase. This information is helpful in revealing that, despite a pretest indicating significant problems controlling French grammar, she was actually close to successful independent performance. Examining her interaction with the mediator brings to light the fact that it was the placement of adverbs in constructions involving negative particles that proved challenging for her. Continued support that specifically targets such constructions is likely necessary for Nicole to reach a level of mastery appropriate to advancing to higher level French courses. Similarly, Chris's performance during

[2] Interaction analysis of face-to-face DA sessions is obviously limited in terms of its scalability as it is rather time consuming and labor intensive to administer such assessments and to subsequently analysis the findings. However, as Qin and van Compernolle (in review) have explored, it is possible to use computerized DA tools, which are easily scalable, and to capture learner interactions with the tools using video screen recording software. Depending on the software used, it is possible to capture (1) what the learner is doing on screen (e.g., cursor movement, clicks, typing) and at the same time (2) audio and video of the learner using the computer's built-in camera and microphone, which enables analysis of such phenomena as reading aloud and private speech.

the instructional phase of DA indicated that he was in fact able to comprehend the meaning of the sentences, despite his low pretest score, and that confusion over the placement of adverbs had created an obstacle that resulted in his initial underperformance. For Chris, simple exercises to practice the placement of adverbs in simple and compound tenses would prepare him to progress to the next level of language study.

Given the importance of not only identifying which learners appear most responsive to instructional intervention but also the specific challenges that individuals will need to overcome in order to be successful, our efforts to build a computerized DA tool for teachers and learners is now concerned with capturing such insights. One possibility is to include an elaborated menu of forms of support that could be made available during the test and from which learners could select; automatically tracking which types of mediation individuals choose (e.g., lexical support, quick tutorials on compound tenses, a review of patterns for adverb placement, etc.) during the test could help to identify areas where future instruction might focus. This information, along with the sets of scores we have discussed, could yield a profile of learner DA performance that would capture both the mediational process as well as learner responsiveness.

Moving beyond the specific details of the project we have described, we wish to conclude by noting that our efforts are in fact part of a larger undertaking that concerns how language and culture are approached by curricula in Modern Language departments at U. S. universities. In this regard, we are reminded of van Compernolle and Williams' (2013) observation that pedagogy, and therefore praxis, includes much more than what occurs in classrooms and in fact reflects institutional and programmatic policies and goals (for related discussions focused outside the U. S., see Chan and Davison, Chap. 7, this volume and East, Chap. 11, this volume). We believe that the kind of collaborative efforts described in our study, including the immediate mediator-learner DA interaction but also the collaboration between DA researchers and faculty who teach literary and cultural studies courses in terms of developing diagnostic tools for assessing advanced-level students' French abilities, is a model for praxis that can be carried out in US university settings. In other words, the research we have reported on here has been driven by a commitment to praxis at the institutional level of a 4-year language program that grants university degrees (i.e., majors and minors) in French and Francophone Studies. The current work is of course a very small step in this direction, but we hope it is one that will foster additional progress toward realizing the potential of DA to promote learner development and individualized instruction across the curriculum.

References

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal, 78*, 465–483.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 39–64). Cambridge: Cambridge University Press.

Davin, K. J., & Donato, R. (2013). Student collaboration and teacher-directed classroom dynamic assessment: A complementary pairing. *Foreign Language Annals, 46*(1), 5–22.

Davydov, V. V., & Radzikhovskii, L. A. (1985). Vygotsky's theory and the activity-oriented approach to psychology. In J. V. Wertsch (Ed.), *Culture, communication and cognition*. New York: Cambridge University Press.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464–491.

Erlam, R. (2009). Elicited oral imitation as a measure of implicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reiders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 65–93). Bristol: Multilingual Matters.

Feuerstein, R., Feuerstein, R. S., & Falik, L. H. (2010). *Beyond smarter: Mediated learning and the brain's capacity for change*. New York: Teachers College, Columbia University.

Frawley, W., & Lantolf, J. P. (1985). Second language discourse: A Vygotskyan perspective. *Applied Linguistics, 6*, 19–44.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice. Clinical and educational applications*. New York: Cambridge University Press.

Holzman, L. (2009). *Vygotsky at work and play*. New York: Routledge.

Ilyenkov, E. (1977). *Dialectical logic*. Moscow: Progress Press.

John-Steiner, V., & Souberman, E. (1978). Afterword. In L. S. Vygotsky (Ed.), *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*. Cambridge, MA: Harvard University Press.

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International, 23*, 112–127.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics, 1*(1), 49–74.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. New York: Routledge.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Leontjev, D. (2016). *Icandoit: The impact of computerized adaptive corrective feedback on L2 English learners* (Unpublished doctoral dissertation). University of Jyväskylä. Jyväskylä, Finland.

Levi, T. (2012). *The effect of dynamic assessment on the performance of students in oral proficiency tests in English as a foreign language* (Unpublished doctoral dissertation). Tel Aviv University. Tel Aviv, Israel.

Lidz, C. S., & Elliott, J. G. (2000). *Dynamic assessment: Prevailing models and applications*. Amsterdam: Elsevier.

Marx, K. (1978). Theses on Feuerbach. In R. C. Tucker (Ed.), *The Marx-Engels reader* (2nd ed., pp. 143–145). New York: W. W. Norton.

Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactive approach to evaluating learning potential*. New York: The Guilford Press.

Novack, G. (1971). *An introduction to the logic of Marxism*. Atlanta: Pathfinder Press.

Ollman, B. (2003). *Dance of the dialectic: Steps in Marx's method*. Urbana-Champaign: University of Illinois Press.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham: Rowman & Littlefield Publishers.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer.

Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly, 43*(3), 471–491.

Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment. *Language Teaching Research, 17*(3), 323–342.

Poehner, M. E., & van Compernolle, R. A. (2016). *From L2 Dynamic assessment to learner receptivity: Consider time as a dimension in classroom assessment validation*. Annual Meeting of the American Association for Applied Linguistics. Orlando, FLA. April.

Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized Dynamic Assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing, 32*(3), 337–357.

Prévost, P. (2009). *The Acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam: John Benjamins.

Qin, T., & van Compernolle, R. A. (in review). *Computerized dynamic assessment of L2 Chinese implicature*. Manuscript submitted for publication.

Siekmann, S., & Charles, W. (2011). Upingakuneng (when they are ready): Dynamic assessment in a third semester Yugtun class. *Assessment in Education: Principles, Policy & Practice, 18*(2), 151–168.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.

Tzuriel, D. (2011). Revealing the effects of cognitive education programmes through dynamic assessment. *Assessment in Education: Principles, Policy & Practice, 18*(2), 113–131.

van Compernolle, R. A. (2016). CA-for-SCT: Dialectics and the analysis of cognition on the ground. *Language and Sociocultural Theory, 3*(2), 173–193.

van Compernolle, R. A., & Williams, L. (2012). Reconceptualizing sociolinguistic competence as mediated action: Identity, meaning-making, agency. *Modern Language Journal, 96*(2), 234–250.

van Compernolle, R. A., & Williams, L. (2013). Editorial: Sociocultural theory and second language pedagogy. *Language Teaching Research, 17*(3), 277–281.

van Compernolle, R. A., & Zhang, H. (2014). Dynamic assessment of elicited imitation: A case analysis of an advanced L2 English speaker. *Language Testing, 31*, 395–412.

van der Veer, J., & Valsiner, J. (1991). *Understanding Vygotsky: A quest for synthesis*. Oxford: Blackwell.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics, 12*(1), 54–73.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky. Volume 1: Problems of general psychology, including the volume thinking and speech* (R. W. Rieber & A. S. Carton, Eds.). New York: Plenum.

Vygotsky, L. S. (1997). *The collected works of L. S. Vygotsky. Volume 3: Problems of the theory and history of psychology* (R. W. Rieber & J. Wollock, Eds.). New York: Plenum.

Vygotsky, L. S. (1998). *The collected works of L. S. Vygotsky. Volume 5: Child psychology* (R. W. Rieber, Ed.). New York: Plenum.

Vygotsky, L. S., & Luria, A. R. (1994). Tool and symbol in child development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 99–174). Oxford: Blackwell.

Zhang, H., & van Compernolle, R. A. (2016). Learning potential and the dynamic assessment of L2 Chinese grammar through elicited imitation. *Language and Sociocultural Theory, 3*, 99–119.

Chapter 10
Reconceptualizing Classroom Dynamic Assessment: Lessons from Teacher Practice

Kristin J. Davin and José David Herazo

Abstract This chapter focuses on three Colombian high school teachers' appropriation of second language (L2) classroom dynamic assessment (DA) to analyze the bidirectional relationship between theory and practice. Situated within Vygotskian sociocultural theory, DA integrates instruction and assessment into unified activity in that teachers assess learners' language ability through a process that involves tailored teaching, or mediation, according to difficulties that learners experience (Lantolf and Poehner, J Appl Linguist 1(1):49–72, 2004). While L2 DA research has documented changes to teacher practice through the implementation of DA, far less work has taken account of insights into DA that may be obtained from teachers and learners. Such a reciprocal relation between conceptual work and practice is a key feature of praxis. In this chapter, we report findings from interviews conducted with three English-as-foreign-language (EFL) teachers who participated in a professional development series on DA (Davin et al., Lang Teach Res 21(5):632–651, 2017). In Phase One of this research, these teachers worked closely with us as they studied and implemented DA procedures in their classrooms. To capture enduring change and offer insight into the further development of DA principles and frameworks, we interviewed each teacher 2 years later. Our findings revealed that teachers recontextualized the concept of DA into a theory of their own practice. Findings suggest implications for DA implementation, but also for the role of situated practice in shaping teachers' appropriation of concepts such as DA.

K. J. Davin (✉)
University of North Carolina at Charlotte, Charlotte, NC, USA
e-mail: kdavin@uncc.edu

J. D. Herazo
Universidad de Córdoba, Montería, Colombia
e-mail: jdherazo@correo.unicordoba.edu.co

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_10

Introduction

Questions of the relationship and interaction between theory and practice and whether or how the two should be integrated to promote research and practice in the field of education have long been debated (Dewey 1902). Poehner and Inbar-Lourie (this volume) situate this debate as reflecting tensions among differing philosophies of science. They explain that some traditions propose a unidirectional flow in which research produces findings that are then 'consumed' by practitioners as they act upon the outcomes of research while other traditions privilege a ground-up development of localized knowledge. As Poehner and Inbar-Lourie point out, both approaches maintain a separation between theory and research, on the one hand, and practice on the other. The orientation they call attention to, and that is elaborated by each of the chapters in this book, positions theory/research and practice in relation to one another, each responding to and informing the other. In the second language (L2) field, one of the theoretical perspectives that advances this relation, referred to as praxis, is Sociocultural Theory (SCT), and it is that theory that frames the work considered in this chapter.

According to Vygotsky (1997), theory should be evaluated by its ability to effect change in the world, and L2 researchers have begun to examine the relevance of the theory to improving teacher professional development (Johnson 2006; Sharkey 2009). Central to this work has been the notion that teacher development occurs when "abstract principles are interwoven with concrete, local experiences" (Smagorinsky et al. 2003, p. 1399). At the same time, L2 researchers have increasingly turned to SCT to elaborate pedagogical frameworks for promoting learner L2 development. Notable here are concept-based instruction and dynamic assessment (DA), both of which hold potential to foster a praxis orientation through partnerships between researchers and teachers working together to bring these frameworks to bear on the realities of L2 classrooms (Davin 2013; Lantolf and Poehner 2011a; Poehner & van Compernolle, this volume; van Compernolle and Henery 2014). This chapter contributes to this literature, and unlike much previous L2 DA work, the perspectives of L2 secondary school teachers are showcased. The inclusion of these perspectives is particularly important given the collaborative nature of praxis, a point that is also brought out strongly in the chapters by Hill and Ducasse and Harding and Brunfaut in this volume.

Approximately 2 years prior to the inception of the present investigation, the authors conducted a professional development workshop series with three L2 teachers in Colombia and three in the United States. During the workshop series, researchers worked in cooperation with the teachers for approximately one academic term to bring DA into their classrooms as a framework to guide their assessment practices. In the original investigation, referred to as *Phase One*, the authors set out to understand how classroom DA implementation changed the discursive practices of the participating teachers (Davin et al. 2017) and the activity settings comprised in their classrooms (Herazo et al. 2019).

The present study represents *Phase Two*, wherein three L2 English teachers were contacted 2 years after the original professional development workshops. This second phase sought to understand in what ways and to what extent the principles and framework of DA continued to affect change in the teachers' behavior after the researchers had departed and there was no 'pressure' to conform to the model put forth in the workshops. While short-term implementation studies are common, too few investigations take this longitudinal approach (Avalos 2011). We sought to understand whether the changes we observed then, and that one of the teachers described as "a challenging change of paradigm", had endured 2 years later. In this regard, we are reminded of Lantolf and Poehner's (2014) argument that the power of theory is determined by its potential for making "a difference in the practical behavior of a community" (p. 27). Conversely, and in the spirit of a reciprocal relation between theory/research and practice that is the focus of this book, the present study also considered how enduring challenges of classroom implementation of DA might feed back into the further development of DA principles.

Background

The focus of the professional development workshop series was DA, a framework emerging from Vygotsky's SCT and Feuerstein's theory of structural cognitive modifiability (Feuerstein et al. 1979; Vygotsky 1978). These theories view cognition as modifiable; that is, able to change through participation in socially mediated activity. Teaching, then, should endeavor not merely to impart knowledge but to make available socially mediated activity necessary to bring about cognitive development. For this to occur, instruction must be tailored to an individual's zone of proximal development (ZPD), defined in the L2 DA literature as the distance between what an individual can accomplish independently and what one can accomplish with more expert assistance (Lantolf and Poehner 2014). In DA, instruction and assessment are unified in that a teacher assesses an individual's emerging abilities, those within the learner's ZPD, and provides instruction that targets the development of those abilities, referred to as mediation. Learners' errors often serve as the impetus for the initiation of L2 DA because they provide a window into learners' partial understanding and their not-yet-fully-developed capacity to control language forms (Davin 2016; Lantolf and Poehner 2011a; Poehner 2008).

Mediation involves the intentional introduction of tools, often by the teacher, to reorganize ongoing activity (Wertsch 2007) and can be either cognitive or metacognitive (Karpov and Haywood 1998). Cognitive mediation provides learners with declarative knowledge about a concept under investigation. In the L2 classroom, linguistic concepts constitute an example of a cognitive tool in that they promote learners' awareness of how the language system works through concept development, allowing them to control their language use (i.e., self-regulate). Metacognitive mediation arises from interpersonal communication and promotes self-regulation (Miller 2011). L2 DA takes place dialogically between a teacher and learners and

typically consists of metacognitive mediation, in the form of prompts, provided by the teacher to learners. The prompts encourage learners to reflect on an utterance, re-examining linguistic concepts to also promote self-regulation of language use (see Excerpt 10.1 for an example).

Most L2 classroom DA approaches share two defining characteristics. The first is that teachers provide mediation through the use of graduated and contingent prompting. Once a learner makes an error, the teacher draws the learner into interaction using verbal prompts that begin as implicit forms of assistance and become increasingly explicit as necessary (Aljaafreh and Lantolf 1994). Prompts do not provide the correct form but allow students the opportunity to self-repair and may include clarification requests, repetition, elicitations, and metalinguistic cues (Lyster and Ranta 1997). Prompts can be scripted in advance or flexible and determined in the moment (Lantolf and Poehner 2004). Scripted approaches offer the advantage of allowing for easy comparisons across learners that highlight the number of prompts required to correct errors. In non-scripted approaches, the mediator uses flexible forms of mediation, calibrated to a learner's emerging needs to promote control of the language system during communication. Direct comparisons of learners are more difficult as the quality of mediation made available to learners may differ markedly, but non-scripted approaches have the advantage of aligning mediation to learner needs in the moment-to-moment dynamics of interaction.

Excerpt 10.1 illustrates a typical interaction during L2 classroom DA and came from the first phase of this investigation (Davin et al. 2017). The exchange took place in a unit in which students were learning present tense verb conjugations to describe their daily routines. In this interaction, the teacher, Martin (a pseudonym), had scripted his prompts in advance although he varied slightly from the scripted prompts when necessary.

Excerpt 10.1

1	S1	When I do arrange my room but my brother disorder*
2	T	Disorganize you mean?
3	S1	(no response)
4	T	Any mistake? Any mistake?
5	S1	(no response)
6	T	No? Listen to this sentence; can you please repeat the sentence aloud?
7	S1	I do arrange my room but my brother disorganize*
8	T	Disorganize… you mean I organize my room but my brother disorganize* it. My question is, what's the mistake?
9	S1	Repeat
10	T	The sentence is: I organize my room or my bedroom but my brother *disorganize** it…Any mistake?
11	S1	Disorganizes
12	T	Very good, third person, because my brother is a…
13	Ss	He

14 T Is a he, very good
15 S1 Third person

Beginning in turn 4, Martin started to provide metacognitive prompts to a learner who failed to add an *-s* to the verb *disorganize* when describing his brother's habit of messing up his room. Martin began with an implicit, or indirect, prompt of asking the learner if he noticed a mistake. In turn 6, he offered a slightly more explicit prompt, asking the learner to repeat his sentence, providing him an opportunity to repair his error. In turn 8, Martin repeated the incorrect portion of the sentence for the learner, again offering a slightly more explicit prompt. In turn 11, the student was able to repair his error, to which the teacher responded with a metalinguistic explanation for the correction, and prompted the learner to elaborate on why the change was necessary. Prompts that are graduated in this way, and contingent upon the learner's needs, engage learners in trying to figure out the reason for their errors, promoting increased control of the language (Rassaei 2014) and learners' sense of agency (Lantolf 2012). Continued learner engagement in figuring out the reasons underlying their errors may ultimately lead to learner autonomy in overseeing their own language use and language learning (van Lier 1996). As in Excerpt 10.1, DA interactions typically conclude when the learner repairs the error or when the teacher has run through the most explicit prompt and must provide an explicit correction and explanation of the error.

The second defining characteristic of L2 DA is that teachers track learners' responsiveness to mediation and employ those data to inform future instruction. A decrease in the quality and frequency of required mediation over time is generally interpreted as an indication that the learner is moving from a reliance on other-regulation (i.e., the expertise made available by the mediator) to self-regulation (i.e., reliance on one's own knowledge and abilities). Put another way, shifts in learner dependence on mediation that is external to him/her signals increasing independence and more autonomous control of the language (Aljaafreh and Lantolf 1994; Lantolf 2012). For example, following Excerpt 10.1, the teacher noted that the learner required four prompts before successful repair. This record indicated to the teacher that the learner still relied on other-regulation and would require more instruction or practice before independently controlling verb conjugations for describing routines, that is before achieving self-regulation in producing this specific linguistic feature of the L2 during oral communication.

Phase One of the Investigation

In a previous article from this data corpus, Davin et al. (2017) examined how four of the six participating teachers appropriated DA immediately following the professional development workshop series. Using data sources that included observations of teaching and stimulated recall sessions, they analyzed the teachers' discursive

practices before and after the workshops to describe any change in those practices deriving from the workshops. Before the workshops, the teachers followed up to students' utterances using almost exclusively recasts, defined as the "teacher's reformulation of all or part of a student's utterance, minus the error" (Lyster and Ranta 1997, p. 46). Following the workshops, however, the teachers began to use more prompts than recasts, giving learners multiple opportunities to reformulate their utterances. These opportunities for reformulation offered learners increased opportunities to reflect on their utterances and progress from other-regulation to self-regulation.

In a second manuscript from the data corpus, Herazo et al. (2019) used a case study to expand the analysis beyond classroom discourse to understand more profound changes to one teacher's instructional and assessment practices following the workshops. The analysis revealed that changes went beyond the teacher's discourse strategies to also include his orientation to assessment and teaching. The teacher, Martin, began to recognize his responses to students' problematic utterances as an opportunity for assessment, adopting a tracking format to assess learners' progress during the lessons as indicated by their responses to his prompting.

Occurring 2 years later, the study described in this chapter sought to deepen our understanding of how the concept of DA 'dialogued' with teachers' concrete local practices; that is, in what ways DA continued to influence the teachers' practices but also how the teachers' understanding of DA as a conceptual framework had changed through their practical experiences. This chapter thus documents whether and how such "praxizing" (Sharkey 2009, p. 126) of DA over a 2-year period might have led to an enduring redefinition of the teachers' theories of instruction and assessment. This investigation expanded beyond just the case of Martin to include two other Colombian teachers of L2 English.

Context and Participants

Phase Two of this investigation was limited to the three Colombian teachers because they had many years of teaching experience, and thus, more flexibility in their instruction and assessment practices. These teachers were originally selected because of their reputation as having committed and open-minded attitudes towards English as a foreign language (EFL) instruction and professional development. Cielo, Margarita, and Martin (pseudonyms) were all experienced teachers of English whose first language was Spanish. They all taught at K-11 public schools in the city of Montería, Colombia, teaching secondary and middle school students, ages 11 through 16. Cielo and Margarita obtained their EFL teaching degree in 2003, from the same university as Martin, and had since taught English in the same school. Cielo held a master's degree and Margarita began one during Phase One of this investigation. They both taught at a Normal school that catered to working class families in which students could opt to become primary school teachers by electing to stay for two additional years of study. After 4 years teaching primary school,

Martin had completed a specialization diploma in English Language Teaching and began teaching older students. His school was located in a poor area of the city. As is common in Colombian public schools (Cárdenas 2006), all three teachers typically had 45 or more mixed-gender students in each class. During Phase One observations, it was not uncommon to observe Cielo, Margarita, and Martin struggling to orchestrate whole class discussions in which students participated in an orderly fashion, respecting each other's turn. Volunteering was not the problem; the challenge was managing 45 students' desire to participate.

The DA workshops occurred at a time of EFL educational reform in Colombia and the two schools were both part of a bilingualism program to improve EFL proficiency in the city. Through the program, which began in 2013, Martin's school increased the amount of EFL instruction from three to six one-hour lessons per week, whereas Cielo and Margarita's school increased from three to eight one-hour lessons weekly. Spurred by globalization and the status of English as an international language, the Colombian Ministry of Education issued standards for foreign language proficiency in 2006 (MEN 2006) and, more recently, curriculum guidelines (Colombia Bilingüe 2016a) to align country-wide EFL curriculum implementation to international L2 standards (cf. Council of Europe 2001). In 2016, a suggested curriculum plan was launched specifying tasks, projects, contents, teaching goals, and methodological and assessment paths for EFL instruction in the country (Colombia Bilingüe 2016b). Of relevance to the present study, the curriculum defined two goals for assessment: assessment *for* learning and assessment *of* learning:

> The former has an educational nature and relates to the follow-up or monitoring that the teacher does of students, which allows understanding the progress they make in their learning. In the meantime, the latter makes reference to the procedures by which the teacher can evaluate the students' performance (Colombia Bilingüe 2016b, p. 24).

These two goals were part of a competence-based orientation in which learners were required to use language for a variety of purposes and situations. Rather than following these guidelines rigidly, Colombian EFL teachers adapted them to suit their teaching needs and school requirements for student grading and promotion.

Workshop Series Description

During Phase One of the investigation, the researchers designed a series of professional development workshops through which to prepare the teachers to implement DA in their classrooms. Teachers were provided with a copy of the second edition text *Dynamic Assessment in the Foreign Language Classroom: A Teacher's Guide* authored by Lantolf and Poehner (2011b). Participants discussed the differences between formative and summative assessment, the characteristics of DA and its origins in SCT, and explored different approaches to DA, discussing the affordances and constraints of each for different classroom contexts. The teachers analyzed

cases of classroom DA use and role-played DA in a way they thought might suit their contextual needs.

Rather than conceiving the workshops as a space where "experts" teach "non-experts", researchers and teachers approached the workshops as a community of study. All three teachers showed an enthusiastic but critical attitude towards DA during the workshops, and questioned whether DA could be implemented in their large classes. Because teacher satisfaction with professional development is closely tied to their ability to mold instructional processes to their own needs and expectations (Nir and Bogler 2008), we continually emphasized the phrase "in your context", expressing our desire that the teachers implement DA in a way most conducive to their own needs. As Canagarajah (2006) writes,

> Teachers in different communities have to devise curricula and pedagogies that have local relevance. Teaching materials have to accommodate the values and needs of diverse settings, with sufficient complexity granted to local knowledge. Curriculum change cannot involve the top-down imposition of expertise from outside the community but should be a ground up construction taking into account indigenous resources and knowledge, with a sense of partnership between local and outside experts (p. 20).

Participants were asked to implement DA over a series of three lessons, each of which was video and audio recorded. They reflected on each instance of DA implementation and received additional feedback from the instructors.

Data Collection and Analysis

Data collection took place approximately 2 years following the professional workshop series when the researchers conducted follow-up interviews. Interviews took place at each teacher's school, lasted approximately 1 h, and were conducted in Spanish. Interview questions probed teachers' philosophy of classroom assessment and instruction and reflections on how their instruction and assessment practices had changed, if at all, since they participated in the DA workshops.

While the data corpus included sources from both Phases One and Two of the research study, the data set analyzed for the present study consisted solely of the interviews from Phase Two. Because the authors sought to understand how the teachers' beliefs and descriptions of their practice aligned with the principles of DA, the authors used theoretical thematic analysis (Braun and Clarke 2006). Following the orthographic transcription of all interviews, the authors repeatedly read through the transcripts, searched for meanings, and discussed emerging patterns. Based on these emerging patterns, the research questions, and the framework and principles of DA, the researchers generated a theory-driven coding scheme (Braun and Clarke 2006). Parent codes (bolded) and accompanying child codes are shown in Table 10.1.

All interviews were coded with this scheme using data analysis software. Codes were sorted into potential themes and the researchers met to discuss how different codes should be combined into overarching themes. Four overarching themes were chosen: (a) *Purpose of assessment*, (b) *Purpose of instruction*, (c) *Change to*

Table 10.1 Coding scheme

Appropriation of concept of DA		Definition
	Concept of graduation	Scaffolding of prompts
	Concept of student agency	Giving agency and control to learners
	Dynamic not static	Providing instruction during assessment
	Tracking students' learning	Tracking students' responses to prompting
	Focus on development	Going beyond error to understand its cause
	Lack of appropriation	Counterexamples to graduation, tracking, etc.
Extent of change		
	Change to curriculum	Changes beyond classroom (i.e., school level)
	Change to instruction	Changes in classroom
	Change to others	Changes to other participants in the setting (i.e., students, colleagues)
Role of situated practice in appropriation		How context influences teacher learning/development/ appropriation
Teachers' current conceptualizations		Teachers' beliefs about assessment

students, and (d) *Lingering challenges*. Data excerpts were collated for each code, and then further parsed into sub themes with particularly representative extracts selected for inclusion for each theme. Selected excerpts were translated to English for inclusion in this chapter.

Findings

Analysis revealed four major themes resulting from the way that Martin, Cielo, and Margarita's local practices entered into a 2-year-long dialogue with the concept of DA. During that period, the three teachers: (1) redefined the purpose of their assessment, (2) developed a concern for depth over breadth in their teaching, (3) became aware of the potential of DA to change students' perceptions of and role in assessment, and (4) began looking for ways to cope with the challenges inherent to bringing DA into their concrete experiences. These themes constitute the building blocks of the three teachers' evolving and local theory of DA.

Change in Purpose of Assessment: Process Over Product

Use of DA in their lessons led Cielo, Margarita, and Martin to rethink the overall purpose of their assessment practices. In line with DA's concern for the process rather than the products of learning, they began to conceptualize assessment as a

way to learn about students' progress (i.e., assessment) and to promote students' development (i.e., teaching) at the same time. An important feature of this dual function of DA, as Martin explained, concerned helping students to realize their own progress:

> I found [DA to be] very positive, well, we are assessing but, what do we do with the assessment? Will the assessment serve just for me to have a judgment or will it help my student grow during the process? In fact, my colleagues and I have written that in our year-long curriculum plan, specifically where it conceptualizes the evaluation strategy and method. We wrote that assessment functions as part of students' learning process…if we want our students to take something away from the assessment and not only the grade…what did [students] learn in the process? In what [way] did it help [them]? In what [way] did it contribute to [their] overall learning process?

Martin's words disclosed his orientation to assessment as process, as an opportunity for students to learn from assessment itself and not only as a way to judge student progress once learning has occurred. Martin reported that this redefinition of the purpose of assessment became formalized through school curriculum documents authored by him and other teachers, helping them to reshape their assessment philosophy. What is more, these conceptualizations helped Martin navigate the proposed Colombian EFL national curriculum. As he put it,

> When I read the methodological principles for assessment in the new curriculum, I remembered immediately the two ways to intervene in the process of dynamic assessment, [when I read] evaluation for learning, I thought, this has to be it, I mean, I assess so that students learn, and I said, 'well, definitely I'm on the right track', and you kind of confirm what you had already learned… during the [DA] project two years ago

The comment revealed Martin's concrete and personal experience with DA. What he had "already learned" became part of the theoretical basis with which he approached the new concepts of assessment proposed by the suggested curriculum.

Attention to process also became a turning point in Margarita's and Cielo's assessment practices as their conceptualizations evolved from a summative to a formative orientation. Margarita defined assessment as "rather a process for students to continuously inform their own learning", whereas Cielo oriented to it as a way to make sure that most students had learned before moving on to a new instructional unit. In both cases, it was clear that they viewed assessment as a means to strengthen the learning process.

Concern for assessment as part of students' learning process also led teachers to incorporate new strategies in their teaching. Cielo, for example, began to use DA for diagnosing students' background knowledge when beginning a new teaching unit:

> well, when beginning a new topic I use it to explore students' previous knowledge, I use it a lot… because some have some background of what we are studying, but others don't, so it is a way to bring students to the same level.

Cielo's comment suggested that, in addition to using DA for diagnosis, she also used it to ensure students developed a shared background knowledge before teaching a new topic.

For his part, Martin implemented what he called "a support plan" to help students learn both during and at the end of each academic term. The plan involved using non-scripted prompting DA during lessons to help learners overcome their misunderstandings and to get information regarding "how the [learning] process is improving from lesson to lesson, for all students in general and for those students who are lagging behind." The plan involved tutorial meetings with students at the end of each term, remedial sessions in which Martin focused on a few students whom he had determined were not keeping pace in the learning process. He reported using scripted-prompting DA during tutorials because he had already identified a problem and hence was able to easily plan his prompting in advance. He commented, "I think it works better planning the prompts in advance, and you say, 'well, I'm going do this and that, if that does not work, then I do this other thing.'"

Although Cielo, Margarita, and Martin's implementation of DA privileged the process over the products of learning, it was intriguing how they each envisioned the purpose of DA differently. While for Cielo the main purpose of DA was diagnostic, it served functions of both diagnosis and remediation for Martin. Interestingly, Martin used non-scripted prompting DA for his day-to-day diagnosis and instruction, whereas he used scripted-prompting DA during his end of semester tutorials designed to support low achievers. Thus, Martin appears to have adopted the notion of approaching interactions with learners as opportunities to mediate their engagement with the language and with instructional tasks and generalized it beyond activities planned as assessments, integrating it more generally into his classroom practice. Margarita, in contrast, used DA as a way to inform students about their own learning and, as we shall see later, aimed at promoting learner understanding of English as a system of form-meaning relations. These three cases show how the abstract principles of DA became interwoven with teachers' local and concrete experiences, resulting in a contextualized (re-)theorization of the uses and benefits of DA in their classrooms.

Purpose of Instruction: Depth Over Breadth

Because instruction and assessment exist in dialectical relation in DA, it was not surprising that a transformation of teachers' beliefs regarding the purpose of assessment was accompanied by a change in their beliefs about instruction. A common theme across all three teachers was that student learning should be deep rather than superficially span a wide variety of topics. Cielo and Martin both explained that they had begun covering fewer units in a year. Martin reported that students needed a "solid base", and that covering all units solely for the sake of getting through the curriculum was "wasting time". Although not stated explicitly, Martin's new conceptualization of the purpose of instruction aligned with DA's focus on fostering development rather than focusing solely on learning (Lantolf and Poehner 2014). In SCT, learning may entail the mastery of skills or the acquisition of knowledge while development refers to the re-organization or transformation of thinking and the

consequences this has for how one sees and acts in the world (Vygotsky 1986; Kozulin 1998). Rather than memorizing a wide array of English vocabulary and conjugations, Martin wanted his students to develop deep and flexible knowledge that would transfer to future use.

Cielo reported a similar change in practice when asked how her teaching strategies had changed, if at all. She stated, "I do not want the student to pass from one course to another with doubts from the previous [course], if I cannot move on to the next topic, I do not, I stay there until at least 80 percent of the class is ready to move on". When asked how she decided when it was time to move on, Cielo responded,

> I cannot be 100% sure that students overcome their difficulties, but with the results of the evaluation of each subject of the test—and not exactly of the test but of the whole evaluation process—with these results I look more or less where we have had failures, and I can decide whether or not to move to the next [unit].

Like Martin, Cielo's comments highlighted her conceptualization of assessment and instruction as intertwined. Her comment revealed that she had abandoned an approach in which all decisions were made on the basis of one summative assessment to instead consider students' performance on an "evaluation process" that included both a test and DAs. Changing from the perspective that assessment results served only to assign grades, Cielo used DA for data-based decision making, determining students' areas of continued need. Rather than understanding the curriculum as material that must be covered in a fixed amount of time followed by an assessment to measure students' completed learning, both Cielo and Martin adopted a philosophy more conducive for language development. They both recognized that language instruction must build on prior learning and that depth of knowledge was more important than breadth.

The theme of depth rather than breadth was also prominent in Margarita's interview, however her philosophy took the purpose of instruction beyond the teaching of English to the development of metacognition. A critique of existing models of L2 classroom DA is that such models take a loose interpretation of the ZPD (Chaiklin 2003; Song and Kellogg 2011) and are learning-focused rather than development-focused (Davin 2016). That is, that these models are more similar to scaffolding (Herazo et al. 2019) because they measure the mediation required by learners to get through a task rather than diagnosing and promoting the development of new semiotic resources that alter individuals' activity in the world. Margarita's conceptualization of the purpose of instruction, and the role of DA within it, was more focused on development than that of Cielo and Martin. She stated,

> It's like not letting go of the opportunity that they can learn a little more, and not only English, which is what I love because it is a cognitive process. They are not only learning English but learning to look beyond. It is when you say, 'And why did you write like this? And why did you use this?' They are given the opportunity to think…

When asked to elaborate upon what she meant by "learning to look beyond", Margarita stated,

> to explore the why of things. The why I am using this, why this can go, why I can answer this way and that pleases me because it develops their thinking. It is not only limited

> learning, what I mean is that many times English is extremely limited in that students learn a little English and nothing else. So I like that English serves to help them learn other things about their lives. So what I do is give them the ability to understand.

Margarita's statement aligned with the perspective in SCT that learning a new language gives an individual new semiotic tools through which to make meaning and act in the world (Lantolf and Thorne 2006). Rather than focusing solely on targeting learners' development of English, Margarita wanted to delve deeper to help learners understand the connection of form and meaning.

Change to Students

Increased Autonomy

The teachers reported that their new philosophies of assessment and instruction influenced students' perceptions of and participation in class, increasing student autonomy. Margarita stated that students came "to be more aware" as a result of her use of DA, referring to their improving skills of self-correction. She suggested that an increase in awareness indicated growth in learners, evidence that "you know your teaching is working when students are able to go back and revise". Similarly, Martin referred to what he called "a change in culture" in his classroom. He said that prior to implementing DA, students often said to him, "Teacher please, teacher please, review me, correct me, you correct me". However, learners became more autonomous as Martin began to implement DA. He stated that previously the culture "was that the teacher corrects me and I know what was wrong and I apply the corrections." However, one goal of his DA use was to make students "more critical of what they are doing, to take their time". He stated,

> Now we have the philosophy of saying, before presenting something, be convinced that you have used all the tools that you have, the knowledge that you have so that when the teacher reviews it, your work is good or when you participate that it is good, then take more time to look at what is created, whether oral or textual, whether it be reading comprehension, whatever, and do it well.

Martin's statement illustrated his desire to increase learners' self-regulated learning. Rather than always relying on him for the answers, Martin's new philosophy of instruction and assessment was designed to increase learners' metacognitive skills and thus promote agency (van Lier 2008).

While Martin's use of DA supported learners' self-regulation, Cielo's appropriation of DA corresponded to her development of a system of shared responsibility in the classroom. Describing the system that she implemented, Cielo stated,

> The students who learned English more easily, who performed better, sponsored one or two students during that period, and gave encouragement to the sponsored students. If their sponsored students overcame the difficulties and their grade was above 3.80, the 'padrinos' (godparents) received a 5 for the period regardless of their other grades… as a result, students exceeded in quantitative terms, many students achieved the goal. I measure the

> qualitative with a survey, I ask each 'padrino' for a report of the work of each sponsored student who he worked with, which tells me whether the student was attentive to the suggestions of the 'padrino'.

Cielo went on to explain that this system worked because students enjoyed working in small groups and because they could choose their own "padrinos". Cielo's system aligned with research into the usefulness of accompanying DA with small group work (Davin and Donato 2013). The interaction between learners during small group work has been shown to be a necessary complement to teacher-led instruction, providing a distinct form of mediation that can be more relevant to learners (Davin and Donato 2013; Guk and Kellogg 2007; Toth 2008).

Decreased Anxiety

Two of the three teachers, Margarita and Cielo, reported that the change in their assessment and instructional practices resulted in decreased student anxiety and stress. Cielo reported that students often did not realize when they were being evaluated in her class. She described that she had timid students who stuck together, but made comments to her such as, "I like my class, teacher". She said that these students were not "the ones strongest at English", but in spite of that, they expressed their comfort in her class, stating "Profe, in your class I feel very good". Cielo attributed these comments to the relaxed atmosphere of her class, to the lack of stress due to the formative nature of assessment in which students did not realize they were being evaluated.

Margarita also noted a change in her students. When asked how her grading procedures had changed, Margarita connected that change to a change in her students. She reported that "they were learning to not be afraid of evaluation". She stated that students feared evaluation because they were afraid of being graded, but that DA was a "less rigid" form of evaluation that students did not realize was evaluative. She described how the decrease in fear of evaluation made the class more enjoyable for them, stating "when students do not feel the pressure of being evaluated, that they enjoy learning more". Interestingly, Margarita was at first worried by the prospect of a classroom environment not motivated by fear. She said,

> They were learning to not be afraid of evaluation as such, of the evaluation process… That is a concept that I had as a teacher, that if the student is not afraid of the 1.0 that the teacher can give, that they will not learn the content. And I discovered in my process with the students that they could be very responsible, that they were not afraid to be evaluated. They did their thing, they did their thing, they did everything they had to do, but they did not panic that they were going to fail.

Her comment suggested a dramatic change to her classroom climate, a change in her beliefs and practices that resulted in lower student anxiety. By implementing DA, Margarita came to the realization that fear of evaluation was not a necessary component of L2 learning.

Lingering Challenges

During interviews, the authors asked the teachers to describe lingering doubts and challenges that they continued to experience in classroom DA implementation. These were probed to understand what insights the teachers might have for further development of DA principles and frameworks. The three most prominent challenges that emerged were the use of DA in large groups, time constraints, and determining appropriate forms of mediation. The first two sub themes were intertwined and are presented together in this section followed by the third theme.

Class Size and Time

All three teachers spoke about time constraints related to DA implementation. The teachers were accustomed to relying on recasts in short Initiation-Response-Feedback interactions in which the teacher asked a question, a student responded, and the teacher recast the utterance in its correct form. In such interactions, the teacher maintained control over the interaction and was able to keep it brief, unlike the more extended interactions that unfolded when the teacher provided graduated prompts. When asked about the challenges of DA, Cielo stated, "It's still time management, it takes a lot of time". Cielo described that time was a challenge due to the size of her classes. She had approximately 45 students in her class and she stated that even when they worked in small groups, "the groups were numerous" and she struggled to provide mediation to each one. She stated,

> If the lesson is not unfolding as I expected it would in the allotted amount of time, if the students have a lot of questions, if it seems very difficult to students, I stay there and I try to find other ways for them to understand me.

Because it was difficult to anticipate students' questions and determine appropriate forms of mediation, Cielo often ran short on time. In light of her reported shift from trying to cover all topics designated for a course to instead making sure that "80 percent of the class is ready to move on", Cielo's comment was not surprising. However, her concern for time suggested that she may not have fully embraced the focus on development emphasized in DA, in which spending additional time fine tuning mediation to learners' needs can be seen as an investment in students' learning that may deepen their understanding.

Margarita also reflected on the difficulty of implementing DA with so many students in constrained amounts of time. When asked about challenges to implementation, she responded, "as I have said from the beginning, time is a challenge as is the number of students". She explained that the difficulty was the varying needs of learners, stating, "you cannot do the same with everyone, you would like for all students to come to you with questions". Margarita's statement aligned with existing research that students have varying ZPDs and thus require different forms of mediation (Brown and Ferrara 1985; Davin 2016). To elaborate, she gave an exam-

ple from a recent unit, describing that, "students come to you with a piece of writing that you know has twenty errors and you would like [to mediate], would like to use [DA] with all students". Reflecting on a recent unit on romanticism in which students wrote poems, she lamented, "if you saw how many errors the first drafts had!" Elaborating, she explained,

> the first five that come up to you and want to do DA and [you ask] 'why these words?' and 'which do you believe would work better here?' but when you continue mediating, my God! It goes on forever and then the line waiting for you to help with revisions! [laughing] How can I have the peace and patience to use DA with a large group?

In addition to indicating the varying needs of learners, Margarita's comments also suggested that she may have struggled to determine which errors to mediate, a difficulty that other participants had experienced in Phase One of this research (Davin et al. 2017). Teachers implementing classroom DA are faced with the complex decision of determining when to provide mediation and when such mediation might instead impede the flow of the classroom or students' attempts at communication.

Like Cielo and Margarita, Martin reported that it was challenging to use DA in such large groups, reporting that he had groups as large as 51 students in one class. Like Margarita's students, his had varying difficulties, as illustrated by his statement that "the need that he has, it is not the same need that the other has". Martin noted "the speed of time". He stated that it was acceptable to provide mediation when an individual's need was the same as that of many others in the group; however, he explained that when it was not, he tried "to do it fast", referring to providing mediation, so that he could "continue to advance the lesson". Martin went on to describe that DA was best used in small groups when you could identify four or five students that required similar mediation. Otherwise, he stated that DA should only be used in the whole group when more than half of the students had similar misunderstandings.

Appropriate Forms of Mediation

As alluded to in their statements above, both Cielo and Martin discussed the challenge of determining appropriate forms of mediation for learners. Cielo's statements that she had "doubts about how to do it", "do[es] not know where to begin", and "stay[s] there and [tries] to find other ways for them to understand" indicated her struggles with mediating students. Cielo's comments alluded to a continuing struggle with contingency in interaction, as did Martin's, who spoke about the challenge more explicitly. When asked about challenges to classroom DA implementation, he responded that, "Sometimes one becomes frustrated because one begins an interaction with [a student] and gives him a tool, gives him a tool but cannot figure out what the student understands, and ends up giving him the answer". His comment suggested that he occasionally terminated an interaction by providing an answer because he was unable to find an appropriate form of mediation to assist learners. The challenge, he described, was finding "the precise way for the student to arrive at the answer"; that is, the appropriate means of mediation.

Discussion and Conclusion

Theory Shapes Practice

The challenge of changing teachers' practice is well documented and a large body of literature exists on how to effect such change (see Johnson and Golombek 2016; Inbar-Lourie & Levi, this volume). Research suggests the need for delayed data collection measuring effectiveness because changes may be fleeting, or as in this case, teachers may appropriate some aspects of the intervention but not others (Avalos 2011). In the present study, analysis of the teacher interviews identified many enduring changes that represented significant shifts in their orientation to and philosophies of teaching. All three teachers moved away from a conceptualization of assessment as a summative, product-focused, one-time event designed to measure students' learning to assign a grade. Instead, they each adopted a conceptualization of assessment as a formative, process-oriented, iterative experience designed to highlight individuals' areas of need so that instruction could be designed to remediate those identified areas. Consequently, all three teachers reported a shift to a perspective of instruction as a way to build a solid understanding of the English language, rather than as a set of units in a curriculum that must be covered. A teacher-centered philosophy in which students had one opportunity to learn what was taught was replaced by a learner-centered philosophy in which decisions about when to proceed to new content were made based on student data. These resulting conceptualizations aligned with the unity of assessment and instruction in DA, which Poehner and Lantolf (2005) refer to as two sides of one coin in that "true assessment is not possible unless it entails instruction and vice versa" (p. 261).

The teachers' appropriation of DA functioned as a conceptual tool that caused a large ripple effect. The initial focus of the workshop series on two components of DA that seemed manageable and learnable, including graduated prompting and tracking students' learning, had more far-reaching influence than anticipated. Following their practice with these two concrete aspects of L2 DA, the teachers' philosophies of assessment and instruction changed and influenced school curriculum as well as learners' enjoyment of English classes. We concur with Johnson (2009) when she notes that it is the development of concepts that may lead to "substantive and significant changes" (p. 4) in the way teachers think of and engage in practice.

Practice Shapes Theory

The teachers' experiences offered insight into aspects of DA principles and practice that might deserve further attention. To be sure, the concept of DA informed the way teachers thought about and approached their practice. They began to focus their

attention on how assessment itself might help to promote development, recognizing the importance of focusing on the process of learning over the product, and on striving for depth rather than breadth of material. Cielo and Martin provided concrete examples of implementation, creatively adopting DA for specific purposes in specific contexts—Cielo for diagnosis at the beginning of a unit and Martin for diagnosis and remediation in end-of-semester tutorials for low achievers. However, each of the teachers experienced tension in their appropriation of DA because of the amount of time that they spent providing learners with mediation. Because interactions were longer, teachers were unable to cover the quantity of material typically designated for their course, causing them anxiety. The issue of DA implementation in large group contexts would benefit from additional empirical and theoretical development.

A second area of concern expressed by the teachers was the challenge of determining appropriate forms of mediation. L2 DA has generally focused on metacognitive mediation that encourages learners to reconsider and reformulate utterances that contain errors. A learner looks inward, consults existing knowledge, analyzes his or her utterance, and makes adjustments. However, cases do occur when prompts and leading questions are insufficient to lead learners to self-correct and instead a teacher may need to determine whether, in that moment, more explicit instruction is appropriate. Such a shift from diagnosing learner current and emerging understanding and abilities to actively promoting new ways of thinking might effectively be pursued through the use of concepts to reshape learners' understanding and use of language (Davin 2016). Of course, such instruction requires that teachers themselves must develop appropriate metalinguistic knowledge of how language works and how form maps onto meaning. To that end, professional development focused on DA might benefit from being paired with professional development on frameworks such as mediated development (Poehner and Infante 2015), concept-based instruction (Negueruela and Lantolf 2006), or systemic functional linguistics (Halliday and Matthiessen 2014).

Conclusion

The evidence of change gathered during the interviews with these three teachers suggested the power of the framework and principles of DA to make a difference in the practical behavior of these teachers' classrooms and schools. Not surprisingly, the changes were not those that the researchers had previously hypothesized, but they surpassed such expectations and had far-reaching consequences for their philosophies of teaching. Although the generalizations that can be made from the present investigation are limited due to the small number of participants, our findings suggest that professional development on DA may result in profound changes to the practice of teachers who hold traditional, product-focused, summative views of assessment.

As Cielo concluded, "Knowing [DA] has helped me to improve teaching, it is not perfect but it has improved the way I do it. I make it more complete, I use prompting more with students, I spend more time."

References

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal, 78*(4), 465–483.

Avalos, B. (2011). Teacher professional development in teaching and eacher education over ten years. *Teaching and Teacher Education, 27*(1), 10–20.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77–101.

Brown, A., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 273–305). Cambridge: Cambridge University Press.

Canagarajah, S. (2006). TESOL at forty: What are the issues? *TESOL Quarterly, 40*(1), 9–34.

Cárdenas, M. L. (2006). *Bilingual Colombia: Are we ready for it? What is needed?* Paper presented at the 19th annual English Australia education conference, Perth, Australia.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 39–63). Cambridge: Cambridge University Press.

Colombia Bilingüe. (2016a). *Pedagogical principles and guidelines: Suggested English curriculum*. Bogotá: Ministerio de Educación Nacional.

Colombia Bilingüe. (2016b). *Suggested curriculum structure*. Bogotá: Ministerio de Educación Nacional.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.

Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom. *Language Teaching Research, 17*(3), 303–322.

Davin, K. J. (2016). Classroom dynamic assessment: A critical examination of constructs and practices. *Modern Language Journal, 100*(4), 813–829.

Davin, K. J., & Donato, R. (2013). Student collaboration and teacher-directed classroom dynamic assessment: A complementary pairing. *Foreign Language Annals, 46*(1), 5–22.

Davin, K. J., Herazo, J. D., & Sagre, A. M. (2017). Learning to mediate: Teacher appropriation of dynamic assessment. *Language Teaching Research, 21*(5), 632–651.

Dewey, J. (1902). *The child and the curriculum and the school and society*. Chicago: Phoenix Books.

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore: University Park Press.

Guk, I., & Kellogg, D. (2007). The ZPD and whole class teaching: Teacher-led and student-led interactional mediation of tasks. *Language Teaching Research, 11*(3), 281–299.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar*. New York: Routledge.

Herazo, J. D., Davin, K. J., & Sagre, A. M. (2019). Corrective feedback and dynamic assessment: Comparisons from an activity theory perspective. *Modern Language Journal, 103*(2), 443–458.

Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly, 40*(1), 235–257.

Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. New York: Routledge.

Johnson, K. E., & Golombek, P. R. (2016). *Mindful l2 teacher education: A sociocultural perspective on cultivating teachers' professional development*. New York: Routledge.

Karpov, Y. V., & Haywood, C. (1998). Two ways to elaborate Vygotsky's concept of mediaton. *American Psychologist, 53*, 27–36.

Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*. Cambridge, MA: Harvard University Press.

Lantolf, J. P. (2012). Sociocultural theory: A dialectical approach to L2 research. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 57–72). New York: Routledge.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics, 1*(1), 49–72.

Lantolf, J. P., & Poehner, M. E. (2011a). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11–33.

Lantolf, J. P., & Poehner, M. E. (2011b). *Dynamic assessment in the foreign language classroom: A teacher's guide* (2nd ed.). University Park: CALPER.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. New York: Routledge.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37–66.

MEN. (2006). *Estándares básicos de competencias en lenguas extranjeras: Inglés*. Bogotá: Ministerio de Educación Nacional.

Miller, R. (2011). *Vygotsky in perspective*. Cambridge: Cambridge University Press.

Negueruela, E., & Lantolf, J. P. (2006). Concept-based instruction and the acquisition of L2 Spanish. In R. Salaberry & B. A. Lafford (Eds.), *The art of teaching Spanish: Second language acquisition from research to praxis* (pp. 79–102). Washington, DC: Georgetown University Press.

Nir, A. E., & Bogler, R. (2008). The antecedents of teacher satisfaction with professional development programs. *Teaching and Teacher Education, 24*, 377–386.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer.

Poehner, M. E., & Infante, P. (2015). Mediated development: Inter-psychological activity for L2 education. *Language and Sociocultural Theory, 2*(2), 161–183.

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*(3), 233–265.

Rassaei, E. (2014). Scaffolded feedback, recasts, and L2 development: A sociocultural perspective. *Modern Language Journal, 98*(1), 417–431.

Sharkey, J. (2009). Can we praxize second language teacher education? An invitation to join a collective, collaborative challenge. *Ïkala Revista De Lenguaje Y Cultura, 14*(22), 125–150.

Smagorinsky, P., Cook, L. S., & Johnson, T. S. (2003). The twisting path of concept development in learning to teach. *Teachers College Record, 105*(8), 1399–1436.

Song, S., & Kellogg, D. (2011). Word meaning as a palimpsest: A defense of sociocultural theory. *Modern Language Journal, 95*(4), 589–604.

Toth, P. D. (2008, June). Teacher- and learner-led discourse in task-based grammar instruction: Providing procedural assistance for L2 morphosyntactic development. *Language Learning, 58*, 237–283.

van Compernolle, R. A., & Henery, A. (2014). Instructed concept appropriation and L2 pragmatic development in the classroom. *Language Learning, 64*(3), 549–578.

van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and authenticity*. New York: Longman.

van Lier, L. (2008). Agency in the classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 163–186). London/Oakville: Equinox.
Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: The MIT Press.
Vygotsky, L. S. (1997). *The collected works of L.S. Vygotsky* (M. J. Hall (Ed.), Vol. 4). New York: Plenum Press.
Wertsch, J. V. (2007). Mediation. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 178–192). New York: Cambridge University Press.

Part IV
Researcher-Teacher Action Challenging Assessment Culture

Chapter 11
Addressing the Possibilities and Limitations of Implementing a New Classroom-Based Assessment of Oral Proficiency

Martin East

Abstract In the context of recent curriculum and assessment reforms in New Zealand's school system, the assessment of foreign language (FL) students' oral proficiency has undergone significant transformation. In line with theoretical arguments recognising both the *learning* benefits of peer-to-peer interaction and the importance of on-going assessments to support teaching and learning, a summative teacher-led interview test (*converse*) has been replaced by a model whereby students collect evidence of several learner-focused peer-to-peer interactions in the context of teaching and learning programmes (*interact*). However, putting theoretical principles into practice in the assessment has proven to be more challenging than anticipated. Findings of a recent research project that investigated stakeholder perspectives (teachers and students) on *interact* in comparison with *converse* have illuminated how additional conceptual understandings regarding both interaction and assessment for learning are needed. This chapter briefly outlines the theoretical rationales for *interact* from both pedagogical and assessment perspectives. It summarises the key findings of the research. It then presents the case of the introduction of *interact* in one school, drawing on data from an interview with the teacher and a focus group with his students (n = 7). The case is presented as an example of how one school is addressing the possibilities and limitations of *interact* in practice. Conclusions are drawn about aspects of *interact* that require revision and development.

M. East (✉)
School of Cultures, Languages and Linguistics, Faculty of Arts, The University of Auckland, Auckland, New Zealand
e-mail: m.east@auckland.ac.nz

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_11

Introduction

In the New Zealand schools sector, the turn of the century has witnessed an increasing emphasis away from a behaviourist, knowledge-based, teacher-led pedagogical approach towards approaches that are more constructivist, learner-centred and experiential. This has been in evidence in substantial assessment and curriculum reforms. As I explain in further detail below, there have been three significant milestones along the path to reform: the introduction (2002) of a new standards-based school assessment system that replaced a traditional summative-examination dominated framework; the publication (2007) of a revised national curriculum for schools that continued a move to encourage a constructivist pedagogy; and the introduction (2011) of modified achievement standards[1] designed to be more aligned with the expectations of the revised curriculum.

The last two decades have therefore been a period of momentous change for teachers in schools, with significant implications for both assessment and teaching practices. In my role as a teacher educator working with pre-service teachers who will go on to teach foreign languages (FL) in schools, the substantial shifts heralded by these reforms have been an impetus to interconnect theory, research and practice in an on-going, reciprocal way. This has included researching the impacts of these reforms on practising teachers' beliefs and pedagogical choices (East 2012, 2016); utilising findings with the beginning teachers I have worked with; researching these beginning teachers' engagement with praxis; and in turn reporting the findings of that engagement (East 2014, 2017, 2019). The enactment of praxis in assessment, as illuminated in this research, implies a fundamental epistemological change in that knowledge construction and refinement occur through collaborations involving researchers and practitioners, and as theoretical proposals are brought into contact with the demands of local contexts, leading to collaborated exploration and practice by experts and practitioners (for more on praxis, see introduction Chap. 1 to this volume by Poehner and Inbar-Lourie).

This chapter documents a component of this cyclical, praxis-oriented research agenda. In the broader context of one teacher's participation in this research (see, e.g., East 2017), this chapter presents findings into this teacher's engagement with the most recent assessment reforms, drawing on data from an interview with him and a focus group with his students (n = 7). The case is presented as an example of how this teacher is addressing the possibilities and limitations of the reforms in practice, with specific regard to the assessment of FL students' oral proficiency, and a new achievement standard known as *interact*. Following this reciprocal relation between theory/research and practice, insights from this case are discussed with regard to their implications for potential revision of the *interact* achievement standard.

[1] Achievement standards are documents that outline what a student should know and be able to do in a discrete named skill at three levels (grades) of performance – *achieved (A), achieved with merit (M),* and *achieved with excellence (E)*. Students who do not reach the standard at a given measurement point receive a *not achieved* outcome. A summary Record of Achievement lists all standards that have been achieved, and when these were achieved.

Background

Up to the turn of the century, New Zealand's high-stakes assessment system for schools was dominated by summative external examinations available at two levels: School Certificate, taken by students at the end of school Year 11 (15+ years of age), and the University Entrance, Bursaries and Scholarships examination (Bursary), taken at the end of Year 13 (17+ years of age). In a move that was hailed as a "revolution of assessment" (Hattie 2009, p. 259), a new assessment system – the National Certificate of Educational Achievement (NCEA) – was introduced at three levels of increasing breadth and depth (1, 2 and 3). Level 1 replaced School Certificate from 2002, and level 3 replaced Bursary from 2004 (level 2 was introduced in 2003). The most significant shift in the new system at all levels was a strong emphasis on classroom-based assessments, created and marked by local teachers (although summative external examinations did remain in place). There was a concurrent shift in emphasis from the assessment *of* learning to 'assessment *for* learning' (Assessment Reform Group 1999, 2002).

Hattie (2009) went on to articulate several perceived benefits of what the new system, and, in particular, the assessment for learning opportunities presented by internal assessments, were conceptualised to achieve:

- a standards-based approach (candidates are measured against clearly articulated standards or criteria);
- a constructive alignment between learning and outcomes (learning goals are clearly aligned to assessment goals);
- opportunities for peer collaborative assessment (peers can work together as they complete an assessment task);
- opportunities for feedback and feedforward (learners' performances can be enhanced as they receive advice from their teachers on the quality of current performances, and guidance about how they might improve their performance on future occasions).

A key New Zealand Position Paper on assessment (Ministry of Education 2011) underscores a commitment to assessment for learning in the NCEA, promulgating a discourse that lauds the powerful benefits of assessments that are firmly embedded within teaching and learning. This reflects a "*deliberate focus* on the use of professional teacher judgment underpinned by assessment for learning principles rather than a narrow testing regime" (p. 4, my emphasis). The NCEA system has given teachers extensive control over how best to organise assessments according to their own contexts and students, and a good deal of freedom to mark and provide feedback on their students' performances in accordance with the published standards.

The constructivist agenda was continued with the publication of a revised national curriculum for schools (Ministry of Education 2007), mandated from 2010, and underpinned by core values such as innovation, inquiry, and curiosity, and key competences such as thinking, managing self, relating to others, and participating and contributing. Beginning in 2008, a process was initiated to review the NCEA

achievement standards to ensure their alignment with revised curriculum expectations. A unique across-the-board approach to assessment renewal involved input from a range of stakeholders, including academics and practising teachers, who were able to bring their different expertise to bear at various stages of the renewal and implementation process. Revised achievement standards were phased in over a three-year period, starting with level 1 in 2011.

Assessing FL Students' Oral Proficiency

In the context of the above recent across-the-board curriculum and high-stakes assessment reforms, the assessment of FL students' oral proficiency has undergone significant transformation.

The original (2002 onwards) NCEA introduced an achievement standard known as *conversation* (hereafter referred to as *converse*). Building on the oral assessments that had been part of School Certificate and Bursary, *converse* was a summative teacher-led interview test, operationalised along lines that have become accepted practice for this kind of oral assessment (Luoma 2004). At the end of the year, candidates would be interviewed independently by their teacher and would respond to a series of questions. The teacher would then provide a grade for each candidate's performance. The assessment was problematic, however. Since the teacher was the examiner who posed the questions, the students often knew beforehand what would be asked, and there was a tendency for students to prepare their responses in advance and then rote-learn these. The assessment was also one-sided (teacher asks; student responds), leading to artificial and restricted 'conversations'. There were limited opportunities for the assessment to replicate the genuine reciprocity and spontaneity of authentic interactions. As such, there was negligible evidence of interactional proficiency, calling *converse* into question as a valid means of assessing oral proficiency. Most particularly, being summative and test-like, the *converse* standard did not match very well against the perceived advantages of NCEA noted by Hattie (2009). Subsequent curriculum and assessment reforms provided the opportunity to review (and replace) the standard (East and Scott 2011a, b).

Informed by a learner-centred and experiential pedagogical approach, the revised New Zealand Curriculum advocated FL courses that would promote significant opportunities for communicative interaction among learners. East (2016) notes that the theoretical drivers for this may essentially be distilled into two key principles: that students learn to use the FL most effectively when they engage in real language use in the classroom (Willis and Willis 2007); and that students learn how to communicate through interaction in the FL (Nunan 2004). Philp et al. (2014) underscore the beneficial *learning* potential of FL peer-to-peer interactions from both a cognitive perspective – e.g., Long's (1983, 1996) interaction hypothesis – and a sociocultural perspective – e.g., Vygotsky (1978) – from which learning is seen as "a jointly developed process and inherent in participating in interaction" (Philp et al. 2014, p. 8). The learning end-goal of such interaction is automaticity in language use

(DeKeyser 2001; Segalowitz 2005), or, as East puts it, "the ability of language users to draw on their knowledge of the FL automatically and spontaneously" (p. 4).

Precipitated by the communicative goals for FL learning encouraged as a consequence of curriculum reform, the process of aligning FL achievement standards with the revised curriculum resulted in the introduction of a new standard (called *interact*) whereby students would collect evidence of several learner-focused peer-to-peer interactions that take place throughout a year-long course. *Interact* was phased in as a replacement to *converse* over the three-year period beginning in 2011. Achievement criteria at NCEA levels 1, 2 and 3 were informed by the can-do statements for spoken interaction of the Common European Framework of Reference for languages or CEFR (Council of Europe 2001).

Theoretically, the new direction for assessment heralded by *interact* enables FL learners to take part in, and to present some evidence of, authentic and spontaneous interactions that represent a range of contexts, which might include genuine interactions with first language (L1) speakers of the target language in real-life situations (e.g., technologically mediated communications such as Skyping and interactions during trips overseas). *Interact* thereby aims to enact the interrelation between the theory informing FL curriculum reform and actual practice in that it seeks to provide the means to introduce, and include as assessment evidence, genuine and spontaneous peer-to-peer interactions in the FL. It was considered that the proposed new assessments would lead to more valid evidence of interactional proficiency because, for example, a level of natural and unplanned reciprocation is required. Also, the fact that assessments associated with *interact* are to take place throughout the year arguably facilitates the enactment of the '*for* learning' benefits of the NCEA noted by Hattie (2009): the integration of assessments into teaching and learning would allow students to work together to complete assessments aligned to teaching and learning goals at different points in the year, and also to receive feedback and feedforward from their teacher.

In the teaching and learning context, teachers and students have substantial autonomy to manage interactions in any ways they choose. Theoretically, all students are free to interact about any topics they wish, at any time, and in any place, and to use any part of those interactions as evidence for assessment purposes. Not all interactions need to be made available for assessment. However, interactions for assessment purposes need to be recorded, either by the teacher or the students, and either as audio or video files.

At the end of the year, students and/or their teachers select from the recordings instances of what are considered best evidence of students' interactional proficiency. A minimum of two samples, lasting in total between 3 min (level 1) and 4–5 min (level 3), are to be submitted for grading purposes. Summative grades, awarded by the teacher, are derived from the overall quality of the submitted work, that is, each component interaction is not assessed individually (NZQA 2016a).

In practice, the considerable freedom that *interact* appears to be offering is constrained by several factors. Although the Position Paper (Ministry of Education 2011) states that "[t]he design of the NCEA is such that it provides *potential* for assessment to be used *formatively* and to be an *integral part* of teaching practice"

(p. 14, my emphases), the word "potential" hints at a problem in practice. Indeed, when the discussion in the Position Paper turns to the NCEA, we are presented with a somewhat contradictory discourse that underscores an *accountability* function that may be at odds with a *learning* function. For instance, reference to formative feedback in the Paper actually points to its use in the process of "*preparing* students for *formal* certification" (p. 15, my emphases).

The Position Paper argues that the NCEA standards "have an important *external* reporting function based on a 'summing up' of where a student's learning sits in relation to an expected standard," leading effectively to "essentially a *summative* use of assessment information" (Ministry of Education 2011, p. 11, my emphases). This has implications for how *interact* is put into practice, and the extent of feedback/feedforward that can be offered. The standards are, after all, "primarily assessment tools" (p. 11), and there is a requirement to demonstrate that any work submitted for assessment is the student's own unaided work. Thus, while the Position Paper appears to support a much stronger relation between assessment and teaching and learning than had previously been the case, the traditional demarking of assessment events from teaching and learning processes is perpetuated. Something of a clash of purposes emerges between assessment *for* learning and the assessment *of* learning. It is worthwhile noting here that such contradictory messages can be observed in other education systems, as is evident from other research reported in this volume (see Chap. 12 by Inbar-Lourie and Levi for a similar phenomenon in the Israeli education system).

Subsequent guidelines that inform how teachers are to implement internal NCEA assessments underscore accountability. So-called 'quality assurance systems' include both internal moderation of assessment tasks to ensure comparability across different classes, and external moderation of samples of work to confirm that both assessment tasks and grading of students' work are consistent with a national standard (NZQA 2016b). There is also an expectation that students will be told in advance when they will be assessed, and on what topic. That is, students are required to be given a task brief that outlines what they will be expected to do, alongside an indication of when the assessment will take place.

With regard to *interact*, we are left with a genuine tension between authentic and spontaneous peer-to-peer interactions in the context of teaching and learning, and interactional tasks that must conform to a set of external and national criteria, and must be clearly flagged as assessments. Having replaced *converse* due to its inadequacy as a valid means of assessing oral proficiency, it seems we are still left with a key limitation – an assessment format that continues to anticipate measurements of oral proficiency through discrete events rather than as fully integrated into teaching and learning. There is an apparent paradox between two potentially irreconcilable assessment aims – a 'professional and learning' objective and a 'managerial and accountability' objective (Gipps and Murphy 1994). That is, the potential of interactions for assessment purposes to contribute to students' learning is countered by the risk that interactions for assessment purposes will be viewed by students as measures of their ability, thereby leading them to focus on performance to the detriment of genuine interaction.

East (2016) reports a substantial two-year project designed to establish how teachers and students were coming to terms with the new assessment as it began to be utilised in real classrooms. The study sought stakeholders' views about *interact* in comparison with *converse*, in two phases: Phase I (2012) gathered teachers' views with regard to NCEA levels 1 and 2 via a national survey and interviews. (A small-scale student survey also collected students' views on *converse* [NCEA level 3] in its last year of operation.) Phase II (2013), which focused on NCEA level 3, elicited teachers' comparative views (interviews) and students' views on *interact* (survey).

A key survey finding was a clear articulation of the learning/accountability paradox. On the one hand, *interact* was acknowledged in terms of its potential to *enhance* students' interactional spontaneity and automaticity. It was reported that *interact* encourages "authentic use of the target language between the students rather than answering the teacher's prepared questions" (teacher of Japanese, East 2016, p. 112) or "helps the students to really interact as they would if they find themselves in [for example] France" (teacher of French, p. 113). This makes *interact* "a more accurate measure of the student's ability to respond to an interaction in a real-life situation" (teacher of Spanish, p. 116). On the other hand, the high-stakes and accountability character of the assessment serves to *undermine* expectations of interactional spontaneity and automaticity. As two teachers of French put it (East 2016, p. 129), when *interact* is seen as "still an examination after all and they want to do well" (French1) spontaneity and automaticity are "too big an ask of our students. They find it almost impossible to do this in unrehearsed situations" (French2). It is therefore "hard not to have students scripting speaking tasks" (French1). Thus, it seems that the challenge for students of producing language spontaneously is exacerbated by the emphasis on performance created by a 'testing' focus.

In what follows, I present findings of a study that took place towards the end of 2015 as a focused follow-up to the extensive study reported in East (2016). The purpose of the follow-up research was to explore the learning/accountability dichotomy from the perspective of both teacher and learners in the context of one school setting, by investigating one teacher's enactment of *interact* at NCEA level 3 and its reception by his students (n = 7). The follow-up research is framed as an instrumental case study (Grandy 2010), aiming, as Wellington (2006, p. 30) puts it, to "provide insight into a particular issue," in this case the implications and outworking of the learning/accountability dichotomy, and to "develop our understanding and knowledge" of this issue.

Introduction to the Case

The school in question is a suburban co-educational secondary college on the outskirts of a major New Zealand city, taking students from Years 9–13 (13+ to 17+ years of age). The class in question is a class of seven Year 13 students of French as FL preparing for NCEA level 3, benchmarked against levels B1 and B2 of the

CEFR. Although the teacher, Frank Smith,[2] had worked with this class since they were in Year 11, this was the first year that he had introduced *interact*, having chosen not to opt into this assessment in Years 11 and 12. The class comprised three girls and four boys. Apart from one boy (François) whose parents were French and who spoke French at home, all students were L1 speakers of English.

In the broader context of a cyclical theory-research-practice research agenda as described at the start of this chapter, the researcher had informally had several discussions with the teacher about how he was dealing with the introduction of *interact* with this class, and the progress he was making. As the students' academic year drew to a close and they were completing their final *interact* submissions, the researcher subsequently invited the teacher and his students to participate in a formal evaluation of how the introduction of *interact* had gone, with a view to identifying aspects of the assessment that might require modification. The teacher took part in a semi-structured interview, and the students were members of a focus group discussion. Both interview and focus group were digitally recorded and later transcribed. The findings presented below illustrate aspects of how the dimensions of accountability and learning play out from the perspective of the teacher and the learners.

Findings

The Teacher's Perspective

When asked about the learning potential of *interact*, Frank asserted that, in his view, the new assessment took his students from clearly rote-learned rehearsed dialogues to situations in which "they were having actual conversations about actual things that they actually cared about." In other words, *interact* "took the kids from French *learners* to French *speakers* in a very genuine, authentic kind of way" (my emphases).

Having first picked up this class in Year 11, Frank recognised that, by Year 12, "they had a real struggle around the talking." At this stage, Frank did not utilise the *interact* assessment, but rather "set up a system whereby I would go on the internet or find an article about a current issue, topic, whatever." Students were required to engage with the material (adapted to their level of language if necessary), to "read it, formulate an opinion about it, and then come back on the first lesson of the following week with something to say about it." The students were free to say what they wished. As a consequence, "some kids said a lot, others said nothing, some struggled, some didn't." Nevertheless, "because it was so open-ended, it forced them into having to formulate their own ideas and then critically take those ideas and express them in French." Through the interactions "they were listening to each

[2] All names in this chapter are pseudonyms.

other, and then they would pick and choose and steal bits off each other." The students "appreciated being able to make mistakes [and] … to not feel like fools, to be able to laugh at each other, with each other, about themselves, and so on and so forth."

Having followed this pattern in Year 12, relationships between the teacher and the students in Year 13 were sufficiently comfortable that "their filters were down … [and] we could muck about, you know, they could make fools of themselves, and I could laugh at them and they could laugh at me, and it was quite relaxed." As a consequence, Frank decided to introduce the *interact* assessment as his students moved into Year 13.

At first the interactions, whether assessed or not, lacked spontaneity and fluency and students relied on pre-learned chunks of language. This "rehearsed stuff" was "the response to the stimulus material they were given, and they would come back with a pre-prepared thing." Frank did not see this as problematic, however. There was, in Frank's view, nothing inappropriate about drawing on pre-learned formulaic language "as a starter or as a means through which they're moving towards that spontaneous [interaction]." After all, "I did that in France and it was an authentic genuine conversation. The French person didn't go 'oh that's a formulaic expression'." To be able to interact spontaneously, "there has to be a whole background of language use in place." Frank concluded, "it's still spontaneous, actually, because they are still having to select the language, they are still choosing what to say from their knowledge." As the interactions began to develop, eventually his students had established "a repertoire of things that they could say … to state their opinions."

The transition from artificial and rote-learned to genuine and authentic in the course of the year was hard to pinpoint. Frank argued, "at some point, goodness knows how to measure when, they [the interactions] stopped being formulaic responses, and they become automated, integrated language use." By the end of the year, his students had reached the stage where the interactions were spontaneous even though they were underpinned by formulaic expressions.

Frank gave an example to illustrate his point. He explained, "there was this catch phrase that … one girl had learned – *choquant*, that is 'shocking'." He noted that at first "she was waiting for the opportunity to use it, spotted the opportunity and inserted it." However, "by the end of the year, she was bandying it about, like, tongue-in-cheek. She was using it as humour and all the kids were laughing." Frank concluded that this example was "quite representative of the shift between something becoming a formulaic response to actually becoming a spontaneous thing."

Despite the positive advances in spontaneity that Frank witnessed in his students, he conceded that a crucial limitation of *interact* was how it was to be operationalised for assessment purposes. He was mindful of the requirement to give students a task brief for the interaction, pre-selected and pre-approved, and likely to elicit the complexity of language required to demonstrate proficiency at the targeted level, that is, "a written task that has been moderated in advance of it being given out and critiqued and … deemed to be at [NCEA] level 3." As a consequence, students immediately became "aware that they are being assessed." In Frank's experience, the interactions therefore became "the spectre that hung over them a little bit." At

least at the beginning, his students "found it really stressful," because "they knew they were being assessed on what they were saying." It was possible to "see them trying to insert a wide variety of language for example, or [knowing] that they'd screwed up the pronunciation of a word." It seemed that Frank's students "really struggled to get away from the fact they might be penalised," that is, get a lower grade than they might wish.

Frank took several steps to mitigate the potentially disinhibiting effect of interactions used for assessment purposes. Since he wanted his students to "stop thinking of it as an assessment," he encouraged them to record a variety of evidence in a range of circumstances. For example, he explained to them that "if you are having a conversation at the beginning of the lesson … just record it on your phone …" He elaborated:

> If you come into French class one day and you think 'hey, I've got my French mojo on today, I just feel like I can talk French,' go talk to your mate about whatever it is you want to, record the whole conversation, change topics five times for all I care. Within that, find a section [where] you think 'hey, that was pretty good that we talked about something that was quite meaty' or you were passionate about or you went a little bit deeper into, and then submit that as your *interact*.

To try to ensure at least some kind of fit to an assessment brief, Frank remarked that "hopefully it's about racism or some of the things we've been discussing in the class." Frank observed that, in practice, it appeared possible to embed more complex language and topics (i.e., language and topics that would meet the required level) within the more 'mundane'. He gave a useful illustration:

> So talking about the school ball and just talking for 15 minutes about all the different dresses that everyone was wearing, they know that is not really going to be evidence of their knowledge. But they might for five minutes or three minutes of that conversation talk about the pressure to buy the most beautiful dress or to be voted the prom king or queen, or the social tension that comes with dating and how that has a real effect on some people, it is almost like passive bullying or whatever, there might be some little gems like that in there, yeah.

Frank concluded, "I just ended up making it as open as they wanted it to be to just take that pressure off them, so they could just have a conversation, which is what the standard is trying to assess."

Frank's means of putting the assessment into practice revealed what he saw as "a real difficult tension to manage" with regard to *interact*, a tension that he elsewhere described as "fundamentally … irreconcilable." That is, "on the one hand the standard is purporting to assess spontaneous conversation." There therefore needed to be "the scope to just hit record on my phone or theirs if … they're having a conversation, then going 'well, actually, I can just submit that.'" On the other hand, "we are prescribing the things that they talk about in pre-prepared tasks with a set topic," and "the requirement for it to be a pre-written task actually kills the spontaneity of that interaction." In other words, "forcing them to talk about a particular subject" was simply "not spontaneous." It seemed that Frank was here wrestling with the realisation that genuinely spontaneous and natural occurrences of interaction were

not being elicited or captured when students were placed into an artificial assessment context that put the spotlight on students' 'performances'.

Added to the tension between spontaneity and pre-determined task was the perceived necessity to "measure their knowledge" with regard to "a skill set … the grammar structures, the vocab, the way of speaking [that] they may not talk about if they are just talking about how they went on a bike ride during the weekend or went and played football at the park with their mate." In other words, there appeared to be an expectation for students to move into a register of speech that would simply not be used in the more genuinely spontaneous and natural interactions of impromptu conversations.

In fact, it is evident that Frank did attempt to resolve the tensions he perceived by allowing his practice to deviate from a rigid assessment brief, even though he observed that, because of that, he "might fail [external] moderation." It was, however, important "to consider how the evidence can be collected to actually capture spontaneous language use."

In line with the '*for* learning' potential of the assessment, Frank used what he described as "little feedforward sheets for them" in order to develop his students' interactive skills. However, in keeping with an accountability requirement that meant that he was not in a position to offer direct and explicit feedback on the language used, his feedback was quite holistic. After each interaction "I would give them an indicative thing … 'needs work', 'good', 'awesome', or 'pay attention to'." The kinds of feedback offered "were things like seeking clarification of other people's opinions, extending and sustaining the conversation, asking questions, stating your opinions, expressing your opinions clearly, talking for long enough, pronunciation, intonation, stuff like that."

With regard to the *timing* of the feedback, "rather than giving it to them after they had done the last one, because they would look at it, file it, forget it," Frank "would give it to them as feedforward right before they did the next one, and say 'hey, this time be thinking about these things. Go away and have a practice now with these things in mind.'" He noted that his students "really appreciated that, they were actively asking for that feedforward, which was good, because they were thinking about the process and trying to shift around to meet the criteria." Frank worked on the assumption that his students were all capable of getting the highest grade – *achieved with excellence*. He therefore gave them feedback "based on the fact that [*excellence*] was what they would be aiming for" in the belief that "for the kids that were going to get *excellence* they would get that, for the kids who couldn't, it still meant they were being pushed."

In the course of the year, Frank remarked that his students' stress appeared to dissipate. An initial source of stress was the "performance" nature of interactions that "they weren't allowed to rehearse; they could practise but not rehearse." This meant that there was "a huge quantity of the unknown in there." Over time, however, his students came to recognise that "spontaneous usage" is "actually the nature of language." His students "just had to deal with that [uncertainty] on the spot." As a consequence of the skills that they developed through doing the assessment, "at the end of the year they went 'yeah, that was a really good standard'." Indeed, this was

"the general feedback and certainly the impression that I got anecdotally, as they became more comfortable with the process … as the year went on, they got better at that and they naturally relaxed around it."

The Students' Perspectives

Students in the focus group articulated perceived benefits of *interact* that resonated with their teacher's perspective. Arguing that oral proficiency was "really what we are trying to learn, I guess, when you take a subject like French," Robert suggested that *interact* was "a really valuable assessment" because it represented "the only really proper opportunity for us to speak together in French." According to François, the value of *interact* was that it "develops your French" and provides evidence that "you can speak spontaneously and think [spontaneously]." In Graham's words, *interact* "tests your knowledge about French speaking on the spot, and having to actually know how it works properly … without rehearsing it." Since, as Anne noted, "to be able to talk it [French] fluently, conversationally, is really handy," *interact* was, in Sandra's view, an assessment "where you actually *apply* what you've learned, yeah it's the *application*" (my emphases). Peter summed up the perceived value of *interact* in these words:

> It is all good and well to be able to read a text and then decipher what it means, but in the real world you need to be able to talk and hold proper conversations. So that is what *interact* helps us to learn.

Several students spoke of the learning potential of *interact* in terms of developing their oral proficiency throughout the year. Anne commented, "it was harder at the start than what it is now." Sandra spoke of reaching "a new level of confidence." She asserted that, initially, "I was not a very confident French speaker at all." However, "with everything practice makes perfect … so I think it is just a case of just practising it lots" because *interact* "gives us that really good opportunity to be more confident in pronouncing French and forming phrases and saying what we think."

Sandra went on to explain that students feel as if they are "in the spotlight at the beginning" because "you are not very used to it and so you begin to doubt everything you've ever learned in French over the last five years and you wonder why you are even here at all." Reflecting back on the past year's interactions, Sandra concluded, "it is a lot easier now because we realise that [we need] to trust ourselves, I guess, to trust ourselves with this language, that we have learned something." Peter noted a broadening of his vocabulary repertoire:

> I think my variety of language definitely increased, you know, conversational French. You find yourself using a lot more tenses, you find yourself instinctively knowing how to say the things that you want to say rather than having to look up what you want to say in the *interact*.

Although students did not speak directly about the stress of interacting in an assessment situation, the reality that students did feel stress and anxiety, together

with the acknowledgment that this stress and anxiety diminished over time, were implicit in their discussion of how the interactions played themselves out in the course of the year. Sandra, for example, spoke of becoming "less scared" as you carried out interactions during the year, "I guess because you have already attempted it and you know you can do it."

In terms of anxiety, the requirement to be spontaneous was perceived as a significant challenge. Mary argued that, to get the higher grades, "the level of fluency … has to be pretty high … but sometimes if you are just thrown into it then it's pretty hard to think of stuff spontaneously." Anne asserted that "even though it was a good thing, it was difficult." This was because "for the whole time you've been studying French you've been rehearsing, like you do speeches, but you write the speeches and then you are doing all tests, but it's all written tests." With *interact*, "you feel kind of self-conscious, 'oh, am I pronouncing this right? Do people understand me?'" Anne went on to comment that actually recording the interactions led to a sense of "pressure" and "oh it's got to be perfect."

At the start of the year, the challenge, for Robert, was "mainly a confidence issue. … I think a lot of us were just thinking we can't do it … and so then it just sort of halted our conversations." However, "when you do it a few times you realise 'okay, this is possible, this is doable' and you feel fine about it." Mary concurred that, before they began the *interact* assessments, "when we were talking in class in group discussions in French I found you slightly hold back because you are kind of afraid [that] people might judge you or whatever." However, as a consequence of interacting, "you feel more comfortable talking in French with your peers."

Another factor that appeared to contribute to students' developing confidence was that, in addition to incidences of dedicated focus on an interaction, the teacher also opened up opportunities beyond a 'formal assessment time' to speak in French. Anne remarked:

> You know, some days you are in a French mood and you are feeling really confident about your knowledge, and so Mr. Smith was really happy for us just to go and have a conversation, even if you are not going to use it [for assessment], just to practise or just talk and also without having the pressure of … 'oh, we are going to do *interact* on this day' or something. It was kind of nice to be a bit more relaxed about it.

Robert concurred, "I think something that really helped was that Mr. Smith just gave us the freedom to go and do an *interact* when we wanted to and when we felt like we were confident enough." This awareness of an apparent development in his teacher's pedagogy as a consequence of implementing the *interact* standard meant that "throughout the year we just had any opportunity we wanted to just go and practise speaking, so yeah, just that freedom." For Mary, the opportunity to "do as many as we want over the whole course of the year" essentially "gives you such a wide range to pick from to actually submit." This, in her view, stood in contrast to a one-off summative interview where "I think if it was still at, like, just one day at the end of the year, with so much going on you would feel so much more pressure, and I think it would be more forced and not as relaxed." Graham agreed that "just being

able to do it when you want" was a distinct advantage. He conceded, "I don't like a deadline of when you have to do it."

The students were also given a good deal of freedom to interact about topics they generated for themselves. Mary concluded, "I think that was better because it was stuff that we were actually interested in, so we were able to be more passionate about it."

Another perceived advantage of undertaking interactions over time was a perception of increasing interactional proficiency. Mary reflected:

> If you compare the ones we did at the beginning of the year to the ones we did at the end of the year, you can just see how much you've learned. And I think it is good not only for the standard, but also for yourself, to just show you how much you've actually learned across the year.

Several factors appeared to contribute to students' perceptions of improvement in oral proficiency in the course of the year, with consequent greater ease in approaching the interactions. François reflected that, over time, the amount of prior preparation to complete an interaction diminished. He explained:

> I remember at the start of the year I did an interaction with Sandra, and preparation-wise it took most of the lesson, maybe three quarters of the lesson, and then we had the last quarter to actually interact. But the last time with Graham and Anne preparation-wise maybe 10 minutes max. We just thought of some points and then spoke for about 5 minutes … we had it all recorded, we just handed it in and it was a lot more flowing.

Another important advantage of on-going assessment, commented on by several students, was the development of strategies to maintain the interaction. For example, not having an adequate language base could be problematic. Mary explained that it could be frustrating when "you knew what you wanted to say but either you didn't have the word for it or you just didn't know how to phrase it together properly so it was actually fluent." To deal with this, Mary explained, "I guess I just said something else or I said [something] that was a bit simpler but in the same context." There was, however, significant learning potential here. As Mary explained, afterwards "I'd go away and learn about it more so that next time I could actually apply it into the *interact*." For Anne, the greater ease of interactions over time was supported by "just more knowledge and more practice, learning different ways on how to respond and to ask questions and to say what you think."

It seemed also that *interact*, being a paired or group assessment, became an exercise in collaborative strategising among the students. Peter observed, "we are all trying to get to the same goal and we are trying to achieve the same thing and get as high marks as possible." As a consequence, "we help each other out because we understand there is difficulty in trying to speak spontaneous French, yeah." As an example, Peter remarked that you could perhaps move a topic onto others if you were stuck:

> There are expressions that we could learn to kind of divert the questions along – 'I think this question would be better to ask the person next to you' or something, and you flick it onto someone else. So there are ways you can avoid halting the whole interaction rather than stopping it and redoing it.

François, as the L1 French speaker in the group, noted, "when I saw there was a halt in the conversation, sometimes I would just, like, pop in and just said something." This was not just because he had the most solid command of French in the group. He commented, "I think that goes for anyone else, you know." François continued:

> You say what you want to say whenever it works. So if Anne found it hard to say something or was thinking about it and Graham wasn't saying anything, you could say something and add to it and then maybe it will spark something for Anne and so she can start talking.

In this way, François concluded, you "build up and snowball together." As Mary explained, "in English, if you are having a conversation and someone just stops and they can't think of a word for something, someone else will just jump in and be like 'oh, do you mean this word?'" She concluded, "so I think we all kind of helped each other out in that sense, like, if we didn't understand something we could try to clarify it for them."

Having François in the group provided an opportunity to assess the students' learning from an L1 perspective. François asserted that, in the space of a year, his classmates' French had "developed a huge amount." He commented that "at the start of the year, since we'd never done *interact* before," initially, his classmates' interactions were "difficult to understand." However, now at the end of the year, and commenting on the students' most recent interaction, he stated that, on the whole, "it was fine. There were a few juddery bits, but it was all pretty much flowing fine." He went on to reflect, "the use of phrases and just building phrases spontaneously is a lot better."

François also provided a valuable L1 perspective on the kinds of language he observed being captured by *interact*, which could stand in contrast to the level of language utilised elsewhere in the course, but which exposes a genuine tension for *interact* in terms of expected language level. He remarked that "with French, like, writing and reading in class, a lot of the time it is really formal and you use tenses and stuff that you'd never use in everyday [life], you know." He noted "since I speak French at home I hear some things that we talk about in French class which are just ridiculous. You never hear these or see them in everyday life." Thus, the language outside of the context of *interact* could be "ridiculously complex," whereas *interact*, in his perception, focused on "just everyday language which people need." He asserted, "I think it is good to be able to speak like a French speaker and not speak like a machine which [you find] a lot of the time in French texts we read in class." He concluded that this was "just silly really." François' perspective here provides a valuable counter-balance to Frank's concern that his students needed to demonstrate a more formal register in the evidence of interaction submitted for grading purposes, as well as insight into what his teacher's practices were actually achieving. It seems that Frank had clearly succeeded in enabling his students to embed more formal language into their broader interactions in ways that led François at least to perceive of the interactions as more natural and (interactionally) authentic than the reading/writing samples the class was required to deal with.

François summed up what he saw as the most positive outcome of *interact* for the class:

> Last year we would never, you know, have spoken in French, the class wouldn't speak, it was, like, English the whole time. Mr. Smith would make an effort to speak to us in French, but nobody would [speak back]. … This year, if Mr. Smith speaks to us in French, I've noticed that the whole class chips in to speak in French, and I think that is due to *interact* as a standard – and, yes, it's good.

Discussion

As a component of a broader research agenda into how curriculum and assessment reforms are received and enacted by real teachers in real classrooms, the case presented above raises several issues regarding the on-going implementation of the *interact* achievement standard.

Frank, the teacher in this context, is clear about both the possibilities and limitations of *interact* as an assessment of oral proficiency. It is seen as a genuine, authentic and successful opportunity to undertake interactions that ultimately demonstrate 'automated, integrated language use.' Nevertheless, operational requirements such as pre-selected and pre-approved task briefs likely to elicit the appropriate complexity of language tend to place the focus on the interactions as assessments, potentially becoming a 'stressful spectre' that 'actually kills the spontaneity of that interaction.' Frank describes this as an 'irreconcilable tension.'

Frank does, however, reconcile the tension in his own practice: he encourages students to take part in and record a range of interactional opportunities beyond more formal assessments, and asks them to consider whether, within those more spontaneous interactions, there is any section that might be submitted for assessment purposes. Frank recognises, though, that this workaround 'might fail moderation' because it deviates from a prescribed assessment brief.

It seems the learners speak of possibilities, but not of limitations. The students do not speak at all from the perspective of accountability (apart from accountability to themselves to do as well as possible in the assessment). Their focus on the clearly perceived positive learning potential of *interact* is not surprising given that they are the learners in this context, and that the requirements around conditions of assessment (e.g., setting a written task brief before the interaction) are an accountability issue for their teacher, but not for them.

It seems that, by the end of the year, *interact* has taken the students 'from French *learners* to French *speakers*' (teacher). The students comment that they have 'developed a huge amount,' could all 'chip in to speak in French,' and have reached a point of 'instinctively knowing how to say the things that you want to say.' The interactions are sustained by learning a range of interactive strategies – students have not only developed 'a repertoire of things that they could say', but are also 'listening to each other' and will 'steal bits off each other' (teacher) and will 'help each other out' (student).

Encouragement to record a range of evidence in a range of circumstances (not all of which would be assessed) gives students the freedom to complete an interaction

whenever they want to – 'making it open to take that pressure off them' (teacher) or just letting them 'go and have a conversation' (student). The teacher particularly capitalises on opportunities that seem optimal for spoken interactions – interaction when 'I've got my French mojo on' (teacher) or when 'you are in a French mood and you are feeling really confident' (student). It seems, from both sets of perspectives, that, by the end of the year, the assessment is appreciated – the students 'got better' and 'naturally relaxed' (teacher), and it was possible to 'see how much you've learned' (student).

Frank recognises the necessity to elicit the complexity of language required for the level of assessment, but, it seems, largely trusts that instances of that language will be in evidence in the context of more general conversations (he cites discussions around the school ball as an example). This, it seems, 'disguises' the more complex language such that François, the L1 speaking student, sees *interact* as avoiding the 'ridiculous', 'silly' and 'really formal' language that might be required elsewhere in the course. This does, however, raise an issue for *interact* around expected complexity of language that requires some resolution.

Frank's management of the *interact* standard provides some valuable insights moving forward. Frank steers a careful path between the accountability expectations of *interact* and freedom for his students to interact about any topics they wish, at any time, and in any place. It appears that this leads to his students' increasing confidence in interacting spontaneously and authentically, whilst at the same time ensuring that samples of language at the required level are elicited and recorded. To achieve this, however, it does seem that Frank is compelled to suspend somewhat the requirement to present his students with a task brief in order to diminish the perception that a particular interaction is for assessment purposes. Frank's stance in this respect is reminiscent of one interview participant in East's (2016) study (Alison, a teacher of Japanese) who argued that having to specify the tasks made the interactions "more artificial," and only taking away the task brief would promote "true authenticity" (p. 166). To solve this dilemma, Alison planned either to "retrofit the tasks" around what her students submitted or "make [the tasks] so generic" that whatever they submitted could fit.

One amendment to the current implementation of *interact*, which would be easy to put in place without having to compromise task moderation requirements, might be to allow teachers to provide open-ended instructions at the start of the year. These instructions might make it explicit to students that, in the context of addressing a range of topics, students will be encouraged to interact with each other about the topics and to record some of these interactions, and that some instances of interaction might be extracted from the recordings for assessment purposes. This would not preclude teachers, from time to time, suggesting to their students that a particular interaction opportunity might be particularly useful to record. This would help to ensure that interactions that lent themselves to particularly 'meaty' language (as Frank put it) were recorded.

Conclusion

The findings from this small-scale project suggest that in several respects the *interact* standard is functioning and achieving the goals it was designed for. Specifically, the peer-to-peer interactions described by Frank and his students have important learning potential (Philp et al. 2014), and indeed the students appear to be learning to communicate through interaction in the FL (Nunan 2004), leading to enhanced linguistic automaticity (DeKeyser 2001; Segalowitz 2005). There is also evidence of a constructive alignment between learning and outcomes, and opportunities for peer collaborative assessment, feedback and feedforward (Hattie 2009).

Nevertheless, "formal certification" (Ministry of Education 2011, p. 15) and the summation of "where a student's learning sits in relation to an expected standard" (p. 11) create tensions for the assessment, and it seems the *learning* potential of *interact* is only enhanced because some of the more formal expectations (i.e., moderation and standardisation of assessment tasks) are occasionally circumvented. It is perhaps at this point, that is, the setting of assessment tasks, that more open-endedness might be beneficial. This does not need to compromise the formal certification dimension of the NCEA because it will still allow for students' proficiency to be demonstrated in relation to the expected standard. However, this would arguably be achieved in a way that shifts the focus somewhat away from a specific interaction as an *assessment* and therefore promotes greater authentic interactional authenticity.

The study reported here is limited in several respects. First, the sample size of one teacher and seven students in one school is very small, and the findings are therefore not generalisable. However, as an instrumental case study (Grandy 2010; Wellington 2006), the findings have merit in highlighting issues around learning and accountability that can inform debates going forward. Second, the study relies on self-report data and does not take account of either classroom observational evidence of interactions, or instances of the interactions the students participated in. However, as a study that informs theory and research around evidence of genuine and authentic interactional proficiency, the complementary evidence from teacher and learners provides corroboration of what appears to be the genuine learning potential of *interact.* Third, it needs to be acknowledged that Frank was arguably in an optimal position to make the most of *interact* by virtue of his participation in a theory- and research-informed pre-service teacher education programme as well as the subsequent opportunities he had to reflect on his own practices in light of theory and past research (see, e.g., East 2017). This highlights the significance of nurturing teachers' assessment literacy through mediated professional development activities and through teacher-researcher encounters (for examples of such work, see Chaps. 2, 7 and 8 in this volume by Michell and Davison; Chan and Davison; and Hill and Ducasse). Other teachers might be less far along the road to successful implementation of *interact*. Nevertheless, Frank's story of implementation, as recorded here, continues the contribution to on-going evidence about real practitioners' engagement with theory- and research-informed initiatives.

The present study provides valuable evidence that, in practice, constraints can be mitigated. The evidence from the present study, taken together with that from East (2016), is that there is considerable potential for *interact* to become a valuable assessment of genuine interactional proficiency. However, as East (2016, p. 199) concludes, we are "left with a fundamental problem for *interact*, and for assessments like it. That fundamental problem is the tension between two different and potentially irreconcilable paradigms for assessment," or, as this chapter has put it, the dichotomy between learning and accountability. In Frank's words, it is an 'irreconcilable tension'. Frank goes on to resolve it somewhat, but only by being willing to challenge the published requirements in his practices.

As *interact* continues to be evaluated, and perhaps revised, reformulated, and improved, through teacher-researcher collaborations such as the one documented here that seek to ensure that there is an on-going interface between theory, research, and practice, evidence can be gathered about the strengths and limitations of *interact* in practice that may become drivers of change into the future. Ultimately, an assessment is required that can elicit the best evidence of learners' oral interactional proficiency. *Interact* has much to commend it in this respect. There are also some issues to resolve.

References

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge Faculty of Education.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved from http://www.assessment-reform-group.org.uk

Council of Europe. (2001). *Common European framework of reference for languages*. Cambridge: Cambridge University Press.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge: Cambridge University Press.

East, M. (2012). *Task-based language teaching from the teachers' perspective: Insights from New Zealand*. Amsterdam/Philadelphia: John Benjamins.

East, M. (2014). Encouraging innovation in a modern foreign language initial teacher education programme: What do beginning teachers make of task-based language teaching? *The Language Learning Journal, 42*(3), 261–274.

East, M. (2016). *Assessing foreign language students' spoken proficiency: Stakeholder perspectives on assessment innovation*. Singapore: Springer.

East, M. (2017). Out with the old and in with the new? The benefits and challenges of task-based language teaching from one teacher's perspective. *Babel, 51*(1), 5–12.

East, M. (2019). Sustaining innovation in school modern foreign languages programmes: Teachers' reflections on task-based language teaching three years after initial teacher education. *The Language Learning Journal, 47*(1), 105–115.

East, M., & Scott, A. (2011a). Assessing the foreign language proficiency of high school students in New Zealand: From the traditional to the innovative. *Language Assessment Quarterly, 8*(2), 179–189.

East, M., & Scott, A. (2011b). Working for positive washback: The standards-curriculum alignment project for learning languages. *Assessment Matters, 3*, 93–115.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
Grandy, G. (2010). Instrumental case study. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research*. Thousand Oaks: Sage.
Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment and higher education student outcomes: Policy, practice and research* (pp. 259–275). Ako Aotearoa: Wellington.
Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126–141.
Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic.
Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
Ministry of Education. (2011). *Ministry of Education Position Paper: Assessment (schooling sector)*. Wellington: Ministry of Education.
Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
NZQA. (2016a). *Languages clarifications*. Retrieved from http://www.nzqa.govt.nz/ncea/subjects/languages/clarifications/
NZQA. (2016b). *School-based assessment rules*. Retrieved from http://www.nzqa.govt.nz/about-us/our-role/legislation/nzqa-rules/expired-rules/assessment-including-examination-rules-2016/6/
Philp, J., Adams, R., & Iwashita, N. (2014). *Peer interaction and second language learning*. New York: Routledge.
Segalowitz, N. (2005). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 381–408). Oxford: Blackwell.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
Wellington, J. J. (2006). *Secondary education: The key concepts*. London: Routledge.
Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.

Chapter 12
Assessment Literacy as Praxis: Mediating Teacher Knowledge of Assessment-for-Learning Practices

Ofra Inbar-Lourie and Tziona Levi

Abstract This study investigates the evolving implementation of a praxis-based language assessment initiative as manifested in a number of school cultures. Drawing on Vygotsky's notion of praxis, wherein theory and practice are tied together and mutually substantiated, this research set out to probe assessment literacy development and practice among teachers in five schools in the periphery of Israel, following the delivery of a teacher in-service course on assessment literacy that was facilitated by one of the researchers. The study examined how and to what extent the knowledge and practice gained in the course were integrated into the schools' assessment culture. The sample included both coordinators of secondary school language departments and school administrators. The data set is comprised of interviews with participants probing changes in the school setting, specifically with regard to assessment knowledge, decisions, and practices, that may be attributed to the in-service course. A link was established between the schools' overall profile (characterized here as either *conservative* or *innovative*) and the extent of integration of course-acquired assessment culture. The study makes a case for differential assessment cultures and their match with local school cultures.

Introduction

This is a story of learning followed by praxis. It traces how acquired Assessment Literacy (AL) principles are performed on site. The settings are busy schools whose perceived mission is to mediate knowledge, encourage thinking, develop skills and

O. Inbar-Lourie (✉)
The Jaime & Joan Constantiner, School of Education, Tel Aviv University, Tel Aviv, Israel
e-mail: ofrain@tauex.tau.ac.il

T. Levi
The Jaime & Joan Constantiner, School of Education, Ministry of Education, Jerusalem, Israel
e-mail: Tzionale@education.gov.il

© Springer Nature Switzerland AG 2020

M. E. Poehner, O. Inbar-Lourie (eds.), *Toward a Reconceptualization of Second Language Classroom Assessment*, Educational Linguistics 41,
https://doi.org/10.1007/978-3-030-35081-9_12

instill values. The teaching force, as well as the schools' administrative staff, carry out this task within organizational cultures founded on a common epistemology, ethos, beliefs and codes of practice shaped by an interplay between individual and community ecosystems (Johnson 2008; Leonard 2011). The issue at hand is the question of reciprocity between theory and practice once a new concept is introduced to the school community. As Poehner and Inbar-Lourie argue in the introduction to this volume, this view of knowledge formation runs counter to models that position researchers exclusively as experts responsible for theory creation. As they explain, this reciprocal, mutually informative relation between theory/research and practice is known as praxis. In praxis, it is not the researcher who is solely responsible for the creation of theories and conceptual models nor is the teacher limited to putting these ideas into practice. Instead, theory and practice are elaborated as researchers and practitioners work together to bring models and principles into contact with actual contexts. This approach is exemplified in other studies in this volume where the practitioner and researcher form a partnership intended to promote mutual learning (see for example, chapters by Baker & Germain, Hill & Ducasse and Chan & Davison).

The focus of this particular story is on the process of implementing an assessment-based educational reform, but the issues we raise can be deliberated within a wider framework of educational initiatives in centralized educational settings and through what we regard as typical reform 'phases.' In the first such phase, the system identifies educational needs in view of new challenges, demographic or pedagogical circumstances or simply a change of guards among policy makers. A new educational policy is drafted to answer these needs. The policy is publicly announced and schools are expected to comply and implement the top-down reform. In the next phase, experts with appropriate academic and applied credentials are called in to articulate theory-based practice via professionalization channels. Following this initial *mediation stage*, a host of dilemmas and questions arise. Will the newly gained knowledge and values underlining the proposed approach (in the case of assessment literacy, the role of assessment in promoting learning) be adopted and integrated by the participating teachers, and will it be shared with colleagues, students and/or other community members? Furthermore, can such post-intervention effects be traced on the level of organizational culture (rather than only in individual praxis), and if so how are they manifested? What elements are embraced, rejected, or ignored in the daily routines and discourse of the institution? Does the field reshape and transform theory, and in what ways? Finally, when looking at different schools, can one detect differential willingness to accommodate the change and let it filter into the school's culture, i.e., does it affect some schools more than others? This is where the interest of the reported study lies.

Our story takes place in a centralized educational system in Israel distraught by external tests. However, recognition of the adverse effects of high stakes examinations is gradually filtering in (Shohamy 2001), with an evident move towards classroom-based formative assessment, similar to other educational contexts (Leung and Rea-Dickins 2007). In an attempt to bring about change and pave a more balanced approach to classroom assessment, and as part of a move towards what is termed "meaningful learning", a new assessment policy was declared in 2014 by the

Israeli Ministry of Education. The policy allows for partial internal school-based assessment at graduation, similar to other contexts such as Hong Kong (Davison 2007). Accordingly, secondary school teachers are allowed to independently choose and assess 30% of the curriculum content using a variety of assessment tools. The remaining 70% continues to be determined by the high-stakes external matriculation, for decades a contested tenet in the Israeli educational system. Although the decision was generally enthusiastically received, perceived to increase school autonomy, improve pedagogy and practitioners' status, it was also accompanied by confusion, mainly due to teachers' lack of knowledge and hence low self-efficacy in using different assessment measures as opposed to the familiar test formats. An array of professionalization initiatives was introduced in order to provide the teachers with the required assessment literacy competencies (The National Authority for Assessment and Evaluation 2016), the focus being on Assessment for Learning (AfL) and theory and practice of assessment (Assessment Reform Group 2002) similar to other educational systems (see East, this volume, on the New Zealand assessment policy). Here we follow the ripple effects of one such initiative, a blended assessment course (face-to-face and online) led by an assessment expert (one of the researchers), who was called in to instruct the teachers in five high schools to help them meet the new demand of creating school-based assessments.

Five courses, one in each of the schools, were conducted simultaneously, each lasting 8 months and comprising theory and practice in assessment. The course outcomes included analysis of assessment tools and the creation of both a performance task and a test intended for the teachers' future use. The task preparation process modeled formative assessment pedagogy whereby the tools (prepared by groups of teachers from the same school) were revised and improved following on-going feedback from the course lecturer and peers. The participants work in schools that are part of a chain of about 200 secondary schools within the Israeli educational system, and are situated mostly in the geographical and social periphery of Israel. The participating language teachers took part simultaneously in a language policy professionalization initiative intended to promote awareness of multilingualism through collaborative practice. We returned to the language teachers who had participated in the course 6 months after its completion and to other professionals in the schools, in an attempt to trace possible dialectic stages of learning development amongst the teachers and enactment of the intervention as part of the school praxis.

The methodology used was multiple-case study, which enables an understanding of the similarities and differences between the cases as well as analysis within and across situations (Yin 2009). Ten semi-structured interviews were held, two at each school, one with a Language Coordinator (English or Hebrew) and the second with a representative of the school administration (Principal, Vice-Principal, Pedagogical Director); in all, five Language Coordinators and five administrative staff were interviewed. The teacher interviewees were invited to reflect on the course content and on the extent to which the assessment beliefs, practices and activities they had been exposed to resonated in their reasoning, in dialoguing with others (colleagues, students, school administration), and in their teaching routines. The teachers and the administrative staff were also asked to reflect on the role of assessment in the school

in general, whether they could detect explicit and implicit changes in the school's assessment vision, policies, and in discussions on dilemmas that arose concerning this topic. In order to contextualize and situate the findings, we will first introduce the relevant terminology and concepts. These include school and assessment culture, assessment literacy (AL, the subject of the course) and Language Assessment Literacy (LAL) in the case of the language teachers. We will also consider the role of mediation in Vygotskian praxis.

School Culture and Assessment Culture

The five schools share some obvious common features: all can be described as situated in the periphery, geographically, socially and economically. All are part of the official Israeli educational system and are therefore required to follow its pedagogical regulations. However, the schools also belong to the large school chain referred to before, with its own vision, agendas and internal culture, and cater to the educational needs of nearly 100,000 students from diverse religious and ethnic backgrounds. The schools thus have a dual commitment, to the national curricula and especially to the matriculation school-leaving exams, as well as to a smaller school community administered by a common body that upholds social equity values, focuses on excellence, especially in science and technology, and on the promotion of learning for teachers as well as students. However, despite many shared large-scale features, each school constitutes a distinct community, the product of past and present traditions and ideologies and current norms and practices.

Deal and Peterson (1999) make a case for using the term "culture" to signify the elusive aura that schools and school communities have, defining "school cultures" as "complex webs of traditions and rituals that have been built up over time as teachers, students, parents and administrators work together and deal with crises and accomplishments" (Deal and Peterson 1999:4). Drawing on school culture research in various locations, they portray successful school cultures as those whose leaders establish and maintain school identity, support the students and teachers in their learning and can manage crises and seek future opportunities. The mixture of characteristics that makes every school culture unique needs to be considered when introducing educational innovations. This is to ensure compatibility between the sets of beliefs that underlie the change initiative and the existing school culture.

It stands to reason that a school's culture affects all school activities, including assessment. The assessment approach that a school chooses to take, whether a uniform test-driven one for assessing achievements or assessment as a means to advance learning, constitutes the school's *assessment culture*. Research findings show that the existence of school professional communities can account for the effective interpretation and application of assessment results within the school's routines (Birenbaum et al. 2011). Birenbaum (2014) takes the discussion a step further in applying complexity theory to consider the attributes that may facilitate or hinder *assessment for learning* (AfL) practices within schools. Focusing on the

reciprocal dynamic connection between two nested learning systems within the institution, classroom learning and teacher professional learning, she notes the cyclic evolutionary process in which change in teachers' beliefs is reflected in their practices and in their students' learning, with the on-going change recurrently feeding into the dynamic learning and enacted transformation of all parties. Schools where AfL change are likely to take place nurture an "assessment culture mindset" which shares the epistemological beliefs and values that underlie AfL as a learning-oriented assessment ecosystem, which upholds "a constructivist epistemology learning orientation, higher-order thinking, mutual trust and respect, caring, collaboration, dialogue, reflection, transparency and tolerance of errors" (Birenbaum 2014: 289). Conversely, implementation of AfL thinking and practices is less likely to prevail in institutions characterized by "positivist epistemology, external motivation, lower-order thinking, low sense of agency, mistrust, intolerance of errors, feelings of stress and frustration, etc." (Birenbaum 2014: 289). This latter school culture is more susceptible to external forces, and particularly to external tests. Recognition of the idiosyncratic nature of school cultures and their unique needs is part of a move to contextualize learning and assessment and situate them within school and classroom contexts, thus creating localized, culturally appropriate assessment "scripts" (Elwood and Murphy 2015). Fleer (2015) presents such an example describing how a group of teachers went through the arduous process of reappraising their assessment practices in seeking an assessment pedagogy that would be aligned with their pedagogical philosophy.

The on-line blended course on assessment that the teachers took was intended to provide them with conceptual understandings transformed into authentic assessment instruments, such as models and tools for analysis of performance-based tasks, tests and rubrics tools. Although the course established a general theoretical foundation, it emphasized the need to consider nested aspects of specific educational contexts – school, subject and learner. The content of the course was guided by the concept of assessment literacy (AL). Since the focus of the research is on language teachers, reference will also be made to a sub-category within AL – Language Assessment Literacy (LAL).

AL and LAL

The term Assessment Literacy (AL) is used to refer to the knowledge and skills assessors need in order to engage in the assessment process, "the knowledge of means for assessing what students know and can do, how to interpret the results from these assessments, and how to apply these results to improve student learning and program effectiveness" (Webb 2002*:* 1). Different perceptions of the central role of assessment, as part of the accountability movement on one hand, and emphasis on formative assessment in the learning process on the other (Black and Wiliam 1998), have had a major impact on how teachers' AL is conceptualized. Research on teacher-trainees' and in-service teachers' AL shows a meaningful lacuna between

the tasks teachers are expected to perform as assessors (both for interpreting and reporting test data and for using assessment for learning procedures), and their acquired competencies. This is not surprising, considering the inadequate opportunities offered to teachers to master AL on the pre- and in-service levels (DeLuca and Klinger 2010; Lam 2015).

Language assessment literacy is derived from the wider AL concept, and "forms the knowledge base needed to conduct language assessment procedures, that is, to design, administer, interpret, utilize, and report language assessment data for different purposes" (Inbar-Lourie 2016:1). Attempts to define the boundaries of such literacy are currently discussed in the language testing research community (see for example the *Language Testing Journal*, Special Issue, 2013 and the theme of the Language Testing Research Colloquium, 2017). The on-going debate focuses on the constituents of knowledge and skills required and their generic or language-specific aspects (Harding and Kremmel 2016), referring also to principles which guide valid, reliable, ethical and fair assessment processes (Davies 2008). Though differential LAL needs are acknowledged with reference to diverse stakeholders (Taylor 2013), most of the literature refers to teachers who are involved in assessment decision-making on a daily basis. The emerging data regarding teachers' LAL in different locations also demonstrates, however, a lack of basic skills in assessing language functions (e.g. Vogt and Tsagari 2014). Based on a survey of both AL and LAL research studies, Xu and Brown (2016) emphasize the importance of contextual macro-micro considerations in teachers' assessment literacy, including disciplinary knowledge and pedagogical content knowledge. They offer a dynamic practice-oriented approach to AL/LAL that includes the ability to decide on assessment methods from a repertoire of available tools, to be familiar with suitable grading methods, with score interpretation and with offering feedback to different audiences. Student activation in the process is a vital part of AL and LAL, as well as awareness of and competence in implementing the principles of ethics and fairness in assessment (Xu and Brown 2016).

The course reported on in this study included the above components presented generically for teachers of various subjects, initiated by the school managements to familiarize the entire teacher staff with AL. The course tasks, however, provided an opportunity for disciplinary practice and consideration of the local school context. Given that the creation of assessment cultures within the school setting can thrive on teachers' integrated professional and practical knowledge, but also on assimilating on-site experience, we will now look at the role of mediation in facilitating such development.

Mediation and Praxis

Relying on a Vygotskian perspective, mediation is recognized as an essential instrument for cognitive change, critical to the generation of higher mental processes such as voluntary attention, intentional memory, logical thought and problem-solving

(Lantolf 2000). Mediators, whether people, objects, or symbols, help transform spontaneous impulses into higher mental processes. The mediation process, according to Vygotsky (1978, 1987), can occur during any interaction, not necessarily in a formal educational setting, although Vygotsky recognized a special value to the kinds of opportunities for cognitive development available through schooling. Hence, with regard to the school context, knowledge is not monopolized by external experts but is an evolving entity jointly created with existing knowledge and ongoing practices.

With regard to theory and practice, Vygotsky challenges us to understand these as a dialectic in which practice is no longer perceived just as the application of theory but rather as a unified equal partner in the creation and transformation of knowing and doing. This notion of the theory-practice dialectic is reiterated by Lantolf and Poehner (2011), who stress that while theory provides the foundation for practical activity, research into teachers' actual classroom practices is essential in order to refine, extend, and thus reshape the theory. In the same vein, understanding is conceptualized as dialectical in nature, combining consciousness (knowledge and theory) with action that results in the creation of an object (Lantolf 2008).

Following Vygotsky, there have been advances in the formation and internalization of concepts as resources or tools with which individuals may regulate their thinking and acting (Galperin 1989; Davydov 2004). Accordingly, new conceptual knowledge that is detached from its context of use will not suffice, and pedagogies that value explicit knowledge must include learning activities that link this knowledge to action. Theory and practice are seen to form a dialectic that dynamically exerts a reciprocal influence with the classroom contributing to the elaboration and development of knowledge. Teachers thus have a highly agentive role, not just as technicians who automatically apply theoretical principles, but also as literate knowledgeable experts whose activities and experiences refine and redefine existing theory (Kinard and Kozulin 2008) and their practice changes accordingly.

Mediation can take different forms, as is documented in this volume. For example, in the chapter by Michelle and Davison, we learn of an online resource system informed by the experiences and practices of teachers in Australian schools and also available to them as ongoing mediation. In the chapter by Hill and Ducasse, the authors describe a teacher assessment literacy (TAL) tool designed for promoting reflection on assessment feedback practices. In the case of the present chapter, mediation is materialized through a professional, blended on-line assessment course. We point out, however, that mediation is not limited to the use of a specific tool and can be perceived on different levels. This is captured by the concentric circles in Fig. 12.1 below. The first act of core direct input mediation transpired in the course itself with the academic expert mediating assessment theory and practice to the teachers in their role as learners. However, what preceded was the Ministry reform followed by the Network's intervention in providing support, connecting the schools and helping with integration of Ministry of Education policy. This encompassing layer created a framework to enable the Ministry's reform, applying internal high-stakes assessment. The second level where mediation meets praxis occurred as the course participants returned to their schools and engaged in reflection and in

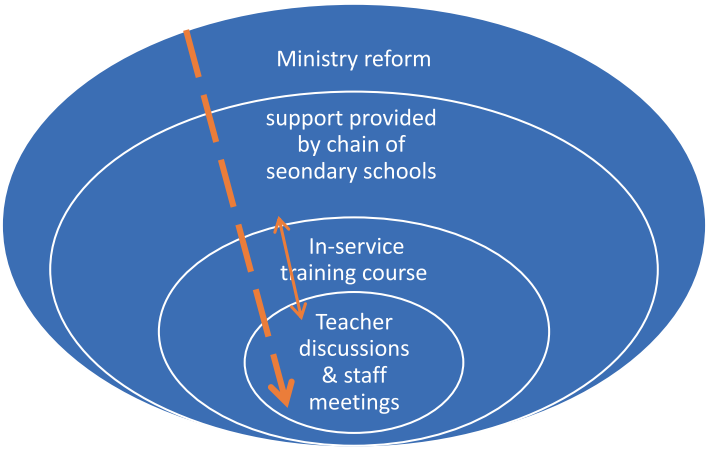


Fig. 12.1 Levels of mediation

classroom activities, mediating the knowledge to colleagues and students, enriching and reshaping their understanding through classroom experience and assessment practices. Teachers focused on the course contents in staff meetings and discussions and assimilated the outside expert input to the local school culture, norms and intrinsic beliefs to fit the local setting, "to inculcate the ability to acquire such knowledge and to make use of it" (Vygotsky 1997:339). Ultimately, at the end of the process, it is the individual teacher who mediates the learners' engagement in the learning activity to make it a source of development.

Research Design

Because school cultures are unique entities, it was assumed that integration of the AL principles and methods acquired in the professionalization initiative would take on different forms in the discourse and practice of each school community, depending on their idiosyncratic pedagogical and administrative features. Based on the literature regarding assessment and school culture, the research question posed was whether variation related to school culture would be detected among the schools in their mediated integration of the course content.

Ten semi-structured interviews were held with school administrators and coordinators of language departments. The interviews, conducted in Hebrew or English, focused on AL in the school culture with regard to the following issues: the organizational and assessment culture of the school; school vision and general pedagogic policy, including degree of teacher autonomy; the school decision-making process; assessment in the school context and its compatibility with the above variables; and the interviewee's vision regarding assessment. Finally, the participants were asked

to reflect in semi-structured interviews on the impact of the assessment literacy course on different facets (organizational, discursive and pedagogical) in the school.

Analysis of Findings

Thematic Categories

The verbal comments generated by the interviews were transcribed, translated from Hebrew when necessary by the researchers, analyzed and coded for emerging themes using a grounded theory approach (Strauss and Corbin 1998). Themes were identified and categorized first as those shared by the whole school sample, and then with reference to individual schools. The themes that emerged were found to correspond with perspectives previously identified as accounting for the reasons why Israeli EFL teachers use or refrain from using alternative assessment tools (Inbar-Lourie and Donitsa-Schmidt 2009, based on a model by Hargreaves et al. 2002). These perspectives were therefore used as categories of analysis in the current study as well. They consist of the following themes: (a) the *technological* perspective, referring to time allotment and availability of resources for conducting assessment, as well as to the teachers' assessment expertise: the extent to which teachers have the necessary knowledge and skills, i.e., assessment literacy, for conducting assessment. The technological aspect also includes possible gaps between what parents expect in the area of assessment (uniform measurement-based tests and grades), and alternative means of assessment; (b) the *cultural* perspective, which relates to the degree to which alternative assessment principles and practices form part of teachers' belief systems and the school culture. This relates to degree of agreement with understandings on the on-going nature of assessment and its integration within teaching learning activities, and what such beliefs entail regarding the role of teachers and learners in the assessment process; (c) the *postmodern* perspective, which critiques the value of assessment beliefs in the postmodern era, and the quality measures of different assessment tools (e.g., portfolios), especially since some are compiled outside the classroom, casting doubt on their authenticity; and (d) the *political* factor, that examines the influence of external top-down accountability systems, particularly high stakes external tests on teachers' assessment practices.

School Profiles

A general description of each of the schools was obtained from the Network's language administration based on continuous interactions and the schools' involvement in other professional development initiatives. The school descriptions were used to create profiles, depicted according to correspondence with four features

identified as conducive or detrimental to "assessment culture mindedness" (Birenbaum 2014): (a) inclination towards developing, adopting and enacting change and pedagogical innovations; (b) centralized versus decentralized administration; (c) collaborative teacher learning communities; and (d) the dominant assessment culture (whether test-focused, emphasizing summative assessment or formatively-oriented). Schools demonstrating an overall profile of reluctance to change, a centralized rather than decentralized administration, where individual work was more prevalent than collaborative modes, and where test-focused pedagogical orientations reigned with a strong emphasis on the matriculation exam, were termed "conservative" or "traditional". Schools that were described as inclined to innovation that functioned in collaborative teacher communities as part of a decentralized administration and upheld a culture of assessment for learning, were termed "innovative". A "school culture continuum" ranging from "innovative" to "traditional" was created with the schools positioned according to this classification. Two schools were placed on the conservative end of the pole, and two on the innovative one, while the fifth school, whose attributes could be classified as conservative as well as innovative with regard to certain aspects, was placed in the middle. The study aimed to see if the degree of mediated integration of the AL contents from the four aforementioned perspectives (technological, cultural, post-modern and political) corresponds with the positioning of each school on the continuum. That is, whether links can be established between the school's general profile and its willingness to adopt and integrate an "assessment culture mindset" through analysis of reports by school administrators and language teachers.

Findings Common to All Schools

All the Language Coordinators who participated in the course, regardless of their school's conservative or innovative classification, attested to having gained meaningful knowledge in various facets of assessment by attending the course, which in turn led to a reconsideration of existing practices. This was evident especially in the cultural category of themes and principles related to assessment: "The knowledge provided in the course was encompassing and varied", and "this knowledge led to rethinking processes and sometimes changed the way we evaluate students" (Language Coordinators).

However, despite their newly gained understandings, all the respondents alluded repeatedly to the dominant impact of the top-down policy dictates within which they operate: the uniform program, the yearly work plan prescribed by the discipline inspectors, and most of all the matriculation exams. This political factor of an external top-down effect on assessment and learning was mentioned repeatedly and perceived negatively as severely limiting autonomy, or in one of the interviewee's words, "there is no breathing space" because "the criteria measurements are locked" with no room for teachers' decision making. This despite the decentralized reform, which grants a certain level of teacher autonomy, and led to the professional

development initiative. Neither the teachers nor the administrators felt any major change in terms of the control exerted on the schools, and all emphasized the test-focused environments: "The Bagrut [matriculation] hovers over everything" (Pedagogic Administrator). One of the interviewees stated that if it was up to her she would cancel tests altogether due to the stress they induce, and introduce more open-ended projects, but "we depend on the general educational system" (Language Coordinator). With regard to the quality of the suggested alternative assessment tools, some of the participants were critical of their psychometric aspects in comparison with the familiar and reliable test format, which replicates the matriculation exam.

All the same, a new assessment literacy discourse began to be evident, with teachers acquiring the "alternative assessment language" (Principal). This was triggered by the top-down reform and especially by the ensuing course, perceived as essential in solidifying the cultural change from an external, test-centered orientation to teachers and schools independently initiating curriculum and assessment moves. The school Principals who were aware of the teachers' AL needs were keen on having their staff participate in the course:

> The course facilitated. Pushed. Triggered discourse in the staff room on the topics of learning and assessment. The teachers understood the significance, though they complained. Without the course, the big change could not have happened. That is why it was so important for me to hold it.

And from one of the Language Coordinators:

> The course improved the procedures. We know exactly where we're headed. How to construct rubrics, how to evaluate certain topics. There is clarity in the tasks. We work according to objectives. The language of assessment is clearer. When speaking about a rubric you know what is meant. Also performance tasks – the terms are clear, as are the expectations.

The technological perspective emerged as crucial in this study as well as in the previous one, where it was found to best predict EFL teachers' alternative assessment use (Inbar-Lourie and Donitsa-Schmidt 2009). Teachers brought up issues of time management, heavy workload, as well as their lack of understanding and appreciation of the value of introducing other forms of assessment: "Implementing anything but tests is full of difficulties and incompatible with school reality" (School Administrator). The change also presented difficulties in encounters with parental expectations: "Parents also have problems with all this creativity. They're wondering: Are we in primary school?"

Language-related issues of the assessment process specific to LAL hardly emerged in the Language Coordinators' comments, the only references focusing on the assessment format for English literature, whether a personal student log or a test, the options allowed by the Ministry of Education. Interestingly**,** however, all the research participants mentioned the role of language as a vital component in the school assessment culture, stressing that all disciplinary teachers need to be involved in the language assessment process not only the language teachers. This cultural value-laden principle can be directly attributed to a previously mediated initiative of on-going language awareness, which also emphasized the need for inter-disciplinary

cooperation among language teachers and other disciplines, that the Network's language teachers participated in.

In summary, the course content was appreciated by all participants and administrators in the different schools regardless of their orientation and despite differences in the perceived utilization of the concepts learned. In addition, all schools attested to the major impact of the external test on the school culture in general and on the assessment culture of the school in particular. However, findings also showed differences among schools, as detailed below.

Differences Among Schools

Input from the administrative staff as well as from the Language Coordinators reflects variability among schools. Findings will be reported according to the emergent themes. Often the themes tend to intertwine, and so are reported here according to the dominant theme which emerged.

The Technological Theme

As was found previously (Inbar-Lourie and Donitsa-Schmidt 2009), the technological aspect, referring to resources and to teachers' AL, was frequently discussed by the interviewees. Change in the more conservatively inclined schools was viewed by one of the language coordinators as "mission impossible", voicing despair at the enormity of the task with regard to the resources needed in all respects – time, professional knowledge, collaboration and students' and parents' perceptions. The difficulties encountered can also be attributed to the epistemological cultural platform of differential learning and assessment and 'assessment for learning' principles, with some of the participants opposing the values of the framework offered: "It's a mental state that needs to be changed. The vision of alternative assessment is not anchored in reality", said one Language Coordinator. The recurring theme of difficulties encountered, however, can also be seen as indicating a willingness to embrace the concepts, despite the ideological and practical hardships. This was more evident in the two schools that are more positively inclined towards innovation and the creation of an assessment culture: "There are difficulties, but assessment is perceived as a process.… The Hebrew language department is attempting to deal with the changes from a number of perspectives. This facilitates the implementation despite all the difficulties involved" (Hebrew Language Coordinator**).** The term "process", often used to portray AfL principles, was applied here to describe the present state of affairs leading to an assessment change. The tone is milder and accepting, despite the overwhelming technological difficulties. In addition, the participating teachers in the more assessment-minded schools felt optimistic about implementing the change due to the professional assessment literacy acquired as part of the course.

This realization emerged also with reference to professionalized test design: "I learnt how to plan the tasks, to work according to objectives also when designing tests" (Language Coordinator).

Two additional features repeatedly referred to in the respondents' testimonies expanded the technological theme or perspective (Hargreaves et al. 2002), apparent more in schools with an innovative profile. The first is the significance of the "team" – the collaborative subject or school professional community that faces the challenges of change in a mutually trusting and supportive relationship, professionally as well as personally: "We have a type of familiarity, we are like a family, involved in different things, but we cooperate with each other", says one of the Language Coordinators. The group acts as a mediator for the change, with team members sharing materials, setting objectives and holding on-going consultations. The learning also occurs among different teams: "I learnt from the other teams' presentations", says one of the language coordinators with reference to the course. Though innovative schools emphasized this point more intensely, it was evident to a certain extent in the other schools, especially within the language team. The second notable feature is decentralization, again typical of innovative schools. This was modeled on different levels: on the administrative level, where the Principal consults with the teachers; on the level of the disciplinary teams, where the Coordinator initiates a joint decision-making process, filtering down to the students' level as the school aims for "increasing learner responsibility and improving independent learning skills" (Pedagogical Administrator).

The Cultural Theme

The epistemological cultural change in terms of how assessment is viewed permeated into schools differentially, from superficial adoption to a more profound organizational cultural change, even though all schools still reported on change in assessment practice as a lengthy process. In schools with a more conservative culture the interviewees' comments reflected the realization that the change has not yet become an integral part of the school culture and is still restricted to the individual teacher or dependent on the specific disciplinary team, despite top-down encouragement within the organization:

> [There is] no peer learning. The departments don't want to collaborate. It's not natural for the veteran teachers to participate. It's difficult, not because of the management – they encourage it. (Language Coordinator)
>
> The professional development gave us tools to carry out the assessment. Each of the teams took it to its own place…. Unfortunately it is still at the team level. (Language Coordinator)

Moreover, the expertise gained in the course was often perceived in these schools as being tool-oriented, e.g., the use of the rubric as a tool.

Conversely, a notable difference can be detected in comments related to the scope of change in the more innovative schools, encompassing the entire school

community and reflecting both a cultural and a technological perspective. In these schools, changes to the assessment culture were perceived in relation to broader school community changes:

> We have discussed this over the last four years. A different kind of teaching. It starts in junior high. We spoke about different teaching and assessment. In the beginning we talked about assessment – how does it impact your teaching? … The teacher is not the source of knowledge… the culture is in the making. We discussed values and skills. (Principal)

The change is not limited to specific subject areas but is evident throughout the school in collaborative planning, as noted by one of the Language Coordinators: "The team is totally invested in carrying out the objectives. With some of the teams we have reached a merging of objectives for a number of teams".

One of the recurrent themes among the more innovative schools was their expressed desire to embrace assessment as part of the change in school culture and align their goals accordingly: "We wrote a new school manifesto that corresponds to the current reality… Transparency… dialogue… teacher autonomy… belief in the process" (Pedagogical Coordinator). In line with this, other forms of assessment were legitimized and integrated into the school routine: "Even the test schedule now mentions other forms of evaluation from grade 7 to 11" (Principal), a technological change that reflects a deeper cultural realization and adoption of principles. The change affected lesson formats and the level of complexity, pointing perhaps also at using higher order thinking skills: "Lessons have become more complex… require more accurate planning" (Language Coordinator). The change was evident in the school culture, indicating a new assessment phase: "For years we had only tests. When they started talking about alternative assessment, we had a period of meaningful learning, not just to pass the exam. Since the reform, followed by the course, alternative assessment is everywhere" (Language Coordinator).

There are still difficulties involved but the attitude is constructive: "I haven't yet made the connection between the course and my work. It's still patchy. In the test assignment, I choose just a few item types and not all kinds. But it was an experience that gave me an advantage", reported a language coordinator from the school positioned in the middle of the continuum. However, reliance on tests was still found in these schools as well, with the exception of the assessment reform: "Tests form most of the evaluation – only the topics included in the reform are assessed via other tools" (School Administrator).

The Post-Modern Theme

This perspective emerged in instances where the interviewees either cast doubt on the quality and authenticity of assessment tools other than tests or reported on this disposition among students and their parents, who were concerned about divergent grading. The dividing line among schools with regard to this perspective was not always distinct. Language Coordinators in more innovative schools were more open

to differences of opinion among the team members, though from the comment below it is not clear whether they viewed this diversity positively or not: "My teachers are not always coordinated in terms of grades because of differences in characteristics, views and education" (Language Coordinator). In terms of parents' and students' acceptance of the new culture, representatives of both school types voiced concern: "Students complain that grading is not uniform" (Language Coordinator). Teachers themselves struggle with uncertainties as to the quality of the assessment tasks produced: "I feel that there is more focus on quantity and less on quality… They [the students] do charming things but in their own time and not as part of skill development in the classroom as part of the curriculum" (Language Coordinator).

Doubts were raised also about the validity of alternative assessments: "The students feel that alternative assessment is like summer camp… Even if it is meaningful and serious, from their point of view it isn't… Students interpret it as freedom not to study and it makes life difficult for the teachers" (Language Coordinator). Yet another Language Coordinator believes that alternative assessment is not conducted properly, leading to undesirable results: "As an educator and as a mother [of school-age children] I feel that alternative assessment has become a burden… a punishment because of the work load. Students learn to hate assignments and that's a shame". Another observation noted that alternative assessment does not reflect students' true abilities, and that students take advantage of the situation: "Sometimes students exploit this situation, so assignments must be done in class with supervision" (Language Coordinator).

The Political Theme

As noted previously, all school cultures and assessment routines were meaningfully impacted by external political forces, especially by the end-of-school exams, directives and formats. The Ministry of Education was the most notable external power referred to, especially with regard to the matriculation exam but also to the power Ministry Inspectors hold in authorizing programs in the different subject areas. The degree of intensity with which these were mentioned by the representatives of the different schools varied however. Top-down phenomena representing the political perspective were mentioned more frequently in the testimonies of teachers and administrators in schools less inclined towards change. In commenting on the school's attempts to create alternatives in assessment, one of the Language Coordinators notes "at the end of the day everything follows the Bagrut [matriculation] requirements."

Little mention was made of the other top-down player, i.e., the group of schools, referred to above as the 'chain' or a Network of secondary schools and its officials, as determiners of assessment policy. Since the Network is committed to and follows Ministry policy, its intervention is less noticeable, perceived as interchangeable with that of the Ministry: "I don't make decisions on my own. The information is fed in from the Ministry's Inspector or from the Network Pedagogical Director to

the subject Coordinator… Certain things are dictated" (Principal). Either one or the other external authority decrees the policy and learning or assessment specifications. The same Principal attests to the fact that in the initial phases of the reform the dual voices of these powers caused "difficulty and chaos".

As was previously noted, more innovative school cultures seemed to attach more value to teacher autonomy: "Teachers have the autonomy to act as they see fit [re assessment practices]" (Language Coordinator). However, even in the most innovative schools that autonomy is dependent on the professional Inspector from the Ministry: "Certain subject areas were not provided with latitude for decision-making" (Pedagogical Administrator).

Discussion

This chapter began with a story of mediated changed praxis following a policy reform. We joined the plot at a certain point in time, 6 months after the intervention ended to take stock of what the protagonists had to say about the mediated process. The interview data enabled an insightful understanding of the impact of situated practice, of school life, its ecology and culture, and the integration of change. The findings demonstrate that although the direct intervention (i.e. the teacher professional development blended course) was identical for all five schools, which moreover shared the same educational macro-environment (i.e. the Israeli school system), it created a different localized effect in each school, an effect that was mediated by each school's existing culture. Institutions more oriented towards change tended to reorient their assessment culture more willingly in belief and practice than schools with a more centralized administration culture that were less pedagogically innovative and more test-driven. The picture that emerges is not dichotomous, however, indicating that certain themes, in particular the institutions' susceptibility to external policy (the political theme) was common to all schools regardless of their culture. It was found that even in the face of reform teachers still relied on external examinations, a cultural ecology that prevailed even though it was at odds with the new situated assessment culture of learner-focused and learning-embedded assessment offered in the course. Despite the reform initiative, schools could not be coerced to induce change that was incompatible with their overall culture. Hence what emerges is the need to consider changes from within, and to construct tailored assessment processes comprising values and norms compatible with the epistemological system of the school community that will be integrated gradually. To be more specific, a staff of teachers in one school, although sharing the same course contents and seemingly the same assessment reform guidelines and policy, still adheres to the self-developed and individualized assessment practices that are part of the school culture. For example, in one of the five schools, the course contents were mediated in groups of teachers working in teams and sharing assessment products, whereas in another school teachers preferred designing their assessments individually and met the final requirements as distinctive entities. This was contrary to

the course task requirements, which opted to model and mediate team work in the different faculties encouraging peer observations and feedback.

The differential reaction to and adoption of a change from above is often dependent upon the interpretation of individual teachers, who engage in decision-making and context adaptation based on their perceptions and assessment knowledge, a phenomenon demonstrated by East (this volume) in the context of a reform in a French speaking examination in New Zealand. Policy makers with an interest in working towards assessment change need to empower the school community to engage in discussion as to their preferred assessment mode and shared assumptions and norms of behavior, a discussion that may also yield consideration of different routes and novel experiences.

With regard to the focus of the professional development, all the course participants expressed favorable views about the course content. However, their responses at the time of the research signaled that despite their agreement with some or most of the beliefs and practices advocated, there were different ways of adapting the ideas to particular assessment cultures (Cowie and Moreland 2015). Thus, when agents of change with a similar agenda merge with different school cultures the result is not uniform and can yield a variety of cultures rather than a single one. In addition, the road to change is lengthy and complex, and doubts and difficulties arise even in schools whose profile shows a strong dedication to embracing change. This suggests that in order to accommodate change more time and support are required, with continued on-going dialogue within a community of practice of researchers and teachers.

In terms of the mediation process, participant testimonies demonstrated diverse mediation channels that were utilized in the change process. The initial mediation in the form of a course provided by an expert did not suffice, though it served as the basis for extended on-site mediated activities. The limitation of the course may have resulted from adherence to a more conventional knowledge formation tradition, which positions the researcher as expert rather than taking a praxis-oriented approach of mutual learning (Poehner & Inbar-Lourie, this volume). At any rate, the subsequent on-site activities were comprised mostly of consultations with colleagues, both from the same and different subject areas. With regard to the language-orientation of the department heads or to special language assessment considerations and ensuing literacy i.e., LAL, the discourse exemplified no such trace seen to be generic. No reference was made, for example, to special language-oriented assessment considerations, such as the knowledge needed to teach and perform evaluations of spoken and written language. This may be explained by the fact that the intervention itself did not focus on subject-related assessment. It contradicts, however, the definition of AL as domain-specific (Xu and Brown 2016), and highlights the scarcity of and need for specialized assessment professionalization training and on-going support for language teachers in the unique features of language assessment as part of pedagogical content knowledge in the context of assessment change.

To conclude, this study corroborates previous research showing that change in assessment culture is contingent upon local settings, and on taking account of the many stakeholders in the schools' ecosystems. The insights gained point to the need

to rethink the practices, goals and intended outcomes of mediated practice at sites where praxis transpires. Specifically, in the field of assessment, harmonious relations between school and assessment cultures need to be established through mutual learning and support provided to schools in the process of changing their assessment identity.

References

Assessment Reform Group. (2002). http://www.nuffieldfoundation.org/assessment-reform-group

Birenbaum, M. (2014). Conceptualizing assessment culture in school. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *The enabling power of assessment: Vol. 1. Designing assessment for quality learning* (pp. 285–302). Dordrecht: Springer.

Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation—Special Issue on Assessment for Learning, 37*(1), 35–48.

Black, P., & Wiliam, D. (1998). *Inside the blackbox: Raising standards through classroom assessment*. https://www.rdc.udel.edu/wp-content/uploads/2015/04/InsideBlackBox.pdf

Cowie, B., & Moreland, J. (2015). Leveraging disciplinary practices to support students' active participation in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*(2), 247–264. https://doi.org/10.1080/0969594X.2015.1015960

Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly, 4*(1), 37–68.

Davydov, V. V. (2004). *Problems of developmental instruction. A theoretical and experimental psychological study* (Peter Moxhay, Trans.). Mosow: Akademia Press.

Davies, A. (2008) Textbook trends in teaching language testing. *Language Testing 25*(3), 327–347

Deal, T. E., & Peterson, K. D. (1999). *Shaping school culture: The heart of leadership*. San Francisco: Jossey-Bass/Wiley.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education, 17*(4), 419–438. https://www.tandfonline.com/doi/pdf/10.1080/0969594X.2010.516643?needAccess=true

Elwood, J., & Murphy, P. (2015). Assessment systems as cultural scripts: A sociocultural theoretical lens on assessment practice and products. *Assessment in Education Principles Policy and Practice, 22*(2), 182–192. https://doi.org/10.1080/0969594X.2015.1021568

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural–historical perspective. *Assessment in Education: Principles, Policy & Practice, 22*(2), 224–246.

Galperin, P. I. (1989). Mental actions as a basis for the formation of thoughts and images. *Soviet Psychology, 27*, 45–64.

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (Handbooks of applied linguistics) (Vol. 12). The Hague: Mouton de Gruyter.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69–95.

Inbar-Lourie, O. (2016). Language assessment literacy. In E. Shohamy (Ed.), *Encyclopedia of language and education* (S. May (Ed.). Volume on language assessment) (pp. 257–270). Springer.

Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of EFL. *Assessment in Education: Principles, Policy & Practice, 16*(2), 185–204.

Johnson, E. S. (2008). Ecological systems and complexity theory: Toward an alternative model of accountability in education. *Complicity: An International Journal of Complexity and Education, 5*(1), 1–10.
Kinard, J. T., & Kozulin, A. (2008). *Rigorous mathematical thinking: Conceptual formation in the mathematics classroom*. New York: Cambridge University Press.
Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169–197.
Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching, 33*, 79–86.
Lantolf, J. P. (2008). Praxis and L2 classroom development. *ELIA: Estudios de Linguistica Inglesa Aplicada, 8*, 13–44.
Lantolf, J. P., & Poehner, M. (2011). Dynamic assessment in the classroom: Vygotskian praxis for L2 development. *Language Teaching Research, 15*(11), 11–33.
Leonard, J. (2011). Using Bronfenbrenner's ecological theory to understand community partnerships: A historical case study of one urban high school. *Urban Education, 46*(5), 987–1010.
Leung, C., & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly, 4*(1), 6–36.
National Authority for Assessment and Evaluation. (2016). *Evaluation of the "meaningful learning program": Students', teachers' and principals' perceptions*. Ministry of Education (in Hebrew). http://meyda.education.gov.il/files/Rama/Lemida_Mashmautit_Report_fm.pdf
Shohamy, E. (2001). *The power of tests*. Harlow: Pearson.
Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). London: Sage.
Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing 30*(3):403–412.
Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402.
Vygotsky, L. S. (1978). Mind in society. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Vygotsky, L. S. (1987 [1934]). *Thinking and speech* (N. Minick, Trans.). In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L.S. Vygotsky. Volume 1: Problems of general psychology* (pp. 37–285). New York: Plenum Press.
Vygotsky, L. S. (1997). *Educational psychology*. Boca Raton: Nova Science.
Webb, N. L. (2002, April). *Assessment literacy in a standards-based urban education system*. Paper presented at the AERA annual meeting, New Orleans, LA. http://facstaff.wcer.wisc.edu/normw/aera%202002/alignment%20analysis%20three%20states%20math%20final%2031502.pdfwww.wcer.wisc.edu/archive/mps/AERA2002/Assessment%20literacy%20NLW%20Final%2032602.pdf
Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162.
Yin, R. K. (2009). *Case study research: Design and methods*. Thousand Oaks: Sage.