The Enabling Power of Assessment 7
Series Editor: Claire Wyatt-Smith

Margaret Bearman · Phillip Dawson
Rola Ajjawi · Joanna Tai
David Boud *Editors*

Re-imagining University Assessment in a Digital World

Springer

The Enabling Power of Assessment

Volume 7

Series editor

Claire Wyatt-Smith, Faculty of Education and Arts, Australian Catholic University, Brisbane, Queensland, Australia

This series heralds the idea that new times call for new and different thinking about assessment and learning, the identities of teachers and students, and what is involved in using and creating new knowledge. Its scope is consistent with a view of assessment as inherently connected with cultural, social practices and contexts. Assessment is a shared enterprise where teachers and students come together to not only develop knowledge and skills, but also to use and create knowledge and identities. Working from this position, the series confronts some of the major educational assessment issues of our times.

More information about this series at http://www.springer.com/series/13204

Margaret Bearman • Phillip Dawson
Rola Ajjawi • Joanna Tai • David Boud
Editors

Re-imagining University Assessment in a Digital World

Springer

Editors
Margaret Bearman
Centre for Research in Assessment and Digital Learning
Deakin University
Melbourne, VIC, Australia

Rola Ajjawi
Centre for Research in Assessment and Digital Learning
Deakin University
Melbourne, VIC, Australia

David Boud
Centre for Research in Assessment and Digital Learning
Deakin University
Melbourne, VIC, Australia

Phillip Dawson
Centre for Research in Assessment and Digital Learning
Deakin University
Melbourne, VIC, Australia

Joanna Tai
Centre for Research in Assessment and Digital Learning
Deakin University
Melbourne, VIC, Australia

ISSN 2198-2643 ISSN 2198-2651 (electronic)
The Enabling Power of Assessment
ISBN 978-3-030-41955-4 ISBN 978-3-030-41956-1 (eBook)
https://doi.org/10.1007/978-3-030-41956-1

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

Foreword

As this book moves into final production, most educators are in the midst of shifting all assessment into online mode, due to the COVID-19 pandemic gripping the globe. Prompted by the unprecedented fear of loss of life in modern times, the entire education sector has lurched into the digital. Assessment traditions hundreds of years old have been set aside. Exams and practical assessments are being reworked so that they can be meaningful for our disembodied, digitally mediated world. We believe that in the years to come, this unsought but seismic shift in people's understanding of digital learning, teaching and assessment will have untold side effects.

This is a time for reimagining. The danger is that in our rush to convert our practices from embodied to digital, that we will simply replicate what has been done. The opportunity is that a whole cohort of educators and students – not the early adopters or tech whizzes – will contribute to the shape of digital learning by sheer necessity. And through this, our sense of what is normal will have changed. We may well discover ways of working with assessment that we have not thought of before.

We hope that we emerge from this crisis – and tragedy – with new ideas about assessment, which take full advantage of digital possibilities. This book offers directions, examples and conundrums to provoke new ways of thinking about assessment in what has become an inescapably digital world.

Melbourne, VIC, Australia
March 2020

Margaret Bearman
Phillip Dawson
Rola Ajjawi
Joanna Tai
David Boud

Acknowledgements

Heartfelt thanks to Mr. Kevin Dullaghan who provided outstanding administration and always kept us on track and to Dr. Helen Walker who organized the 2017 symposium with flair.

Contents

Contributors

Chie Adachi Deakin Learning Futures, Deakin University, Melbourne, VIC, Australia

Rola Ajjawi Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia

Tiffani Apps School of Education, University of Wollongong, Wollongong, Australia

Aneesha Bakharia UQx, Institute for Teaching and Learning Innovation, University of Queensland, Lucia, QLD, Australia

Margaret Bearman Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia

Karley Beckman School of Education, University of Wollongong, Wollongong, Australia

Sue Bennett School of Education, University of Wollongong, Wollongong, Australia

David Boud Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia

Andrew Cain Deakin University, Melbourne, VIC, Australia

Doug Clow Open University UK, Milton Keynes, UK

Linda Elizabeth Corrin Learning Transformation Unit, Swinburne University of Technology, Hawthorne, VIC, Australia

Joanne Dargusch School of Education and the Arts, Central Queensland University, Noosa, QLD, Australia

Phillip Dawson Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia

Chris Edwards Open University UK, Milton Keynes, UK

Mark Gaved Open University UK, Milton Keynes, UK

Lois Ruth Harris School of Education and the Arts, Central Queensland University, Rockhampton, QLD, Australia

Trina Jorre de St Jorre Deakin University, Melbourne, VIC, Australia

Yoon Jeon Kim MIT Playful Journey Lab, Cambridge, MA, USA

Simon Knight Faculty of Transdisciplinary Innovation, University of Technology Sydney, Ultimo, NSW, Australia

Siva Krishnan Deakin University, Melbourne, VIC, Australia

Rosemary Luckin UCL Knowledge Lab, University College London, London, UK

David Marshall Faculty of Arts and Education, Deakin University, Melbourne, VIC, Australia

Sandra Kaye Milligan Assessment Research Centre, University of Melbourne, Parkville, VIC, Australia

Marcus O'Donnell Deakin University, Melbourne, VIC, Australia

Abelardo Pardo UniSA STEM, University of South Australia, Adelaide, Australia

Edd Pitt Centre for the Study of Higher Education, University of Kent, Canterbury, UK

Peter Reimann The University of Sydney, Centre for Resarch on Learning and Innovation, Sydney, Australia

Bart Rienties Open University UK, Milton Keynes, UK

Jekaterina Rogaten Open University UK, Milton Keynes, UK

Louisa Rosenheck MIT Playful Journey Lab, Cambridge, MA, USA

Joanna Hong-Meng Tai Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia

Laura Tubino Deakin University, Melbourne, VIC, Australia

Naomi Winstone Department of Higher Education, University of Surrey, Guildford, UK

Part I
Introduction

Chapter 1
Introduction

Phillip Dawson, Rola Ajjawi, Margaret Bearman, David Boud, and Joanna Tai

Abstract This brief introductory chapter outlines how this book, *Reimagining University Assessment in a Digital World,* explores the implications, opportunities and challenges of the digital world for assessment in higher education.

This is a book about the implications, opportunities and challenges of the digital world for assessment in higher education. It has been produced as a collaboration between international researchers, catalysed by a 2017 symposium at the Centre for Research on Assessment and Digital Learning (CRADLE) at Deakin University in Melbourne, Australia. At that symposium, we set a challenging brief for our colleagues: to fundamentally reimagine the intersection of assessment and digital learning in higher education.

When we say university assessment, we refer to a broad and inclusive conceptualisation. Assessment serves multiple purposes, and this book emphasises the need for assessment to prompt and sustain learning as much certifying achievement. While the authors in this book take a range of views about assessment and focus on various purposes or aspects, these align with the broadest view: assessment as making judgements about what someone is capable of, based on some sort of demonstration or product. This ranges from high-stakes examinations to low-stakes formative tasks, and it includes judgements made by educators, students, their peers, and others. Assessment is therefore also concerned with the feedback information that is produced as consequence of these judgements.

When we say the digital world, we mean a world where the digital is pervasive and all around us. This isn't a naïve 'digital natives' style argument; we don't think students are magically wonderful with technology, nor that we should pander to some imagined desire for high-tech learning. We do however recognise that life in general, and education in particular, has been changed as our world has become more digital.

P. Dawson · R. Ajjawi · M. Bearman (✉) · D. Boud · J. Tai
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: margaret.bearman@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_1

The digital world offers a range of technologies that can enable us to do things we've always wanted to do with assessment at a massive scale. It also enables us to assess differently, or make sense of masses of data to understand the efficacy of existing approaches. This book explores these and other potential benefits of the digital world for improving assessment.

The digital world also poses fundamental challenges not just to how we assess, but to what we assess. As artificial intelligence advances, computers are increasingly able to undertake tasks that were once the domain of humans. As students engage with social media, they become exposed online in ways that previously only applied to media figures or celebrities. This poses challenges for assessment, which are explored in this book.

Assessment in a digital world is thus a broader concept than the narrower concept of 'e-assessment' that tends to occupy the intersection between technology and assessment. E-assessment is largely concerned with technologized assessment, and at its best it leverages what new technologies can do to enable new forms of assessment that are better for learning and judging what students can do. While there is benefit in making assessment digital, we think that there are much greater opportunities and challenges posed by the digital world for assessment. Some of these may not even involve technologized assessment at all; instead they may involve preparing learners for tasks that are distinctly human. These are concerns beyond the tradition of e-assessment, and they require a broader set of disciplinary voices. This book features chapters by scholars from the disciplines of philosophy, computer science, media studies and psychology (to name but a few) – as well as assessment and educational technology.

This book is broken up into five parts; this chapter and the next comprise *Part I: Introduction*. You can read the chapters in any order, however we strongly encourage you to read the next chapter first. That chapter provides some foundational concepts about how assessment can or should change to suit the digital world. It situates the question of university assessment within broader debates about the place of technology in the university and the place of the university in a digital world.

Part II: The changing role of assessment in the digital world asks what shifts are required of assessment in order for it to be fit for purpose in the new digital world. Two of its chapters deal with the ethics of re-imagined assessment: one with a focus on the new and emergent ethical issues posed by the digital world (Bearman et al., Chap. 3), and the other revealing challenges of inequity that persist despite the digital world (Harris and Dargusch, Chap. 8). Keeping with the theme of persistent challenges, this part also explores how feedback might not actually have been reimagined as it has been technologized (Pitt and Winstone, Chap. 7). This part also includes two chapters that focus on artificial intelligence: one which focuses on the different roles people and computers will play in assessment, and what really needs to be assessed in a world of artificial intelligence (AI) (Bearman and Luckin, Chap. 5); and another which asks how student use of AI tools should be incorporated into assessment (Dawson, Chap. 4). A final theme in this part is how assessment spans

the boundaries between learners' digital work, study and social worlds. How learners portray themselves online, and the role assessment has in mediating this, is the focus of one chapter (Ajjawi et al., Chap. 6); this complements a focus on how assessment can represent the realities of digital design work and shift away from the textual bias so present in current assessments (O'Donnell, Chap. 9).

Part III: The role of big data in reimagining assessment asks how data can be used to improve assessment. A view of big data as 'augmenting' assessment and feedback practices is presented in this part (Knight, Chap. 10), in which analytics support (but not supplant) the human players in assessment. The next chapter asks how the huge datasets generated by online students can be used to identify problems with assessment (Rogaten et al., Chap. 11). The following chapter identifies ways technology can broaden the repertoire and scaling possibilities for assessment, and how it can help assessors observe learning processes (Pardo and Riemann, Chap. 12). The part ends with a chapter on 'metrolytics', a combination of educational measurement and learning analytics approaches (Milligan, Chap. 13). Where the previous part was focused on possibilities of what could or should be (as well as what should not be), this part explores the bleeding edge of what is possible with data and assessment.

Part IV: Practical exemplars provides a grounded snapshot of what expert educators are already doing in reimagining assessment, and what is possible with current technology. It begins with a chapter focusing on self and peer assessment in self-paced online courses (Corrin and Bakharia, Chap. 14). This is followed by a chapter exploring what is currently possible in self and peer assessment, including a table showing the capabilities of current tools (Tai and Adachi, Chap. 15). The next chapter gives an exemplar of a reimagining of assessment in support of constructive alignment that would only be feasible through technology (Cain et al., Chap. 16). One implication of the digital world on assessment has been increasingly open-ended tasks, and the next chapter in this part explores the challenges this poses for educators and students (Apps et al., Chap. 17). This is followed by a chapter documenting how assessment can be not just reimagined, but redesigned through play (Kim and Rosenheck, Chap. 18). This part ends with a chapter on how the representation of assessment can be reimagined through digital credentials (Jorre de st Jorre, Chap. 19). Common throughout this part is a focus on what is possible, through practical exemplars educators can explore.

The book concludes with *Part V,* a chapter that builds on common threads from throughout the book to establish an agenda for future work on assessment in a digital world.

As you read this book, we would like to issue you with the same challenge we issued the chapter authors: to broaden your thinking about the intersection of technology and assessment beyond just e-assessment toward fundamentally reimagining assessment in a digital world.

Chapter 2
New Directions for Assessment in a Digital World

Margaret Bearman, David Boud, and Rola Ajjawi

Abstract Assessment exists within a series of pedagogical, administrative and technological legacy practices. It tends therefore to reflect the needs and concerns of a previous time. However, this does not align with a digitally enabled world with rapidly expanding information and an increasingly dynamic view of knowledge. This chapter explores how to reimagine assessment in a time of digital change. Firstly, it investigates how assessment designs can draw on technology to move beyond the status quo, providing an example of a programmatic e-portfolio. Next, it introduces the idea that assessments should enable graduates to work in the digital world. This is discussed through an illustration of how students can build effective digital personas, including social media profiles, within their assessments. Finally, it looks to the inter-relationship between assessment and knowledge in a digital world, suggesting co-construction as a guiding principle. Often assessment is overlooked in discussions of educational technology. Reimagining university assessment is a key way to make teaching and learning applicable to a digital world.

2.1 Introduction

Traditions cast long shadows. No matter when you were educated, you would find many aspects of today's typical university assessments familiar, if not identical, to those held within your alma mater. Even if you attended university in the 1960s, you would find that many of the current undergraduate exams and the essays are still similar to the ones completed in your youth. However, many things have changed in the last 50 years outside of university assessment. Teaching and learning is being transformed. The collective content of the academy has exponentially expanded, particularly in the last two decades. The advent of many different technologies

M. Bearman (✉) · D. Boud · R. Ajjawi
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: margaret.bearman@deakin.edu.au; david.boud@deakin.edu.au; rola.ajjawi@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_2

means that work and lifestyles found in the post-World War II period are distant memories. So it remains curious that even though universities may use digital technologies to conduct assessment, it is often in the same form as in previous generations, albeit more efficiently, at distance and at scale.

The digital era we live in has been called the 'fourth industrial revolution' (Schwab 2016, p. 26). There is no question that computing power and digital connectivity have fundamentally changed our daily lives and our civic institutions. There has never been so much information available to so many, nor so many opportunities to disseminate individual views. The digital technologies that underlie these changes have much to offer university assessment. However, as described, traditions are remarkably persistent and new tools have tended to reinforce the same educational practices (Selwyn 2014, 2016). While exciting innovations exist, these mostly occur outside of mainstream educational practice. The dominant modes of university assessment remain much as they were: demonstrations of knowledge, such as quizzes, tests or essays, which can be easily compared and graded. In a time where it is easy to access information rapidly and accurately, our assessments appear to be somewhat missing the point. They too often require a high degree of recall and offer little opportunity for student input or choice. Our overall impression is, in higher education, the digital has locked in an old set of ideas about assessment.

New paradigms of assessment are helping chart new directions, but these conversations can be absent from most implemented educational technology systems. Contemporary views, familiar to most, suggest that high quality assessment should promote *learning* through demonstrated attainment that meets *standards* (Boud 2017). Therefore well designed assessment requires well designed feedback practices that support diverse student trajectories throughout their courses of study and across their working life (Bearman et al. 2016). However, there is little discussion regarding how these contemporary views of assessment can mesh with the now ubiquitous digital technologies, educational or otherwise. Assessment is often overlooked in educational technology conversations despite its pivotal role within the academy. As we explore later, institutions and educators reveal their values both by *what* is assessed and *how* (Knight 1995). We suggest that significant questions remain largely unaddressed: How can technology support forms of assessment that emphasise learning? How can assessments support living and working in a digital era, with its networked and rapid access to both people and information?

This chapter considers what reimagining assessment might look like in a time of digital change. In order to leave behind unhelpful legacies, we need to create a space to imagine what more we can do. Otherwise, our assessments and hence our universities may become increasingly misaligned with the digital world in which we all live and work. We commence by taking a closer look at what Selwyn calls the "digital university": institutions enmeshed in both mundane and innovative technologies at a time when "historical changes in the processes of production of knowledge [are] characterised by high velocity and dizzying excess…" (Dahlgren 2018). We then offer three types of reimagining that can help shape new ways of assessing and hence learning in this digital world. We firstly consider *reimagining the status quo:* how we can design our assessment processes better to make use of the technology

we currently have. The second type of reimagining explores *reimagining new assessment designs that take account of the digital world*. Finally we consider *reimagining assessment designs for new ways of knowing.*

2.2 The Need to Reimagine

Technology has always been a part of human existence and technologies have both shaped and been shaped by social relations. In this sense, humanity swims in a sea of technology and for the most part we are blind to it. However, new digital tools have profoundly changed higher education: imagine a university without electronic journal access, electronic time sheets and learning management systems. Moreover, university students are constantly engaging with their own digital worlds (Aagaard 2015); students' digital selves are necessarily omnipresent during their studies. When they graduate they will likely draw on artificial intelligence (AI) to supplement work and leisure activities. University staff, students and alumni alike live and work in the digital economy. So, higher education cannot avoid digital technologies, irrespective of whether deliberate attempts to harness technology to improve education are successful.

So how are universities coping in this new digital era? Two scholarly works discuss how universities can situate themselves in the emerging technological landscape. The first deliberately situates in a kind of realpolitik pessimism to explore how technology and the university mutually shape each other. The second orients itself within a productive optimism and concerns itself with the role of the university during a time of epistemic crisis. Both illuminate the challenges facing the university due to the rise of digital technologies.

In his 2014 book, *Digital Technology and the Contemporary University: Degrees of Digitisation,* Selwyn interrogates the current status of technology within the university. He writes: "the pessimistic commentator [accepts] that digital technology is *not* bringing about the changes and transformations that many people contend…" (p. 18). He calls for "the very difficult conversations that need to take place about what digital higher education is, and what digital higher education should be." From Selwyn's perspective, technology is not neutral: it serves a variety of different and often competing agendas. He notes, as we have, that digital technologies tend to maintain the status quo. Most of the book dissects the often grim realities of digital technologies in our universities and outlines a range of ways in which administrative and educational technologies "reproduce long-standing and deep-rooted structures and arrangements" (p. 141). However, in closing, Selwyn suggests we need to *reimagine* higher education in this digital era. He calls for changes in how we talk about educational technology, the physical spaces surrounding technologies, and the need for guiding principles which outline moral rules of engagement.

The second scholarly work of interest is Barnett and Bengtsen's 2017 paper: *Universities and Epistemology: From a Dissolution of Knowledge to the Emergence of a New Thinking.* In contrast to Selwyn's pessimistic stance, the authors "argue

that an optimistic university is a university-in-the-world and thinks the future from the world." However, this call for optimism is set against an acknowledgement of how the digital age is shifting the notions of knowledge and truth: "…with the onset of digital reason, perhaps a new kind of truth value is emerging. What is in evidence in this regime is unstable truth, flickering truth, pictorial truth, and truth as digital picture, always likely and instantaneously to be replaced by another digital picture of the world" (p. 6). If the very notions of truth and knowledge have changed, then this necessarily impacts upon institutions whose core purpose is knowledge production. The authors therefore suggest that the university has a mandate to assist broader society to make sense of digital knowledge and truth. In line with both this and Selwyn's work, we suggest that reimagining assessment in a digital university is about considerably more than technology. It is about new forms of truth and knowledge and what this means for our students' learning, our own pedagogical practices and the place of the university in society.

We offer three different approaches to reimagining assessment pedagogies. We firstly consider how educators might redesign status quo assessment, with a particular focus on reimagining administrative systems more than pedagogy or technology. The second line of exploration considers the kind of assessment needed to take account of living in a digital world; we provide an example that focusses on the types of skills necessary to manage a professional social network. The final approach considers the place of assessment pedagogies in a broader context of changing truth and knowledge in a digital age, using the example of institution-student co-construction of assessment.

2.3 Reimagining the Status Quo

What is the pedagogical work that we want assessment to do but which seemed impossible prior to technological enablement? As educators and course designers, we have been shaped by administrative systems that were originally bounded by the constraints of paper communication. However, many of these constraints now exist purely by organisational convention. This can limit possibilities both for what we can do and what we can imagine as possible. Assessment within a university occurs within a series of administrative as well as pedagogical legacy systems. For example, systems may require a mark out of 100, even when this is invalid or no longer appropriate. This is in part because universities have highly regulated themselves and assessment has many (sometimes unnecessary) policy restrictions. This has created layers of bureaucracy, which are slow to shift. This is further compounded by some assessments being reified in technologies, such as Learning Management Systems (LMSs), which themselves become legacy systems almost as soon as they are rolled out into practice.

This is fundamentally a question of redesign and focusses on bringing together what it is we currently wish assessment to do, the affordances of the latest technology and looking beyond the limitations of current administrative structures. There

are many potential examples of this. We consider one in particular: electronic submission of work that spans a course and how we could improve assessment pedagogies as a consequence of this ubiquitous change in university assessment.

Example: Reimagining Marks Through the Electronic Learning Portfolio Students once had to physically submit assignments and collect the marked copies, usually from departmental offices, with marks manually entered into ledgers. Now electronic submission of assignments and submission of grades is taken-for-granted. This enables: submission by students at any time and in any place; composing and delivering comments without faculty or students having to go to a physical location; and grades and comments which are kept as a permanent record, avoiding loss of work. Pedagogically, this represents almost no change from a paper-based system. The gains have been efficiency-based: electronic submission greatly adds to the convenience of all parties and potentially saves each considerable time and effort on travel and handling. In addition to convenience, there are also organisational benefits: in the case of queries, dates of submission and return are precisely recorded and there is a permanent record of both student work and tutor response. However, this very capacity to electronically submit work and comment and retain it have far reaching pedagogical implications, which we are yet to take full advantage of.

At the simplest level, the tracking of students' performance occurs through the collection of marks for different assignments, their weighting and averaging for each course unit and the assembling of these across a course, ultimately leading to a final student transcript. However, this administrative system has not yet taken account of courses that are standards-based, which is now required by most organisations. Instead, the legacy recording of grades and a grade point average provides normative data when criterion-based data is needed (Boud 2017); these marks do not represent attainment of a particular standard but rather a questionable statistical representation across a collection of non-commensurate tasks. If courses and course units are organised around course and unit learning outcomes, then it is attainment relative to these that need to be tracked. Currently, we use grades, which in themselves are merely weighted averages of marks on discrete assessment activities, with no account taken of whether they can be meaningfully combined. Dressel's (1957, p. 6) critique of grades from 60 years ago still stands: "… an inadequate report of an inaccurate judgement by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite amount of material."

Once performance on learning outcomes for each course unit can be recorded, then it becomes possible for both students and educators to see progress towards overall learning outcomes across a program of study. This is entirely possible at present. Digital tools, such as ReView™, enable students and tutors to track performance for every criterion on every assessment task for an entire degree program and for these to be displayed visually for students both in terms of their tracking towards course-level learning outcomes while also providing feedback information for each task (Thompson 2016). Thus, students can direct their attention to those program

learning outcomes most in need of development. This form of tracking also enables self-assessment to be readily incorporated and for students to receive feedback information on their judgements as well as their substantive performance over time (Boud et al. 2013).

This is clearly not a significant *technological* challenge. Many contemporary LMSs can be configured to allow much better tracing of assignments and feedback information than is possible with paper-based approaches. For example, students and educators should be able to readily see information provided to a student on previous assignments. Therefore, what is most urgent is to conceptualise how assessment can be constructed over an entire program. While pedagogy is the focus, this also involves an administrative reimagining – how might we think of assessment that spans a program rather than a unit? How can assessment be reimagined in order to provide coherence to the student about what they are trying to achieve and how? The technological fix is trivial compared to the imaginative one. We illustrate both this triviality and complexity with the example of the course-wide electronic learning portfolio.

Learning portfolios are based on an old idea that has renewed merit. The keeping of portfolios by students is a long-standing practice in the expressive disciplines, but they have now been taken up across the board. At their most basic, portfolios are repositories of student work curated for various purposes. They are used to aid learning as well as to demonstrate progress, for example, through the approach known as a patchwork text in which students weave together existing written work collected as a new work to form a capstone achievement (Winter et al. 2003). Originally, these were physical portfolios, often overly voluminous and difficult to manage. Such portfolios are rapidly being displaced by those stored electronically (albeit with material objects sitting behind images where necessary), often within an LMS.

Such electronic portfolios or e-portfolios can take a variety of forms but they are, generally, digital repositories of student work, some of which has been validated by formal assessment and some of which has not. The collection permits additional opportunities for students to manipulate, comment upon and curate materials contained in it. Secondary processing is much easier within the digital space: students can select from readily available and previously assessed work, which demonstrate how they have met learning outcomes. This kind of secondary processing is the key to the e-portfolio and where redesign can be maximised. In addition to students selecting material that meets learning outcomes, they can also assemble representations of their achievements that they can curate and annotate for different purposes: applying for a job or gaining a research scholarship or developing a digital persona. They can add new meta-cognitive texts to these curated objects, which include thoughts in relation to the artefacts. In this way students can access and monitor their own work over time, both to observe their own progress in achieving greater competence in a particular professional area, to reflect on it, and to gain insight into what further work is needed to attain particular outcomes (Clarke and Boud 2018). However, the main barrier to the enhanced use of eportfolios is not the technology, nor educational design. It is the procedural structures of the university (such as

modularised self-contained units of study) which impede coherence across courses. A programmatic approach to the e-portfolio would allow a degree of student *control across their course* through their own choices in sourcing and manipulating representations of their work across their program. It also allows students to track their own progress over time, without reliance on *marks*.

This example illustrates the challenges facing redesign. In many cases, the technology is available but also requires thoughtful pedagogy and most importantly, ways of rethinking how assessment is conducted beyond a single course unit. This means we need reimagined administrative and governance structures.

2.4 Reimagining Assessment to Take Account of the Digital World

Society exists within a constantly connected and partially automated digital network. Traditional assessment designs, which have tended to emphasise knowledge and recall, may no longer be suited to this dynamic and networked digital world. This is emphasised by the focus on the so-called twenty-first century skills such as problem-solving, teamwork and communication (see for example Griffin and Care (2014)). Application of knowledge rather than recall is deemed more critical in an era where information is so readily available. There is an opportunity to look beyond accustomed assessment designs to reimagine assessments for entirely new ways of thinking and working. For example, our assessments should take account of our smartphone tools or the algorithms that already guide our thinking through search engines. We provide one particular example here: the need to portray a digital self and the implications for assessment.

Example: Reimagining Assessment to Incorporate Constructing Personas Creating a digital presence allows individuals to represent themselves within their various personal and professional online communities and it is increasingly critical that individuals and employees have the skills to present digital personas. As is developed further in Ajjawi, Boud and Marshall (Chap. 6, this volume) a digital persona can be conceptualised as a series of portrayals within various social networks. As with any celebrity or *linkedin* profile, these representations are both real and constructed. A persona is not merely a digital presence, it is the performance of the person within and to the social network.

This notion of deliberately constructing a persona within a social or professional network prompts different kinds of assessment designs. One such design could invite students to portray their self, image and achievements in digital assessment to different audiences, therefore changing what they choose to include in order to align with expectations of the particular network. Such strategic constructions of persona can be encouraged to distinguish individual student accomplishments rather than to

conceal their achievements through aggregation within a grade. Constructed personas represent achievement in a way that a transcript cannot.

Assessment designs that support persona development could encourage students to engage with different social media platforms across the span of their degree. This might start with limited tasks for smaller audiences, then expand in complexity and across wider networks as the student nears graduation. Consider the portfolio design described in the previous section. There are a number of social media performance artefacts that could contribute to the patchwork. In early years, students could produce a short video oriented to their colleagues talking about themselves, expectations of the degree and aspirations for their future. As they progress through their degree, students might 'follow' leading figures on Twitter and thereby identify the kinds of groups they wish to associate with professionally. They might analyse social media conversations to identify what matters to such groups and the ways in which the leaders and other represented there are positioned within it. This may prompt critical thinking with respect to joining such conversations. Closer to the end of their degree, students could present their own personas, emulating relevant others' digital constructions on applications such as LinkedIn. The portfolio could be selectively opened for feedback comments from peers, educators and/or industry partners.

Another way of thinking about personas, is how the assessment artefacts within a portfolio represent the student to particular audiences. To give a discipline-specific example, an education student might develop a teaching plan utilising professional craft knowledge and their own teaching philosophies. The plan will then form part of their portfolio, engaging feedback information from peers and the educator. As students then undertake a placement later in their degree, they introduce the revised plan to their supervisor, discuss its applicability and then enact it in the placement. Reflection on the enactment and the students' developing teaching philosophy might then be constructed as another blog post for the next work-placement, which could be made open to comment to a wider audience. Yet another part of the portfolio could include students' reflections on how they use different hashtags in communicating with their networks, what aspects of self are advanced and why. While these artefacts do not necessarily constitute a 'profile', they nonetheless represent a digital presence.

This iterative interaction with digital artefacts, industry and academia distributed through the portfolio enables students to construct different personas as they develop and reconstruct their achievements for different audiences. As illustrated, the influence of different networks on students' identity construction and who has access to what can be explicitly discussed and reflected on through the design of the assessment tasks. The use of modelling – both as a pedagogy and a practice – can become explicit and distributed. This alerts students to how personas are created within workplaces and encourages them to take account of the 'rules of the game'. By encouraging students to portray themselves as part of their assessment they become more agentic, although still bounded by sociocultural constraints, and therefore can begin to influence their positioning for future opportunities. This allows them to

graduate with greater awareness of their role in the strategic construction of self and what they can offer their professional communities.

2.5 Reimagining Assessment for New Ways of Knowing

The previous examples illustrate that assessments prescribe more than *what* students should know, they also denote *how* they should come to know as well as showing *what* they know and can do with this knowledge. In the nineteenth century, knowledge was seen as a relatively static truth, and assessments reflected this (Perry 1968). In the twentieth century, there was a shift to multiple ways of knowing and relative truths; assessments reflected this societal shift (Perry 1968). Now, in the early twenty-first century, views of knowledge and truth are also changing with the rise of social media, algorithms and instantaneous information (Barnett 2017). This has implications for universities and university students. While the previous sections have described how reimagined assessments might take account of new technologies and emerging digital practices, this section explores the dynamic relationship between knowledge and assessment before illustrating the implications for assessment in a digital world.

Assessment exerts what Tan (2004, p. 654) calls "epistemological power", as it presents a particular view of knowledge and the students' relationship to that knowledge. Compare the patchwork portfolio with the multiple choice question examination, which is still highly emphasised in some disciplines.

> …By utilizing multiple choice questions, assessors are predetermining that the knowledge in a given field can be demonstrated through restrictive means and that students should not or need not be given the option to express their fuller views on the said question. This confines the student to a closed interaction and reinforces the idea that someone else knows the answers to the question, so original interpretations are not expected. The teacher exercises epistemological power over the student by stipulating a fixed number of outcomes the student can consider and by insisting that only one of the options is valid. (Tan 2004, p. 656).

Underlying the multiple choice format – and often the typical compare-and-contrast essay – is a singular premise: the student passes the assessment when they reproduce the correct way of knowing. However, as mentioned previously, this is often (but not always) an ill fit for our current era, where information can be plentifully provided with a single click. The patchwork portfolio or activities that involve persona construction demand that students must manipulate presentations of knowledge, which is arguably a highly important skill for the digital age.

These examples draw attention to the possibly outdated pedagogical assumptions that underlie our assessment processes. In the current era, we contend that assessments should help students navigate dynamic knowledge and truth structures which are the hallmark of the digital world. Next, we provide an example of an assessment pedagogy suited to the digital world to illustrate our contention: co-construction of assessment in criteria.

Example: Co-construction of Assessment Criteria Different perspectives on co-construction exist. However, two key tenets of co-construction are collaborative problem solving and a shared problem space. Both activities require constant negotiations and expansion of meaning (Reusser and Pauli 2015). This process is inherently fragile and often asymmetric, none more so than during assessment activities where power dynamics come most into play. To do it well, pedagogy needs to position students as proactive thinkers, ones who make judgements about the quality of their own work and progress rather than relying on unilateral judgements of their teachers (Tai et al. 2018).

Co-construction promotes: knowledge as dynamic; educators as knowledgeable but not all-knowing; and students as capable of making their own judgements about quality of work. For example, when students work with educators to develop criteria for assessment, such as co-constructing a rubric (Kilgour et al. 2019), it may provide an opportunity to understand how notions of quality are constructed (Bearman and Ajjawi 2018). However, these types of activities are less common in technology-mediated assessment. This is an irony: the pedagogies we most need to account for the digital world are the ones least likely to be presented in online only spaces and are limited to single unit solutions. However, co-construction of assessment criteria could also occur in virtual spaces. There is no reason that small online discussion groups can work on identifying criteria, which are then presented to the wider cohort. These could be the source of debate and reflection for both students and educators through a series of online activities. Finally, a mix of voting and educator selection could identify final criteria. While this might be logistically challenging within a large cohort, it is achievable through the affordances of digital technology. Through doing this, these types of pedagogies become accessible beyond single units and can become integrated into course designs at a larger level.

As this example illustrates, co-construction of assessment allows students to contribute to shared understandings of knowledge. This provides students with an opportunity to navigate learning and knowledge in ways that will serve them well throughout their lives. Meaning is derived through dialogues; the mutuality of exchange forms a key part of living and working in the digital world. It is about who is convinced as much as what is said; the act of persuasion changes both you and those around you.

2.6 Conclusions: Assessment Reimagined

In this chapter we have sketched a range of reimagining that underpins ways we can think about assessment. As with all works of the imagination, what we propose necessarily builds on what is currently possible. Indeed, there are examples in all universities that have come a good way towards many of the ideas presented within our examples. It remains a challenge to shift from the small change within a unit, to a broader change across courses. However, it is increasingly imperative that we

reimagine how we design and administer our assessments. Without doing this, our assessments will become increasingly irrelevant to the digital world.

References

Aagaard, J. (2015). Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & Education, 87*(Supplement C), 90–97. https://doi.org/10.1016/j.compedu.2015.03.010.

Ajjawi, R., Boud, D., & Marshall, P. (This volume). Chapter 6: Repositioning assessment-as-portrayal: What can we learn from celebrity and persona studies? In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world*. Dordrecht: Springer.

Barnett, R. (2017). *The ecological university: A feasible utopia*. London: Routledge.

Barnett, R., & Bengtsen, S. (2017). Universities and epistemology: From a dissolution of knowledge to the emergence of a new thinking. *Education Sciences, 7*(1), 38. https://doi.org/10.3390/educsci7010038.

Bearman, M., & Ajjawi, R. (2018). From "seeing through" to "seeing with": Assessment criteria and the myths of transparency. *Frontiers in Education, 3*(96). https://doi.org/10.3389/feduc.2018.00096.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the Assessment Design Decisions Framework. *Teaching in Higher Education, 21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217.

Boud, D. (2017). Chapter 2: Standards-Based Assessment for an Era of Increasing Transparency. In D. Carless, S.M Bridges, C.K.Y Chan, & R. Glofcheski. (Eds.). *Scaling up Assessment for Learning in Higher Education*, pp. 19–31. Singapore: Springer.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education, 38*(8), 941–956. https://doi.org/10.1080/02602938.2013.769198.

Clarke, J. L., & Boud, D. (2018). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International, 55*(4), 479–486. https://doi.org/10.1080/14703297.2016.1250664.

Dahlgren, P. (2018). Media, knowledge and trust: The deepening epistemic crisis of democracy. *Javnost-The Public, 25*(1–2), 20–27. https://doi.org/10.1080/13183222.2018.1418819.

Dressel, P. (1957). Facts and fancy in assigning grades. *Basic College Quarterly, 2*, 6–12.

Griffin, P., & Care, E. (2014). *Assessment and teaching of 21st century skills: methods and approach*. Dordrecht: Springer.

Kilgour, P., Northcote, M., Williams, A., & Kilgour, A. (2019). A plan for the co-construction and collaborative use of rubrics for student learning. *Assessment & Evaluation in Higher Education, 45*, 1–14. https://doi.org/10.1080/02602938.2019.1614523.

Knight, P. (1995). Introduction. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 13–24). Birmingham: Kogan Page.

Perry, W. G. (1968). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.

Reusser, K., & Pauli, C. (2015). Co-constructivism in educational theory and practice. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (Vol. 3, 2nd ed.). Oxford: Elsevier.

Schwab, K. (2016). Four leadership principles for the fourth industrial revolution. Retrieved from https://www.weforum.org/agenda/2016/10/four-leadership-principles-for-the-fourth-industrial-revolution

Selwyn, N. (2014). *Digital technology and the contemporary university: Degrees of digitization*. London: Routledge.

Selwyn, N. (2016). *Is technology good for education?* Cambridge: Polity.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Tan, K. H. K. (2004). Does student self-assessment empower or discipline students? *Assessment & Evaluation in Higher Education, 29*(6), 651–662. https://doi.org/10.1080/0260293042000227209.

Thompson, D. G. (2016). Marks should not be the focus of assessment—But how can change be achieved? *Journal of Learning Analytics, 3*(2), 193–212. https://doi.org/10.18608/jla.2016.32.9.

Winter, R., Parker, J., & Ovens, P. (2003). The patchwork text: A radical re-assessment of coursework assignments. *Innovations in Education and Teaching International, 40*(2), 112–122. https://doi.org/10.1080/1470329031000088978.

Margaret Bearman is Research Professor within the Centre for Research in Assessment and Digital Learning, Deakin University. Margaret's research focusses on learning and teaching in higher and professional education. She is particularly interested in ways of improving feedback, the sociomateriality of standards, and what it means to be an online learner in a digital world.

David Boud is Professor and Foundation Director of the Centre for Research in Assessment and Digital Learning, Deakin University, Professor of Work and Learning at Middlesex University and Emeritus Professor at the University of Technology Sydney. He has been a pioneer in developing learning-centred approaches to assessment across the disciplines, particularly in student self-assessment, building assessment skills for long-term learning and new approaches to feedback.

Rola Ajjawi is Associate Professor in Educational Research at the Centre for Research in Assessment and Digital Learning, Deakin University where she leads a program of research into learning and feedback in workplace settings. She is Deputy Editor of the journal *Medical Education*.

Part II
The Changing Role of Assessment in the Digital World

Phillip Dawson

This section deals with the fundamental shifts in assessment that are afforded, required or made inevitable by the digital world. It deals with questions that are of interest to practitioners, researchers and society at large: what should we assess in the age of artificial intelligence? Who owns the outputs of digital assessment? Is technology just providing a repackaging of traditional educational practices? How can we bridge the online worlds of education and the professions? These are questions that go beyond a technological determinist view of assessment and educational technology. Rather than just asking "what cool things can we do with technology?", this set of chapters examines the broader social, ethical and pedagogical implications of the digital world.

There are four sub-themes in this section. The first deals with ethical matters. The second addresses feedback. The third, artificial intelligence. And the fourth, matters of digital practices and portrayals.

The Ethics of Re-imagined Assessment

In society more broadly, the digital world has both led to new ethical challenges and the exacerbation of existing ethical challenges. This pattern applies to assessment as well, perhaps even more so, due to power differentials, high stakes, and the significant time invested by educators and students. Bearman, Dawson and Tai (Chap. 3) offer an exploration of the moral, ethical and social dimensions of assessment in this new world. Their chapter prompts us to question how much agency learners and educators really have in digital assessment, and who owns assessment products when platforms of education are corporatized. They also raise concerns about diversity, including the 'invisible' barriers that educational technology presents to learners. This last point is further expanded in Harris and Dargusch's chapter, which draws on a study with low socioeconomic status undergraduate students at an Australian university (Chap. 8). They argue that assessment inequity existed before, and it appears to persist in the digital age due to three key causes. Firstly, student

self-efficacy with respect to assessment appears to still be a powerful factor in the digital world. Secondly, the digital world has not obviated the challenges of preparedness for study amongst diverse cohorts. Thirdly, students continue to have a diverse range of other roles outside the university, and these still place pressure on their ability to participate in assessment. Taken together, these two chapters urge us to not just focus on the new ethical challenges posed by the digital world, but to also carefully consider any claims that old challenges have been technologized away.

Feedback in a Digital World

Being both the site of immense dissatisfactions, and a key driver for learning, feedback has been the topic of much educational technology research and development. But how far have we really come? Pitt and Winstone (Chap. 7) focus on the subset of the technology enhanced feedback literature that studies audio, video and screencast feedback. They identify three dilemmas. Firstly, they ask if these new forms of feedback are really about enabling feedback dialogues, or if they are largely about replicating existing monologic feedback practices. Secondly they ask if this technologizing of feedback is driven by pedagogy, or practical and logistical realities. Thirdly, they ask if these new feedback modalities really lead to improvements in student learning, or if they are largely just a way to improve student satisfaction. While their chapter focuses on feedback, its fundamental questions are useful more broadly in considering assessment in a digital world, as they question why we do what we do, and how we evaluate our practices.

Artificial and Human Intelligences in Assessment

In a world where people are legitimately concerned that robots will take their jobs, we must ask: will robots take our assessments? These two chapters consider what the role of humans and computers will be as artificial intelligence becomes increasingly sophisticated and able to take on roles previously done by people, both teachers and students. Bearman and Luckin (Chap. 5) begin by exploring and defining what artificial and human intelligences are, and what they are good at. They provide examples to show the roles computers can play in assessing student learning, and what sorts of learning still need to be assessed in a world where computers can do so much. They provide an exemplar assessment task that requires students to demonstrate uniquely human capabilities, and conclude by arguing that artificial intelligence may force us to focus on what really counts in assessment. This theme is taken up by Dawson (Chap. 4), who focuses on cognitive offloading, which occurs when people use physical actions to make mental tasks easier; this can be as simple as writing something down to remember later, or it can involve more sophisticated technologies like translation tools or equation solvers. The chapter asks if students

should be allowed to use cognitive offloads in assessment, and if so, how they should be incorporated into assessment designs. Four principles for cognitive offloading in assessment are proposed. Firstly, that learning outcomes should be scoped with respect to the cognitive offloads allowed. Secondly, that lower-order outcomes should be allowed to be offloaded once mastered. Thirdly, that capability with cognitive offloading should be considered a desirable learning outcome. And finally, that students should learn how to evaluate the quality of the products of cognitive offloading. The view presented by these two chapters is not one of a robopocalypse for assessment in a digital world, instead, it is one of optimism and consideration for what makes humans special.

Assessment as Digital Portrayal and Digital Practice

In an era of social media and technologized work, the boundaries between learners' work, study and social worlds are becoming blurred. This sub-theme argues that the digital world demands a re-casting of assessment: as an act of portrayal, and as a digital practice. Ajjawi, Boud and Marshall's chapter argues that digital assessment places great demands on how students present themselves (Chap. 6). Drawing on celebrity and persona studies, they suggest that learners strategically negotiate the versions of themselves they present in different contexts, including in assessment. Students need support to navigate their public and private online worlds and identities. Scaffolding students to manage their personas through assessment might act as a 'bridge' between higher education and the world of the discipline, profession and workplace. O'Donnell's chapter (Chap. 9) also focuses on the digital world of assessment as a connector to the world of work, through authentic tasks. While assessment is already routinely digital, O'Donnell argues that much practice is simply the computerisation of existing approaches to assessment, which are heavily biased toward text; this echoes Pitt and Winstone's arguments about digital feedback modes (Chap. 7). Instead of continuing to privilege composition, the chapter argues we should employ real-world digital design tasks. A detailed example is provided of a digital design task involving blogging. Common to both chapters is a concern for digital assessment as a way to support students in being and becoming members of their profession or discipline.

Chapter 3
Digitally Mediated Assessment in Higher Education: Ethical and Social Impacts

Margaret Bearman, Phillip Dawson, and Joanna Tai

Abstract As assessments become increasingly mediated by technology, we tend to focus on the immediate benefits of digital innovation. However, by utilising technology, assessment is itself changing and this has implications for society at large. What are the social impacts, moral responsibilities and ethical dilemmas presented by digitally mediated assessments? There are likely to be real challenges for students, individual educators and institutions. As assessments are increasingly delivered with technology, it is worth considering: who controls assessment work; who might be excluded by the new forms of assessment; and who benefits from assessment labour. This chapter reviews the broader social and ethical landscape of digitally mediated assessments and presents practical propositions to help navigate pressing ethical challenges.

> Every day we see the emergence of new technologies. And every day we see a widening gap between progress and society's ability to cope with its consequences. Klaus Schwab (2016)

3.1 Introduction

New technologies present social and ethical challenges. Technology is not neutral; the interaction of people with technologies creates new ways of perceiving, thinking, doing and relating. Higher education assessment is no exception; old assumptions cannot be relied upon and a range of hitherto unforeseen ethical and social dilemmas are emerging as universities slowly embrace technology-mediated assessment. This chapter explores the social impacts associated with overlay of assessment and digital technologies and considers some of the ethical choices afforded to us as academics, assessors, researchers and learners. Ethical considerations permeate all higher education contexts; while they may not be anyone's initial and

M. Bearman (✉) · P. Dawson · J. Tai
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: margaret.bearman@deakin.edu.au; p.dawson@deakin.edu.au; joanna.tai@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_3

immediate concern when thinking about assessment, they are important to ensure we are not harming our students, ourselves and our societies.

Society is being changed by digital technologies. We live in a digital world, where technology is often heralded as being innovative, exciting and future-oriented. At the same time, chatbots tweet offensive and malicious falsehoods (Neff and Nagy 2016) and Facebook data can be used without consent (Solon 2018). Reflecting this tension, discourses of technology in higher education polarise into doom and hype (Selwyn 2014). However, irrespective of pessimism or optimism, "digital education is undeniably a 'big deal'" (Selwyn 2016, p. xviii) Moreover, not all technologies are necessarily *educational*; university students' off-task use of technology may be ubiquitous (Aagaard 2015). In short, digital technologies are both impactful and omnipresent in higher education settings.

Assessment also has significant impacts upon students, albeit of a more traditional kind. Assessment's importance is illustrated by the often repeated mantra "assessment drives learning". However, this nostrum doesn't capture the social complexity underpinning assessment and assessment practices of both educators and students. Assessment concerns standards, knowledge and certification. As such, it shapes university students in many different ways (Hanesworth et al. 2018) and similarly, it also shapes academics (Raaper 2017). Assessment also 'legitimises' certain forms of knowledge (Shay 2008). At a simple level, assessment acts as a 'gate' – to pass an assessment may provide entry to future employment, status and other opportunity. In short: assessment is power. It legitimises knowledge and has the capacity to include, exclude and 'discipline' (to borrow Foucault's term) students and academics alike (Raaper 2016).

The intersection of assessment (powerful) and technology (omnipresent) therefore deserves close scrutiny. Waelbers (2011, p. 138) suggests there is a need to take moral responsibility: to "take up the task of enquiring into the social impacts of technologies". Others have taken up this challenge with respect to the ethical and moral use of technology in higher education (e.g., Selwyn 2014) with possibly the strongest focus in the area of data analytics (Slade and Prinsloo 2013), which argues that there is a moral imperative to act as well as to avoid using student data (such as assessment data) without permission. There is also a substantial literature on critical digital literacies, which reveal significant and diverse thinking about the ethical, political and ideological nature of digital literacies across all levels of education (Pangrazio 2016; Potter and McDougall 2017). However, we seek to conceptualise issues particularly pertinent to the intersection of technology and university assessment, which we believe have not been discussed elsewhere.

We think this is particularly pressing due to the tremendous power exerted by assessment. At the same time, this is not an easy task because new technologies mean there are no precedents. Waelbers (2011, p. 67) notes: "falling back on established moral rules is difficult when developing new technologies or employing new technologies in new manners." She exerts us to use our "moral imaginations". Therefore, we encourage readers, as they consult other chapters in this book, to be mindful of what else might occur, outside of the promising and positive results, and to think of alternate scenarios which might eventuate within different contexts. If

we are not looking for the challenges, the barriers, the equivocal results and the potentially damaging outcomes, then these unanticipated and undesired consequences may become invisible.

The aim of this chapter is to explore some of the social impacts, moral responsibilities and ethical dilemmas presented by digitally mediated assessments. These affect students, educators, administrators, policy makers, software developers, researchers and the broader communities which the academies seek to enhance. Ethical frameworks in healthcare are concerned with notions of the capability of individuals to make decisions about themselves; seeking to provide benefit and avoiding harm; and the just distribution of benefits (Beauchamp 2007). We see these having particular relevance to digitally mediated assessment in higher education around three topics: a) *agency*, b) *diversity* and c) *labour*. Embedded within these, we present three practical propositions to inform ethical design and implementation of digitally mediated assessments. Our concerns and propositions are clearly not exclusive, there are many others and there are many deeper philosophical and moral readings of our current digital world (see for example Waelbars (2011) or Pangrazio (2016)). Our thoughts are strongly influenced by its framing within our own context (higher education in metropolitan Australia) and we suggest readers could explore the social and ethical conundrums within their local contexts.

3.2 *Agency*: Managing the Challenge of the 'Black Box'

In higher education as in healthcare, the ability of individuals to make decisions about themselves is an important ethical principle. Recognising the interdependent nature of higher education, we believe this goal is better framed as agency. In general, agency is practiced when individuals or their communities "exert influence, make choices and take stances in ways that affect their work and/or their … identities" (Eteläpelto et al. 2013, p. 62). We regard agency for students as being a pedagogically and socially important aspiration. It is not controversial to suggest that students should have a degree of agency over their assessment work in accordance with calls for students as "responsible partners" in assessment (Boud and Associates 2010). This also aligns with Slade and Prinsloo's discussion of ethical learning analytics (2013). They contend that student data and associated conditions should always be made known to students and students should be given a degree of agency over how this data is generated, manipulated and used (Slade and Prinsloo 2013). So there is general agreement from both the assessment and technology literatures that student agency is both desirable and necessary.

Enacting these principles presents significant challenges. Technology-mediated assessment are not necessarily enabling better practices. The affordances and constraints provided by particular platforms are frequently obscured. We suggest therefore that with respect to technology-mediated assessment, agency is not just a matter for students but also for educators.

Educators can struggle to understand educational technologies. Sometimes educators are aware of an *explicit* 'black box', such as automated marking software (see Bearman and Luckin, Chap. 5 this volume). More frequently, the constraints of the system simply become part of the taken-for-granted nature of technology. For example, learning management systems and plagiarism software can *implicitly* constrain both educators and students through enabling certain (often simpler) types of assessment and making it more challenging to innovate or expand different kinds of assessment. Thus quizzes become even easier and groupwork becomes even harder to implement. In general, the constraints of technology may not be apparent to academics who have a complex and sometimes romantic view of technology in assessment (Bennett et al. 2017). If educators are struggling to acknowledge and understand the technologies or algorithms underpinning their assessments, how can they begin to explain them to their students? We contend that a necessary step towards agency is for both educators and students to sufficiently understand technology and assessment to make informed choices.

3.2.1 Proposition 1: Build Necessary Skills and Literacies Within Technology-Mediated Assessment

In order to make informed choices about technology-mediated assessment, educators and students must be able to 'read' both the technology and the assessment formats. There is an urgent need for such capabilities. As highly complex forms of technology such as automated marking become more prevalent, the opportunity for "black box" algorithms becomes greater, which increases the risk of unknowing inappropriate use, particularly as educators and students do not have any prior experiences with these digital innovations. Our choices as educators and students are generally predicated on previous experiences of assessment. For example, an educator who has only ever set paper tests for undergraduate school-leavers may have a reduced set of design considerations, when compared to a colleague who has spent time considering the assessment needs of a diverse cohort, with a blend of online and face-to-face students. Similarly, a student who does not understand what a wiki is nor why it might be used for assessment, may struggle to approximate the specific genre presented to them, and therefore resort to mechanistic approaches in order to maximise achievement. In this way, the data created by learning analytics or automated marking processes have the potential to be misunderstood and misapplied by all parties as there are no prior contexts to help shape understanding.

Building skills and literacies is a means to promote greater agency for both educators and students. In this way, educators and students can both draw from a full repertoire of choices when doing assessment work. For the educator, it avoids the trap of developing assessments according to technological imperatives rather than pedagogical purposes. For the student, it provides increased ownership over the assessment process.

So what might these skills and literacies look like? At a simple level, 'literacy' can be considered the ability to 'read' and 'write'. Elmborg (2012, p. 80) writes: "…To read something skilfully is to be literate. To be able to write something is to master the literacy." We refer to this notion of literacy broadly as "skills". At a more complex level, Freire and Macedo discuss how literacy is about reading "the world"; that being literate should permit "learners to recognize and understand their voices within a multitude of discourses in which they must deal" (Freire and Macedo 2005, p. 97). Literacies can be seen as a form of social change as much as skill development (Freire and Macedo 2005; Pangrazio 2016). This is a significant point and we don't wish to promote a 'transmission' model of assessment or digital skills. Instead we take the view that digital and assessment literacies are about understanding the plurality of meaning within assessment forms and digital technologies. Literacies therefore are necessarily about social, ethical and political concerns within a context of skill development.

Students and educators already come with varying degrees of digital and assessment skills but it is important not to make assumptions. There is a view that digital skills are somehow different across generations. The myth of the "digital native" has been busted several times over (Bennett 2016; Margaryan et al. 2011; Selwyn 2009) and we should avoid misguided opinions of what "young people" like and are familiar with. However, what students and educators need to know about digitally-mediated assessment may not be easily itemised. Joint orientation to, and discussion of, technology and its purposes and functions in the assessments provides a useful way forward. However, building skills and literacies development into our assessment pedagogies may produce better traction.

We offer two contrasting examples to illustrate both the possibilities and the limitations. The first example – PeerWise –has been extensively scaled and serves to develop both digital and assessment skills. The second example – the Learning Analytics Report Card (LARC) – is a smaller scale approach, but which develops critical digital- and assessment-literacies.

The PeerWise software develops assessment skills through engaging students in the process of making and refining multiple choice questions (MCQs) (Denny et al. 2008). In this system, MCQs are used for learning purposes. All MCQs are student generated, and feedback can be provided to the question writers if there are errors within the question. When provided with such agency to contribute to assessment tools (albeit in a formative setting), students demonstrated the ability to detect errors and also agreed with faculty on the quality of questions (Denny et al. 2009). Subsequent work identified that students who participated more frequently in the process of creating MCQs also performed better on end-of-semester assessment (which did not contain MCQs) than those with low participation. This association was present for both high and low achieving students (Bates et al. 2012). The involvement of students in assessment creation is therefore likely to contribute to shared knowledge regarding assessment. The advantage of this approach is it develops skills regarding MCQs as well as how technology can mediate student contributions; the limitation is that it perpetuates a traditional form of assessment, which comes with many drawbacks.

The LARC is an example of building digital literacies that may hold promise for assessment (Knox 2017). In short, 12 students were given access to limited learning analytics about their behaviours during the course (e.g., attendance, interactions, time on task) and the capacity to manipulate what was presented back to themselves and at what time. The LARC offered regular reports, with language that deliberately lacked nuance and therefore had the air of being slightly provocative: e.g., "you have usually been extremely social during the course but in week 12 you interacted less with others." The students were offered an opportunity to reflect on each 'report' and through this make personal meaning regarding the potential for this type of assessment. The LARC pilot aimed "to highlight the entangled and complicit conditions of data capture and surveillance, and to offer ways for students to both experience and reflect on forms of participation within the [learning analytics] process…" (Knox 2017, p. 50). The pilot data suggests that educators and students alike learnt through the LARC – about the constraints and affordances of learning analytics and computer generated feedback on assessment tasks. Indeed, the literacies are more than shared, they can be considered 'co-constructed' or jointly developed. However, the LARC is designed for those studying technology and society; this may limit its broader application.

These two examples indicate how skills and literacies inform agency with respect to technology-mediated assessments. They enable informed choices, both in the short and longer term. We suggest that it is worth considering the following when researching or designing technology-enabled assessments. See Box 3.1.

Box 3.1: Practical Considerations Regarding Skills and Literacies That Promote Agency with Respect to Technology-Enabled Assessment

- *What do educators and students know about the constraints and opportunities of the technology – and the assessment design?*
- *What do educators and students understand about the broader social constructions of technology-mediated assessments?*
- *How can low skill levels be accounted for?*
- *What opportunities do educators and students have to build their skills? How about their literacies?*
- *What impact might technological and assessment skills and literacies have on how assessment is designed, implemented and undertaken in a digital world?*

3.3 *Diversity*: The Challenge of Inclusive Technologically-Enhanced Assessments

The promise of technology is that it can ensure greater access to higher education (Devlin and McKay 2016), particularly to those who have been excluded on the basis of: class, race, ethnicity, gender, disability, citizenship status or rurality. Technology-mediated changes such as online learning or text-to-voice readers may reverse some of this exclusion. However, there appears to be limited evidence gathered to date to support some claims. In other words, the availability of technology does not automatically enable diversity, although there certainly is the potential to reach greater audiences.

Many universities now have policy on accessible design for technologies. This is often based on the considerable literature exploring universal design for learning. Universal design offers both principles and processes for designing inclusive and fair assessments through thinking flexibly and providing alternatives to accommodate individual differences (Almond et al. 2010). This is important and laudable work but it may not be enough; over a decade ago, Hanafin et al. (2007, p. 438) wrote in the disability literature: "[higher education] assessment has both intended and unintended consequences, and consequences are greater for some students than for others."

Unintended consequences are difficult to track, especially in relation to assessment. Discrimination already occurs in assessment; for example Esmail and Roberts' (2013, n.p.) study of over 5000 candidates sitting a postgraduate medical examination indicated that "subjective bias due to racial discrimination …may be a cause of failure…" Such biases may be 'invisible' to educators and other students and therefore there is particular danger when it comes to technology. Moreover, technology itself can be a source of discrimination, as technology is not cheap and while there is an assumption that it is ubiquitous, this is only an assumption.

The power imbalance inherent in the system is unlikely to facilitate diverse students' ability to appeal against or even point out these concerns by themselves. This power imbalance may be exacerbated by a rise in computational support to generate grades or provide feedback, as the underpinning algorithms tend to be derived from historical data. Though we consider the digital to be a facilitator of inclusivity, artificial intelligence tends to determine solutions according to previously submitted work. Thus it may continue to disadvantage diverse populations who do not fit previous patterns of achievement. How can we meaningfully ensure that inclusion is built in to our technology-mediated assessments and ensure that access to higher education is as broad as possible?

3.3.1 *Proposition 2: Community Engagement Helps Identify 'Invisible' Barriers*

Inclusivity benefits all (Hanafin et al. 2007). However, educators may be unaware of the potentially deleterious impact of our assessments, due to the necessary constraints of individual experiences as well as institutional systems. This means ethical frameworks such as the Open University's call for data modelling and analysis to be: "sound and free from bias" (Prinsloo and Slade 2017) have some limitations. In plain terms, we do not know what our own (inevitable) biases are. It is therefore important to work with all potential stakeholders to consider potential barriers to participation, which may be the result of digitally mediated assessments. The engagement of diverse lay people in the design and implementation of technology mediated assessment – most likely at a course or departmental level – may help take account of those 'invisible' cultural and contextual factors. There is a long tradition of lay involvement in health care (Entwistle et al. 1998) and there may be lessons for higher education. Lay people may be better able to illuminate and question practices which are taken for granted within the academy, but less so in general society. Of course, consultation should also include students, across the demographics of the enrolled cohort, and faculty, such as tutors/markers, academic developers, technology support staff, lecturers, administrators and so on.

In addition to lay perspectives, the academy may also usefully seek information from industries outside education to inform better technology-mediated assessment practices. Health, military and other high-stakes fields have ongoing assessment requirements (Bennett Jr. et al. 2002; Castanelli et al. 2016), so may have innovated different forms of assessments which better deal with particular access issues. In some areas, they may gain from insights into university processes and so such consultation may provide bidirectional benefits. However, the 'industry of assessment' – the private enterprises who seek to make a profit out of assessment – also may present barriers to community engagement. Not all assessment is controlled by educators or even universities. For example, outsourcing assessment or grading to private enterprise is increasingly common. We suggest that the university sector must consider how to engage with these providers.

In summary, educators and institutions must examine themselves, in order to identify crucial dilemmas and potential pitfalls. This is not a fixed state as social norms shift and change; we may find old barriers fall and new ones emerge. Under these circumstances, the most ethical approach may be through repeatedly and genuinely interrogating technology mediated assessment practices.

This proposition leads to three questions for all those involved in digital assessment as outlined in Box 3.2.

Box 3.2 Practical Considerations Regarding Identification of 'Invisible' Barriers to Diversity in Technology-Mediated Assessments

- *What does enabling access really mean for technology-mediated assessment?*
- *How do we prevent ourselves from promulgating entrenched discriminations in the process of developing or implementing technology-mediated assessment?*
- *What opportunities are there for diverse stakeholders, including those outside of the academy, to have involvement in our assessment processes?*
- *How do we support our increasingly diverse cohorts through technology-mediated assessment?*

3.4 *Labour:* The Challenge of Assessment Products in a Digital Landscape

Assessment has traditionally involved students engaging in academic labour, the fruits of which were seen by educators and then kept by the student or discarded. Technology has changed this, by making it easier to store and duplicate the work students produce. This has created opportunities for software vendors to use student work to enhance the assessment process. However sometimes these players have additional motives. For example, academic integrity software vendors provide text-matching services, which have greatly reduced incidence of copy-paste plagiarism, however to do so they have kept copies of student work. Students have taken legal action against this practice, unsuccessfully claiming that it infringed upon their intellectual property (Bennett 2009). Business models are also being built around students sharing their assignments online with their peers; concerns have been raised about the legality of those approaches and the boundary between the student's property and the university's property (Rogerson and Basanta 2015). Both of these cases raise the question of who should be allowed to profit from students' assessment labours.

In a digital world, students' assessment labours can have effects beyond the boundary of the academy, and even beyond the domain of education and technology. While it is heartening to see students' work have an impact for the community, this impact is not always positive. For example, Wikipedia has seen many student contributions, and there is a small body of literature investigating the effects of contributing to Wikipedia on students (Di Lauro and Johinke 2017). Despite the noble intent of those setting Wikipedia assignments, students often do more harm than good to Wikipedia, creating enormous work for long-term Wikipedians (Wikipedia 2018).

Assessment is intended to evidence what students are capable of, however students may need to be aware that in a digital world it may also form a permanent record of their mistakes. There may be secondary uses of this evidence against students. This has been recently illustrated by the case of Dr. Bawa-Garba, a junior doctor whose reflections collected as part of an educational e-portfolio may have influenced a court case regarding lack of fitness to be a medical practitioner (Dyer and Cohen 2018). This case illuminates the potential consequences when asking students to post on Twitter, create websites, or upload videos to YouTube for assessment purposes. Future employers of those students, and indeed, society, may look favourably or unfavourably upon these artefacts, depending on their content and tone. How then can we manage the social and commercial impacts of assessment labour, in a way that minimises their harms and maximises their benefits to students and society?

3.4.1 Proposition 3: Technology Mediated Assessment Should Inform, and Be Informed by, Institutional Policies

Unlike other challenges discussed thus far, questions of the ownership, access to, and impacts of assessment products are generally answered through institutional policies. These are essentially legal questions and concern intellectual property (IP) arrangements. With respect to IP the World Intellectual Property Organization (http://www.wipo.int/portal/en/) offers comprehensive access to IP policies across countries that describe how institutions manage IP arising from assessments. On a related note, these institutions generally describe strict privacy protocols as well.

However, there are still pitfalls. The use of technology for assessment will be constantly developing and changing well into the future and some types of uses (and potential benefits and harms) will not be covered by current policies. Innovative educators who work with technologies outside their institutionally provided tools can find themselves in grey areas as policy catches up (Bennett et al. 2017). There are also points where the university policy is relatively powerless. For example, a range of software, including plagiarism detection software and learning management systems have end user license agreements that require the "donation" of student data. In these situations, students do not have full control of their intellectual property. Moreover, legal frameworks are not the same as ethical frameworks. Recent events such as the Facebook and Cambridge Analytica data breach (Solon 2018) are clear examples of the evolution of what is considered appropriate to share, and the realisation that data are not as private or safe as they were thought to be. Constant vigilance may be required to stay abreast of new developments and modifications to license agreements.

Through considering the questions in Box 3.3, we urge educators and designers of digitally mediated assessments to think about how the digital products of assessment labour are being used.

Box 3.3 Practical Considerations Regarding Assessment Labour, Institutional Policies and Technology-Mediated Assessment

- *What interactions do students have with the broader community as part of this technology-mediated assessment?*
- *What digital work do students produce as a consequence of this assessment and who else might see this?*
- *What are the institutional policies, ethical or legal frameworks that may influence how this technology-mediated assessment is implemented and how assessment products are stored?*

3.5 Conclusions

Technology will continue to mediate and influence the format and processes of assessment into the future. One of the greatest challenges facing those who work with technology, is how they take account of an ever innovating landscape. Technological determinism might suggest that all human actors are shaped by technologies (Oliver 2011) and that the digital controls us, rather than the other way around. We reject this view. Our belief is that individuals have many opportunities to exercise agency, and therefore should seek to be moral, and to 'do good' through their practices.

In this chapter we have presented three concerns facing us as academics, assessors and learners. There are many others. Readers may wish to invoke their "moral imaginations" to apply these to the other chapters within this book, or other work about assessment and technology. As you do so, we suggest you ask yourself three questions: *Who controls? Who accesses? Who benefits?*

References

Aagaard, J. (2015). Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & Education, 87*(Supplement C), 90–97. https://doi.org/10.1016/j.compedu.2015.03.010.

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., et al. (2010). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *The Journal of Technology, Learning and Assessment, 10*(5).

Bates, S. P., Galloway, R. K., & McBride, K. L. (2012). Student-generated content: Using PeerWise to enhance engagement and outcomes in introductory physics courses. In N. S. Rebello, P. V. Engelhardt, & C. Singh (Eds.), *2011 physics education research conference: AIP conference proceedings* (Vol. 1413, pp. 123–126). Melville: American Institute of Physics.

Bearman, M., & Luckin, R. (This volume). Chapter 5: Preparing university assessment for a world with AI: tasks for human intelligence. In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world*. Dordrecht: Springer.

Beauchamp, T. L. (2007). The 'four principles' approach to health care ethics. In R. Ashcroft, A. Dawson, H. Draper, & J. McMillan (Eds.), *Principles of health care ethics* (2nd ed., pp. 3–10). Chichester: Wiley.

Bennett, M. G. (2009). A nexus of law and technology: Analysis and postsecondary implications of A.V. et al. v. iParadigms, LLC. *Journal of Student Conduct Administration, 2*(1), 40–45.

Bennett, S. (2016). The digital revolution in higher education: Rhetoric and reality. In J. E. Cote & A. Furlong (Eds.), *Routledge handbook of the sociology of higher education* (pp. 328–327). Abingdon: Routledge.

Bennett, W., Jr., Schreiber, B. T., & Andrews, D. H. (2002). Developing competency-based methods for near-real-time air combat problem solving assessment. *Computers in Human Behavior, 18*(6), 773–782. https://doi.org/10.1016/S0747-5632(02)00030-4.

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology, 48*(2), 672–682. https://doi.org/10.1111/bjet.12439.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Castanelli, D. J., Jowsey, T., Chen, Y., & Weller, J. M. (2016). Perceptions of purpose, value, and process of the mini-clinical evaluation exercise in anesthesia training. *Canadian Journal of Anesthesia/Journal Canadien d'anesthésie, 63*(12), 1345–1356. https://doi.org/10.1007/s12630-016-0740-9.

Denny, P., Luxton-Reilly, A., & Hamer, J. (2008). The PeerWise system of student contributed assessment questions. In *ACE '08 proceedings of the 10th conference on Australasian computing education-volume 78* (pp. 69–74). Darlinghurst: Australian Computer Society.

Denny, P., Luxton-Reilly, A., & Simon, B. (2009). Quality of student contributed questions using PeerWise. *ACCE '09 Proceedings of the 11th Australasian Conference on Computing Education-Volume 95* (pp. 55–63). Darlinghurst: Australian Computer Society.

Devlin, M., & McKay, J. (2016). Teaching students using technology: Facilitating success for students from low socioeconomic status backgrounds in Australian universities. *Australasian Journal of Educational Technology, 32*(1), 92–106. https://doi.org/10.14742/ajet.2053.

Di Lauro, F., & Johinke, R. (2017). Employing Wikipedia for good not evil: Innovative approaches to collaborative writing assessment. *Assessment & Evaluation in Higher Education, 42*(3), 478–491. https://doi.org/10.1080/02602938.2015.1127322.

Dyer, C., & Cohen, D. (2018). *How should doctors use e-portfolios in the wake of the Bawa-Garba case?* (p. 360). BMJ: British Medical Journal (Online). https://doi.org/10.1136/bmj.k572.

Elmborg, J. (2012). Critical information literacy: Definitions and challenges. In C. Wilkinson & C. Bruch (Eds.), *Transforming information literacy programs: Intersecting frontiers of self, library culture, and campus community* (pp. 75–80). Chicago: Association of College and Research Libraries.

Esmail, A. & Roberts, C. (2013). Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: Analysis of data. BMJ 347;f5662. https://doi.org/10.1136/bmj.f5662.

Entwistle, V. A., Renfrew, M. J., Yearley, S., Forrester, J., & Lamont, T. (1998). Lay perspectives: Advantages for health research. *BMJ, 316*(7129), 463–466.

Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational Research Review, 10*(Supplement C), 45–65. https://doi.org/10.1016/j.edurev.2013.05.001.

Freire, P., & Macedo, D. (2005). *Literacy: Reading the word and the world*. e-version. Routledge. https://doi.org/10.4324/9780203986103.

Hanafin, J., Shevlin, M., Kenny, M., & Neela, E. M. (2007). Including young people with disabilities: Assessment challenges in higher education. *Higher Education, 54*(3), 435–448. https://doi.org/10.1007/s10734-006-9005-9.

Hanesworth, P., Bracken, S., & Elkington, S. (2018). A typology for a social justice approach to assessment: Learning from universal design and culturally sustaining pedagogy. *Teaching in Higher Education*, 1–17. https://doi.org/10.1080/13562517.2018.1465405.

Knox, J. (2017). Data power in education: Exploring critical awareness with the "Learning Analytics Report Card". *Television & New Media, 18*(8), 734–752. https://doi.org/10.1177/1527476417690029.

Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students' use of digital technologies. *Computers & Education, 56*(2), 429–440. https://doi.org/10.1016/j.compedu.2010.09.004.

Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication, 10*, 4915–4931. doi:1932–8036/20160005.

Oliver, M. (2011). Technological determinism in educational technology research: Some alternative ways of thinking about the relationship between learning and technology. *Journal of Computer Assisted Learning, 27*(5), 373–384. https://doi.org/10.1111/j.1365-2729.2011.00406.x.

Pangrazio, L. (2016). Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education, 37*(2), 163–174. https://doi.org/10.1080/01596306.2014.942836.

Potter, J., & McDougall, J. (2017). Dynamic literacies. In J. Potter & J. McDougall (Eds.), *Digital media, culture and education: Theorising third space literacies* (pp. 15–36). London: Palgrave Macmillan UK.

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un) charted. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 49–57). SOLAR.

Raaper, R. (2016). Academic perceptions of higher education assessment processes in neoliberal academia. *Critical Studies in Education, 57*(2), 175–190. https://doi.org/10.1080/17508487.2015.1019901.

Raaper, R. (2017). Tracing assessment policy discourses in neoliberalised higher education settings. *Journal of Education Policy, 32*(3), 322–339. https://doi.org/10.1080/02680939.2016.1257160.

Rogerson, A. M., & Basanta, G. (2015). Peer-to-peer file sharing and academic integrity in the internet age. In T. A. Bretag (Ed.), *Handbook of academic integrity* (pp. 273–285). Singapore: Springer Singapore.

Schwab, K. (2016). *Four leadership principles for the Fourth Industrial Revolution*. Retrieved from https://www.weforum.org/agenda/2016/10/four-leadership-principles-for-the-fourth-industrial-revolution.

Selwyn, N. (2009). The digital native – Myth and reality. *ASLIB Proceedings, 61*(4), 364–379. https://doi.org/10.1108/00012530910973776.

Selwyn, N. (2014). *Digital technology and the contemporary university: Degrees of digitization*. London: Routledge.

Selwyn, N. (2016). *Is technology good for education?* Cambridge: Polity.

Shay, S. (2008). Beyond social constructivist perspectives on assessment: The centring of knowledge. *Teaching in Higher Education, 13*(5), 595–605. https://doi.org/10.1080/13562510802334970.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366.

Solon, O. (2018). The Cambridge Analytica Files: Facebook says Cambridge Analytica may have gained 37m more users' data. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/apr/04/facebook-cambridge-analytica-user-data-latest-more-than-thought.

Waelbers, K. (2011). *Doing good with technologies: Taking responsibility for the social role of emerging technologies* (Vol. 4). London: Springer Science & Business Media.

Wikipedia. (2018). *Wikipedia: Student assignments*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Student_assignments.

Margaret Bearman is a Research Professor within the Centre for Research in Assessment and Digital Learning, Deakin University. Margaret's research focusses on learning and teaching in higher and professional education. She is particularly interested in ways of improving feedback, the sociomateriality of standards, and what it means to be an online learner in a digital world.

Phillip Dawson is an Associate Professor and Associate Director of the Centre for Research in Assessment and Digital Learning at Deakin University. He holds degrees in education, artificial intelligence and cybersecurity. His research focuses on assessment and learning matters like feedback, rubrics and academic integrity, often with influences from computer science.

Joanna Tai is a Senior Research Fellow at the Centre for Research in Assessment and Digital Learning, at Deakin University. Her research interests include peer learning, self and peer assessment, evaluative judgement, student perspectives on learning and assessment, and research synthesis.

Chapter 4
Cognitive Offloading and Assessment

Phillip Dawson

Abstract Cognitive offloading refers to using tools like notes, calculators or spell-checkers to reduce the cognitive demands of a task. Assessment has a patchy history in attending to cognitive offloading. In some settings, such as exams, there are explicit rules that relate to cognitive offloading, such as the allowance or prohibition of textbooks and notes. However in other settings, particularly authentic open-ended tasks there is less clarity. This chapter proposes principles for incorporating cognitive offloading into assessment, with a focus on transparency, programmatic assessment, evaluative judgement and authentic assessment.

When you write something down to remember it later, you are engaging in 'cognitive offloading': "the use of physical action to alter the information processing requirements of a task so as to reduce cognitive demand" (Risko and Gilbert 2016, p. 677). Although the term 'cognitive offloading' is relatively new, people have been cognitive offloading for a long time, with tools like pencils, calculators and word processing software. Assessment has also long been designed to allow or prohibit various forms of cognitive offloading, for example, an examination might allow or prohibit students to bring notes or other memory aids.

However, recent developments in both assessment and cognitive offloading technologies have created problems. Assessment has been shifting away from highly controlled and contrived tasks like in-person invigilated (proctored) pen-and-paper examinations toward other forms which are performed at a time and place of the student's choosing under circumstances that are less tightly controlled. This provides many more opportunities for students to use cognitive offloading tools outside of the view of their assessor, making it difficult to know if the work students produce was done independently or with the help of tools.

P. Dawson (✉)
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: p.dawson@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_4

Our current digital world is characterized by rapid advancements in artificial intelligence and related technologies, which are providing more cognitive offloading tools. One example is Computer Algebra Systems (CAS), which can be used to offload the solving of algebraic problems. Recent media reports have highlighted situations where students have used CAS tools against their assessors' wishes (Biddle 2017). Similarly, students who have difficulty in paraphrasing sources can use online tools instead, ranging from translation software (Jones and Sheridan 2015) to custom-made paraphrasing software (Lancaster and Clarke 2009; Spinbot.com 2016). These tools enable students to be certified as having met outcomes they have not demonstrated themselves, and which they may not be capable of demonstrating at that particular point in time. Premature or inappropriate use of these tools also poses a real threat to students developing those skills. Without guidance, students may become dependent on these tools, but unable to make judgements about the quality of the outputs they provide.

At present a limited range of learning outcomes can be offloaded to tools. However, one of the long-term goals of the field of artificial intelligence is the 'Robot College Student Test': the development of software that is capable of independently completing the requirements of an undergraduate degree (Goertzel et al. 2012). As progress toward this goal is made, further cognitive offloading tools are likely to become available that are capable of performing the work to demonstrate achievement of a larger range of increasingly sophisticated learning outcomes.

Given the threats cognitive offloading poses to assessment and learning, it may be tempting to ban some cognitive offloading tools. However, this may be misguided given the increasing role that cognitive offloading is having in the workplace. With the rise of artificial intelligence and other technologies, employees are increasingly concerned that their jobs may be taken by these systems (Brougham and Haar 2017). Although media have hyped the prospect of job losses, some professions are instead reconsidering what the role of human professionals will be as technology starts to take on roles that were traditionally human. For example, Jha and Topol (2016) suggest that radiologists and pathologists may shift toward being 'information specialists' who exercise judgement on the results of work undertaken by computers. This broader societal trend suggests that blanket bans on cognitive offloading at university may stymie the preparation of students for working in a world where artificial intelligence is rapidly developing. If assessment is to prepare students for a digital world where cognitive offloading is the norm, assessment needs to embrace it.

The challenge then is not how to stop students from undertaking cognitive offloading, but how to design assessment in such a way that it: (a) still ensures validity of judgements made about what students can do; (b) uses cognitive offloading to help students achieve broader and deeper outcomes; and (c) helps students make judgements about the products of their cognitive offloading. This chapter uses a set of familiar assessment concepts to develop principles for addressing cognitive offloading in a digital world. It commences with a review of some fundamental concepts and findings from cognitive offloading research.

4.1 Cognitive Offloading

Cognitive offloading research is largely situated within the field of cognitive psychology. Most of the research so far has been concerned with two key questions (Risko and Dunn 2015, 62): how do people decide if they are going to cognitive offload or not, and what are the effects of cognitive offloading on cognitive processing. Most of what can be said empirically about cognitive offloading thus falls under those two broad questions.

Much cognitive offloading research has come from the perspective that the human mind is cognitively lazy (Risko and Dunn 2015). If given the choice to use our minds, or to offload, we'll usually choose to offload, even if this means we need to expend physical energy. Several research studies have been designed using the 'choice/no-choice' paradigm (Risko and Gilbert 2016), where participants are either required to cognitively offload (no-choice external), required to not cognitively offload (no-choice internal), or free to choose if they cognitively offload (choice). When given the choice, people usually opt to cognitively offload, even for very undemanding tasks where there is no benefit to offloading. For example, when asked to remember two letters of the alphabet (e.g., "PJ") for a few moments, many participants opted to write the letters down despite little benefit to doing so (Risko and Dunn 2015). One assessment implication of this sort of cognitive laziness is that if we give students the choice they may be likely to cognitively offload even if they expect performance will not improve.

While it is possible to observe physical actions that result from choosing to cognitively offload – like writing down a series of characters – cognitive offloading involves both the physical action and the non-use of mental resources. For example, participants who typed information into a computer remembered much less of the information if they thought the information would be saved (Sparrow et al. 2011); the physical action was the same, but participant belief that they did not need to remember resulted in cognitive offloading. However this result is not uniform across all media: in another study, participants who took a photograph of an object remembered it worse than those who did not take a photograph, even when the participants immediately deleted the photograph (Soares and Storm 2017). Other studies have similarly found that for different types of tasks, requiring different mental capabilities, decisions to cognitively offload work differently (Eskritt and Ma 2014). Current thinking in cognitive offloading research is that both cognitive laziness as well as metacognitive beliefs about performance drive decisions to cognitively offload or not (Risko and Dunn 2015). If someone thinks cognitive offloading will be less mentally taxing, or result in better performance, they will likely choose to do it.

Cognitive offloading can also take the form of deferring, if we know that someone or something else will do the work for us. In one study, participants who were tasked with solving anagram problems performed much more poorly and were less persistent if they were told that the answers would be available to them (Risko et al. 2017). However, they did not believe this affected how persistent they were. This suggests that the common formative assessment design of asking students to solve

problems, and letting them check the solutions themselves, may lead to students choosing cognitive offloading in the form of checking the answer when the mental effort required becomes too high.

A danger in cognitive offloading is that we can start to believe we are capable of more than we actually are. If we know we can search the Internet for the answer, we may think we actually know the answer ourselves. In one study involving nine experiments, participants who had searched for information themselves remembered it much more poorly than people who had information provided to them (Fisher et al. 2015). The participants who searched also had an over-inflated sense of competence with the information. The researchers made the suggestion that the search participants never bothered to truly learn because they felt they could find the information themselves. This suggests there may be a tendency to include cognitive offloads in judgements about one's own capabilities.

Cognitive offloading has some similarities with earlier work on cognitive tools. That body of literature was focused on tools that support learning and it informed the development of educational designs and technologies (Jonassen 1992; Herrington and Parker 2013; Iiyoshi et al. 2005). Cognitive tools research is largely situated within the educational technology research literature. In contrast, the cognitive offloading literature is largely concerned the use of tools that have not been chosen for their educational merits, and the use of such tools outside of a formal educational setting. Where cognitive tools presented opportunities for educational designers to build better courses, cognitive offloading presents assessors with the challenge of students themselves choosing to use tools that may hamper the effectiveness of the task.

There are two significant gaps in the cognitive offloading research that serve as caveats for the arguments in this chapter. Firstly, the majority of the empirical work on cognitive offloading has focused on memory or performance in small-scale tasks in lab settings; cognitive offloading is rarely studied in the context of large complex tasks that are used in authentic assessment. Secondly, although the research has much to say about how individuals choose to cognitively offload, it is relatively silent about how the choice should be made by someone else (e.g., an assessor). This chapter thus takes the conceptual core of cognitive offloading as a tool to explore contentious issues like the rise in artificial intelligence and outsourcing of work, rather than a set of empirically-derived propositions.

4.2 Cognitive Offloading and Assessment

As a term in education, 'assessment' has a variety of meanings, and these meanings have shifted over time (Cookson 2017). This chapter takes Joughin's definition that assessment means "[making] judgements about students' work, inferring from this what they have the capacity to do in the assessed domain, and thus what they know, value, or are capable of doing" (Joughin 2009, p. 16). However it is also important to note that assessment serves multiple purposes (Boud 2000), and that cognitive

offloading may have different implications for formative, summative and sustainable assessment. This section considers the implications of cognitive offloading on different purposes and concepts within assessment.

4.2.1 Cognitive Offloading and Summative Assessment

Assessment's summative or certification purpose involves making valid and reliable judgements about what students are capable of based on what they have demonstrated. When students use cognitive offloading in ways not intended by the assessment designer, they threaten the accuracy of those judgements; for example, if an engineering assignment is meant to be solved without a Computer Algebra System (CAS) tool, but a student uses one, the judgements made about that student may be inaccurate. This relates to the choice/no-choice paradigm in cognitive offloading research. Given the results about cognitive laziness and the overall preference for cognitive offloading people appear to have, if assessment designers implicitly intend for students to do a task themselves (a 'no-choice internal' design) but students know they can cognitively offload, they probably will do so.

While cognitive offloading researchers design experiments around choice/no-choice and internal/external binaries, summative assessment is only sometimes conducted with explicit rules for these binaries. The most common example of explicit rules is in exams, which are typically conducted in a no-choice internal design (for closed-book exams) or a limited-choice design (for open-book exams), or even designs that allow for much more choice, such as open-book, open-web exams (Williams and Wong 2009) that are choice designs. Cognitive offloading may provide a useful construct for educators in choosing between closed and open book exams; students are less likely to remember the information in their textbook if they can bring it with them into the exam, however by offloading the remembering of information to the textbook they may be able to use cognitive resources to demonstrate learning outcomes that cannot be offloaded to a textbook.

Outside of the context of examinations, the rules around cognitive offloading in summative assessment are somewhat less clear. Take, for example, the common task of paraphrasing information from a source. Paraphrasing is a core academic skill, and it is essential to be able to produce many genres of academic writing. However, it is easier to copy-paste text into online tools like 'SpinBot' (Spinbot.com 2016) and have the text instantly paraphrased than it is to actually do the paraphrasing yourself. It may also produce higher quality paraphrasing if your own paraphrasing skills are not well developed. There are dozens of other ways to cognitively offload academic skills, including approaches we may now think of as relatively benign, like bibliography generation, spell checking and grammar checking. The rules around which cognitive offloads are acceptable and which are not acceptable are rarely made explicit to students.

The primary purpose of summative assessment is to certify that students have met particular learning outcomes. However, if assessments are designed in such a

way that we do not know the extent to which the student has cognitively offloaded, we do not know if the student themselves has actually met the outcome. The practice of specifying learning outcomes without reference to the people or tools involved means that we generally do not know if the learning outcome was meant to be achieved with or without cognitive offloading. This is akin to certifying that someone can change gears in a car by testing them in their choice of a car with an automatic or manual transmission. For maximum transparency, where students are given choice with respect to cognitive offloading the learning outcomes and assessment need to specify this.

4.2.2 Cognitive Offloading and Assessment for Learning

The simplest definition of assessment for learning is that it is assessment which privileges assessment's learning purpose over other purposes (Carless 2017). Given previous results from cognitive offloading research suggest (a) that people will offload if given the chance, (b) that offloading hurts performance in short-term memory tasks when the offloads are unavailable, and (c) that it can lead to blind spots in knowledge, cognitive offloading may pose a threat to assessment for learning. For example, imagine an English-speaking student who is learning Spanish, and undertaking a piece of formative assessment which requires them to translate a paragraph of text. Automatic translation tools like Google Translate are becoming more ubiquitous and the quality of the translations provided has increased to the point that they approach the quality required for admission to university, even for long and complex texts (Groves and Mundt 2015). There is concern that use of these translation tools may hamper student development in language learning, and also that policy is not keeping pace with technological development (Mundt and Groves 2016). However, tackling this problem on its own will not address the broader threat that cognitive offloading poses for assessment for learning.

The cognitive offloading literature would suggest that, given the choice, students are going to offload assessment for learning tasks. In the presence of tools that will solve the entire problem like Google Translate, or problem solutions at the back of the book, students may persist with difficult problems for less time, and not even know they are being less persistent (Risko et al. 2017). This suggests that the choice/no-choice paradigm may be very useful: if students are to use assessment for learning to develop skills that could actually be offloaded, then they need to be told not to use those offloads.

However, cognitive offloading may also offer benefits to assessment for learning in that it may increase the efficiency of students when undertaking work that cannot be offloaded. If students have already mastered some particularly time-consuming algebra that is fundamental for the assessment of a more advanced learning outcome, then offloading the time-consuming algebra could support students in spending more time on the advanced learning outcome.

4.2.3 Cognitive Offloading and Sustainable Assessment

Boud's (2000) concept of sustainable assessment proposes that assessment needs to serve students in the long term, and that our short-term obsession with marks and grades can lead to a long-term dependence on others to evaluate our performance. A core concept in sustainable assessment is the development of 'evaluative judgement', which is students' understanding of what quality work looks like and their ability to see quality in their own work and the work of others (Tai et al. 2018). Where cognitive offloading may threaten assessment's learning purposes, it may actually help assessment's sustainable purposes through requiring students to make regular judgements about the quality of the work produced from offloading. For example, as CAS tools become more sophisticated, students will need to be trained more in determining if the solution they have been provided actually addresses the problem; their ability to make sound evaluative judgements may become more important than their ability to do the work themselves.

However, if evaluative judgement itself is offloaded, this may similarly disrupt students' development of this capability. The artificial intelligence community has long worked toward developing systems that can conduct evaluative judgements; examples include Google's PageRank system that evaluates web pages, or automated essay scoring systems that mark student work (Dawson 2018). If students start to depend on these systems to understand if their work is good enough, they may become unable to judge the quality of work in the absence of such support.

4.2.4 Cognitive Offloading and Authentic Assessment

When given the opportunity to cognitively offload, people tend to do so. This includes not only students, but also professionals. How, then, should the decision be made to permit students to undertake cognitive offloading or not – to set assessments as choice or no-choice tasks? Authenticity may serve as a useful lens to view this decision through; the argument being that cognitive offloading should be permitted to students to the extent that professional practitioners themselves engage in cognitive offloading.

Gulikers et al. (2004, p. 69) define authentic assessment as "assessment requiring students to use the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in professional life." They proposed a five-dimensional framework, with one of the dimensions, 'physical context', having potential for the inclusion of cognitive offloading. Gulikers et al. argue that the physical context of an authentic assessment task should reflect the tools and resources that practitioners use when conducting the same work. While not explicitly mentioned, cognitive offloading could be considered as part of the physical context of a task because it is part of the resources a practitioner may use. There is thus room to accommodate cognitive offloading as part of authentic assessment. An

authentic assessment view of cognitive offloading suggests that students should be allowed to engage in cognitive offloading if professionals do so. Taking the metaphor of choice/no-choice from cognitive offloading research, choice is authentic, and no-choice is inauthentic.

4.2.5 Cognitive Offloading and Programmatic Assessment

Taking a programmatic view of assessment involves broadening thinking about assessment beyond just the immediate task or unit of study toward the entire degree or program of study (Schuwirth and Van der Vleuten 2011). Programmatic assessment is concerned with scaffolding students toward achieving degree-level outcomes, and certifying that students have achieved those outcomes. A programmatic view allows assessment designers to reduce over-assessing, by identifying instances where outcomes have already been sufficiently demonstrated. Taking a programmatic assessment perspective on cognitive offloading, students need not do all of every task themselves. As long as an outcome is sufficiently assessed in the degree program without cognitive offloading, students should be free to use cognitive offloading at other times.

One way to incorporate cognitive offloading in programmatic assessment may be to use it as an instructional scaffold. At the start of a degree, students may be freer to choose the cognitive offloading they want to use, but by the end of a degree they may be required to do the entirety of a task by hand. Alternatively, a reverse-scaffolding approach (Chase and Abrahamson 2015) could be taken, whereby students are only allowed to use cognitive offloading once they have mastered particular outcomes. This might enable the achievement of more challenging learning outcomes because the lower-level tasks have been offloaded; this is similar to primary school students learning their times tables before being allowed to offload them to calculators so they can solve more complex problems. Differing approaches may need to be taken with respect to different outcomes within a program, depending on the level of automaticity and consistency expected of a graduate with respect to the outcome: are they expected to demonstrate an outcome without help quickly and correctly every time, or is the ability to undertake that outcome at all enough?

4.2.6 Cognitive Offloading and Academic Integrity

Academic integrity could be fruitfully rethought through the lens of cognitive offloading. In addition to the ethical element of academic integrity that is usually focused on (Bertram-Gallant 2015; Davies and Howard 2016), there may also be a cognitive offloading element. In the future, students may be required to cite not only the materials they have read, but also the cognitive offloading they engaged in.

Rather than focusing on each possible academic integrity blunder students can make, academic integrity as a field could usefully abstract away to the concepts of choice/no-choice and internal/external. Is there any difference from an integrity perspective if a student has purchased a bespoke essay from a website, or if they have used a computer program to produce one? These are both examples of students undertaking a no-choice internal task as if it were a choice task.

Returning to the 'Robot college student test' discussed earlier (Goertzel et al. 2012), there are significant implications for academic integrity if this goal is met; any task a student is expected to do will be offloadable to technology. This again suggests the need to act in a principles-based way rather than separately responding to translation tools, CAS, paraphrasing tools and other offloads individually. When a computer can meet any learning outcome on behalf of students, which learning outcomes do students still need to meet?

4.3 Five Principles for Incorporating Cognitive Offloading into Assessment

This chapter has argued that cognitive offloading is an unavoidable part of learning and working in a digital world, and that assessment should be designed with cognitive offloading in mind. This section proposes five principles for researchers and assessment designers to apply cognitive offloading to their assessment work.

1. Specify learning outcomes and assessment with reference to the cognitive offloading allowed

Clear specification of learning outcomes and assessment is fairer and safer for students, assessors, and the community (Biggs and Tang 2007). The current practice of specifying outcomes without reference to the assistance allowed is not transparent, and will become less transparent as cognitive offloading technologies become more sophisticated. Learning outcomes need to make mention of what sort of cognitive offloading is allowable. Disciplines, institutions or departments may wish to establish lists of acceptable cognitive offloads for particular levels of study and make mention of these in official course documentation and in marketing the outcomes of degree programs.

2. Programmatically offload lower-order outcomes once mastered

Cognitive offloading is a natural partner to programmatic assessment, because it provides a vehicle to avoid repetitive over-assessment of some lower-level outcomes. Taken together, cognitive offloading and programmatic assessment imply that once mastery has been demonstrated to the level required of the degree program, the student should be free to offload. This raises the challenge of specifying the level of automaticity and consistency required of students before they are assessed as having reached the required level of mastery to offload an outcome.

3. Treat capability with cognitive offloading as an outcome

If assessment is to prepare students to function in a digital world, students need to learn about cognitive offloading. This includes not just learning how to use particular tools, but developing a conceptual understanding of cognitive offloading as applied to their discipline. In our students' lifetimes the capability of cognitive offloading technology will likely continue to progress, so their ability to make decisions about which tools to use, when to offload and when to do it themselves is key. Learning about cognitive offloading is not a purely technical skill and may include concepts like privacy, attribution of work and efficiency.

4. Extend the development of evaluative judgement to include the products of cognitive offloading

More complex cognitive offloading tools undertake work on behalf of students, such as paraphrasing or algebra solving. For students to use these tools effectively they need solid evaluative judgement, otherwise they risk building their own work on a foundation of incorrect or poor-quality offloaded work. Evaluative judgement has previously been defined as "the capability to make decisions about the quality of work of self and others" (Tai et al. 2018, p. 471). The term 'others' could be seen to imply persons, however if students are to master cognitive offloading they will need to learn to apply evaluative judgement to the work of computers as well.

5. Use cognitive offloading authentically

Echoing earlier arguments about authentic use of technology in education (e.g., Herrington and Kervin 2007), cognitive offloading needs to be thought of as an authentic part of work. For an assessment to be considered authentic in a digital world, by default it should allow the same cognitive offloading that a professional would use.

References

Bertram-Gallant, T. (2015). Leveraging institutional integrity for the betterment of education. In T. Bretag (Ed.), *Handbook for academic integrity* (pp. 979–994). Singapore: Springer.

Biddle, P. (2017, July 5). AI is making it extremely easy for students to cheat. *Wired*. Retrieved from https://www.wired.com/story/ai-is-making-it-extremely-easy-for-students-to-cheat/

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university*. Maidenhead: Open University Press.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. https://doi.org/10.1080/713695728.

Brougham, D., & Haar, J. (2017). Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization, 24*(2), 239–257. https://doi.org/10.1017/jmo.2016.55.

Carless, D. (2017). Scaling up assessment for learning: Progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 3–17). Singapore: Springer.

Chase, K., & Abrahamson, D. (2015). Reverse-scaffolding algebra: Empirical evaluation of design architecture. *ZDM: The International Journal on Mathematics Education, 47*(7), 1195–1209. https://doi.org/10.1007/s11858-015-0710-7.

Cookson, C. J. (2017). Assessment terms half a century in the making and unmaking: From conceptual ingenuity to definitional anarchy. *Assessment & Evaluation in Higher Education, 43*(6), 930–942. https://doi.org/10.1080/02602938.2017.1420138.

Davies, L., & Howard, R. M. (2016). Plagiarism for the internet: Fears, facts and pedagogies. In T. Bretag (Ed.), *Handbook for academic integrity* (pp. 591–606). Singapore: Springer.

Dawson, P. (2018). Exemplars, feedback and bias: How do computers make evaluative judgements? In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education : Assessment for knowing and producing quality work*. London: Routledge.

Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory and Cognition, 42*(2), 237–246. https://doi.org/10.3758/s13421-013-0362-1.

Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General, 144*(3), 674–687. https://doi.org/10.1037/xge0000070.

Goertzel, B., Iklé, M., & Wigmore, J. (2012). The architecture of human-like general intelligence. In P. Wang & B. Goertzel (Eds.), *Theoretical foundations of artificial general intelligence* (pp. 123–144). Paris: Atlantis Press.

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes, 37*, 112–121. https://doi.org/10.1016/j.esp.2014.09.001.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67. https://doi.org/10.1007/bf02504676.

Herrington, J., & Kervin, L. (2007). Authentic learning supported by technology: Ten suggestions and cases of integration in classrooms. *Educational Media International, 44*(3), 219–236. https://doi.org/10.1080/09523980701491666.

Herrington, J., & Parker, J. (2013). Emerging technologies as cognitive tools for authentic learning. *British Journal of Educational Technology, 44*(4), 607–615. https://doi.org/10.1111/bjet.12048.

Iiyoshi, T., Hannafin, M. J., & Wang, F. (2005). Cognitive tools and student-centred learning: Rethinking tools, functions and applications. *Educational Media International, 42*(4), 281–296. https://doi.org/10.1080/09523980500161346.

Jha, S., & Topol, E. J. (2016). Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA, 316*(22), 2353–2354. https://doi.org/10.1001/jama.2016.17438.

Jonassen, D. H. (1992). What are cognitive tools? In P. A. M. Kommers, D. H. Jonassen, J. T. Mayes, & A. Ferreira (Eds.), *Cognitive tools for learning* (Vol. 81, pp. 1–6). Berlin/Heidelberg: Springer.

Jones, M., & Sheridan, L. (2015). Back translation: An emerging sophisticated cyber strategy to subvert advances in 'digital age' plagiarism detection and prevention. *Assessment and Evaluation in Higher Education, 40*(5), 712–724. https://doi.org/10.1080/02602938.2014.950553.

Joughin, G. (2009). Assessment, learning and judgement in higher education: A critical review. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 13–27). Dordrecht: Springer Netherlands.

Lancaster, T., & Clarke, R. (2009). *Automated essay spinning – An initial investigation*. 10th annual conference of the subject centre for information and computer sciences, Kent, UK.

Mundt, K., & Groves, K. (2016). A double-edged sword: The merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education, 6*(4), 387–401. https://doi.org/10.1080/21568235.2016.1172248.

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition, 36*(Suppl C), 61–74. https://doi.org/10.1016/j.concog.2015.05.014.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Risko, E. F., Huh, M., McLean, D., & Ferguson, A. M. (2017). On the prospect of knowing: Providing solutions can reduce persistence. *Journal of Experimental Psychology: General, 146*(12), 1677–1693. https://doi.org/10.1037/xge0000334.

Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher, 33*(6), 478–485. https://doi.org/10.3109/0142159X.2011.565828.

Soares, J. S., & Storm, B. C. (2017). Forget in a flash: A further investigation of the photo-taking-impairment effect. *Journal of Applied Research in Memory and Cognition, 7*(1), 154–160. https://doi.org/10.1016/j.jarmac.2017.10.004.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745.

Spinbot.com. (2016). *Spinbot: Article spinning, text rewriting, content creation tool*. Retrieved from https://www.spinbot.com/

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology, 40*(2), 227–236. https://doi.org/10.1111/j.1467-8535.2008.00929.x.

Phillip Dawson is an Associate Professor and Associate Director of the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. He holds degrees in education, artificial intelligence and cybersecurity. His research focuses on assessment and learning matters like feedback, rubrics and academic integrity, often with influences from computer science.

Chapter 5
Preparing University Assessment for a World with AI: Tasks for Human Intelligence

Margaret Bearman and Rosemary Luckin

Abstract Artificial intelligence (AI) is changing how people work and live. There is an urgent need for higher education to recognise this change and respond. Assessments are key to this response. Old models of assessment draw on pattern recognition and recall, which will be of considerably less use to newer forms of employment. This chapter defines artificial and human intelligences, then distinguishes between them, outlining the things that humans can do that machines cannot and vice versa. In particular, we suggest personal epistemology ("meta-knowing") and evaluative judgement (making judgements about quality of work) are peculiarly human. Consequently, higher education assessments should focus on these capabilities. We offer the example of a critical appraisal assessment as an illustration. By deliberately focussing assessments on areas where humans think better than machines, we can enhance graduates' abilities to navigate through an AI-enabled world.

5.1 Introduction

Artificial intelligence (AI) is currently changing the face of professional work. For example, tax companies employ deep learning systems to compile tax returns (IBM 2018) and police use facial recognition to identify suspects (Anthony 2017). If future employment relies on a possibly exponential integration of intelligent machines into the workplace, then it is critical that universities recognise this both in their curricula in general and in their assessment in specific. This

M. Bearman (✉)
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: margaret.bearman@deakin.edu.au

R. Luckin
UCL Knowledge Lab, University College London, London, UK
e-mail: r.luckin@ucl.ac.uk

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_5

chapter explores the implication of the rise of AI for assessment in higher education.

The authentic assessment literature suggests that assessment should correspond to real-life conditions, thereby developing a relationship between the university and the workplace (Villarroel et al. 2018). AI is already part of this authentic landscape. Moreover, sustainable forms of assessment are those that develop skills which can be used into the future, not just for immediately passing the unit (Boud and Associates 2010). Ideally, completing any assessments should increase learners' long-term capacity to work and learn. Therefore it is critical that assessments also reflect future employment and practice environments, which will be supplemented with AI.

Given the considerable implications of AI, this chapter takes some time to consider the broad landscape. The first section defines both AI and machine learning and provides a concrete example of the current use of AI in assessment: automated marking. Next, we turn to describe human intelligence, and from this discussion focus on assessments that can develop meta-intelligences. In particular, we focus on assessments that develop both students' understanding of knowledge itself (personal epistemology) as well as their understanding of what constitutes quality (evaluative judgement). Finally, we provide an example of a re-imagined assessment that would prepare learners for a society, where AI is integrated into the fabric of working and living.

5.2 AI and Machine Learning

5.2.1 What Is AI?

AI is a term that is frequently used but not always understood. As Scherer (2015) notes, it is very difficult to define AI, partly because we lack a good definition of "intelligence"; consequently, there is little consensus. A Wikipedia entry on AI notes that it is simply "intelligence demonstrated by machines" (Wikipedia 2018). By this definition, almost every technological device we own may be considered intelligent from our old-fashioned calculators to our map apps on our smartphones. The intelligence of these machines is constrained by their hardware as well as their software.

In this chapter, we focus on a particular type of 'more intelligent' AI, which demonstrates what Russell and Norvig (2016) might call 'rational'. This type of AI seeks to take the "best possible action in a situation" (Russell and Norvig 2016, p. 30). We can intuitively understand the difference between using a calculator to compute a statistic (designed to be a tool) and software that implements a statistical calculation without input from humans (designed to take the best possible action).

In AI terms, this is also reflected in the fact that we have two basic implementations of AI: AI that does not learn and the actions that must be programmed in advance by the system developer; and AI that learns and that can change its behaviour on the basis of new information. The latter systems use "machine learning" techniques. Machine learning refers to those algorithms and systems that "improve their knowledge or performance with experience" (Flach 2012, p. 3). Machine learning is frequently reported in the popular press. Examples include AIs that beat humans at games such as go or chess, or that drive autonomous vehicles, or are able to respond to our spoken requests using Natural Language Processing, or can decide if the photo of us taken in the e-passport gate at the airport is the same person as our passport photo. Machine learning is also used by almost every email provider to filter spam; every email we receive has some kind of intelligent software that is constantly identifying new forms of 'junk'.

Turing once famously proposed an "imitation game" whereby an interrogator had to judge from a transcribed response whether a machine was imitating a human, or was in fact a human (Turing 1950). However, this masks some of the more important features of AI. Machine learning is not the same as human learning, nor is machine intelligence the same as human intelligence. There are fundamentally three main components to machine learning AIs. Firstly, there are *data* about concepts or phenomena to be learnt, which act as input to a computational system containing one or many learning algorithms. Secondly, the system iteratively calculates outputs, based on these inputs, often many millions of times, and thirdly, the system uses necessary power and memory capacity, frequently entailing modern processing and cloud storage. It is this 'perfect storm' of the coalescence of advanced machine learning algorithms, large data sets and affordable powerful computer processing and memory that has resulted in the surge of AI applications being scaled across the globe.

These three components that comprise machine learning may be easiest to understand with the example of email filters, which are so pervasive that we barely notice them. All incoming email forms the *data*. The software classifies email (*inputs*) as either 'junk' or valid communication (*outputs*) based on its learning from millions of emails, keywords and email features. The system classifies email many millions of times, drawing on necessary *computational power and memory capacity*. During this process, users can flag "junk" and the email filter will take account of this. However, while AI outputs can 'imitate' human outputs (and vice versa) this does not mean that human intelligence and AI are equivalent. In our email filter example, if we were to scan by hand, we might remove more junk but be significantly less efficient as we would be overwhelmed by the amount of junk email we would receive. It is a task well suited to AI not human intelligence. It is AI's capacity to undertake certain tasks more effectively and efficiently, that is shaping how humans work.

Work is being irrevocably changed by AI and machine learning. The popular press sometimes predict a fearful future: "…advances in AI are disrupting industries and posing a threat to the job security of millions of workers worldwide. The jobs of office clerks, receptionists, customer service reps, analysts, marketers, doctors, attorneys, underwriters and creatives could be replaced by AI in the next decade" (Kucheriavy 2018). While there is no question that there is digital disruption, this underplays the advantages we have enjoyed thus far (such as the email filters). We would like to take a more nuanced approach and outline the relative advantages of the different types of intelligences.

5.2.2 The Value of AI and Machine Learning

AI works at scale and at speed. Given appropriate infrastructure, AI is always available and can be responsive at any time at a constant cost. AI can draw from, what is for humans, unimaginably large data sets. Likewise, machines can perform tasks – and therefore learn – seemingly endless repetitions. Additionally, AI is not consciously bounded by social norms or prepared to be satisfied with notions of 'good enough' outcomes, which can limit human thinking (Scherer 2015).

Given these broad advantages, there are particular characteristics of tasks that are likely to be supplemented or replaced by machine learning. These are tasks that have large scale data availability, are useful to many people and have clear conditions denoting success. The latter may be identified via rules (as with games such as go or chess) or it may be available through other means such as analysis of a large amount of human data. The criteria for success however may be more nuanced than first appears and we humans need to ensure that we make very careful and thoughtful judgements about success. For example, Alpha Go built by Google Deepmind beat world champion go player Lee Sedol (Newcomb 2016). By this we mean that Alpha Go won more games. However, if go is a game that is about building social relationships with other people in one's community, then one might legitimately ask if AlphaGo did in fact 'win'. In other words, if we look beyond the scoring of games to the broader cultural contextualisation of the game then we may judge AlphaGo to have been less successful, since it did not build any social relationships with fellow go players or build cultural capital beyond AI.

Despite the rapid rise of AI, the fundamental principles and therefore the fundamental limitations underpinning machine learning have not changed a great deal over many years. In our view, the AI revolution has been prompted by: a vast improvement in physical computational power; the overall shift to a digital landscape enabling large stores of data; and innovations in developing simultaneous computational processes. While there is no question that AI continues to expand its capabilities, there is as yet, no indication that AI can do *everything* the same as or better than we humans. Of course, we cannot account for the future, however we

speculate that the next shift in AI will not be to make machines more human, but to provide additional capability for us humans that we cannot yet envisage.

5.2.3 *Example of AI in Higher Education: Machines as Assessors*

Machine learning is already having a radical impact on higher education assessment processes. Possibly the most prominent use of machine learning is with respect to the automated scoring of written texts, a field that is around two decades old. We will use this as an example to illustrate the relative merits of machine learning and human intelligence and why reimagining assessment is so important in the age of AI.

Automated essay scoring is based on machines learning the characteristics of essays associated with grades through a corpus of already marked material. A submitted essay is compared with this pre-defined set. The essay is given a numerical grade based on linguistic similarities between the submission and the corpus of material (Kovanović et al. 2015). Under these circumstances, automated essay scoring is shown to be comparable, and possibly better than, human graders (Landauer 2003). More recently, work is using similar types of systems to provide information to students about their work; the key advantage here is that the student does not have to wait for a human to have time for human feedback, they can get comments as they need it. A study of 1.3 million student essays indicated that providing students with formative feedback scores is correlated with an improvement in essay quality (as marked by the automated system) (Foltz and Rosenstein 2015). This indicates both the potential for such systems and the scale of their reach.

The limitations of automated scoring systems are considerable. There needs to be a relatively constrained stable set of responses, which means that the essay prompts have to be restricted to particular topics and tasks. A recent study suggests that such systems are bounded by the quality of the original data that was used to train the machine to associate grades with certain linguistic patterns (Wind et al. 2018). Finally, the algorithms cannot provide an explanation or justification of the grade: and this accountability is often one of the key demands of educators (Luckin 2018).

Taken together, this snapshot of a field of research indicates the strengths and weaknesses of current approaches to machine learning for use in education. With respect to strengths, machine learning techniques can be very accurate and used at very large scales to both assess and improve student essays. The efficiency gains and the capacity for information-as-needed are key advantages. On the other hand, as the systems are based on patterns derived from a particular data set, these advantages are subject to the topic and task being amenable for this type of grading and the original human grading to be appropriate. This means that this approach is not useful for assessing innovative or unique types of thinking and the in-built biases of humans may become amplified in grading processes. Further, feedback is

conceptualised as a score, and does not involve a contextualised explanation of the reasons for the score. Finally, and possibly most importantly, automated essay scoring illustrates how current machine learning approaches align with the assessment we already do, rather than promoting the assessment we will need into the future.

5.3 Human Intelligence

While AI approaches to current assessment may help us deliver the assessment we are already doing, this chapter is concerned with exploring assessment that will be useful to us in an AI world. It is worth taking a slight detour to reflect on what human intelligence has to offer, particularly in an increasingly digital world.

Intelligence is an altogether tricky term. On the one hand, it has traditionally been cast primarily as a "mental ability", even when taking account of environmental factors (Sternberg 1997). However, in the era of AI, it forces us to consider 'mental' as being somewhat different to an abstract cognitive state. As humans, our thinking intersects with our perceptions, emotions and experiences. We have often over-emphasised the rational (Buchanan 1992); our most productive and creative thinking is sometimes unconscious (Lawson 2006). While we sometimes think of our bodies as limiting (being tired or angry or stressed), they also provide huge advantages. Humans have sensory inputs that allow them to experience and think differently at different times, and respond to different contexts. Human intelligence is intertwined with lived experience, thinking draws on memories, interpersonal relationships as well as an infinite array of materials and contexts. Today's thinking becomes yesterday's memory. Altogether this means that human thinking is remarkably diverse with influences depending on the time of day, the state of the body and the social pressures of the era.

In 2018, Luckin proposed a model of human intelligence with seven interwoven elements, which takes account of the differences between people and machines. She writes (Luckin 2018, p. 64): "… I use the word *element* because it characterises something that is essential and significant. Each element in this interwoven intelligence model does not come with expectations of being a particular size or shape, or of fixed dimensions. Throughout our lives we develop our intelligence within and across the seven elements that I propose, some more quickly than others. I must stress that these are not separate intelligences, but different elements over a complex interwoven whole." Luckin's model includes *academic intelligence* or ways of understanding and applying knowledge and *social intelligence* or how interacting with others helps us understand and apply knowledge, both as individuals and as communities. It also proposes five forms of *meta-intelligences* – ways through which we develop and regulate a knowledgeable understanding of ourselves (Luckin 2018). These are what allow us to: direct our efforts; reflect on our experiences; set goals for the future and explain these goals to others; and identify deficits or strengths. The five meta-intelligence elements are: meta-knowing (knowing what knowledge is and how to use it); meta-cognition (knowing and directing our mental

activity); meta-subjective intelligence (knowing and regulating our motivations and emotions); meta-contextual intelligence (knowing how our physicality interacts with our particular context); and perceived self-efficacy (judging how well our intelligence can equip us to succeed in particular situations). Looking at this list, it is clear that human and computer intelligence are very divergent.

One clear attribute of human intelligence, which is not shared by AI, is being able to define what constitutes success. While algorithms can alter themselves, and hence derive new and innovative combinations, machines as yet can't define and defend quality. Machine learning achieves 'best possible' through pre-defined indicators or by deriving indicators based on patterns of previous human behaviours. Success therefore, depends on what is already known or what has already happened. On the other hand, humans create indicators of success in the present, with an eye to the future. Human intelligence can reconcile competing views of success by finding a new solution, and then persuade others as to why this is better. This presents us with some guidance as what might be useful to assess in humans, as we explore in the next section.

5.4 Assessing Human Intelligence

Many assessments are drawn from a time when machines could not perform any cognitive functions and focus therefore on *academic intelligence* alone. While academic intelligence still holds a very important place in working and learning, the relative emphasis by assessments on this single element of intelligence presents a particular misalignment with modern life. AI and machine learning's strengths are in applying knowledge – and machines are increasingly better at this than humans. But they have significant limitations in other areas. We suggest therefore that we should be directing our students towards things that they can do better than machines.

Students adapt their behaviours because of their assessments (Swanson et al. 1995). If we want students to develop certain intelligences then it is highly important that we assess these. More than this, we want *sustainable* assessment: assessment that will assist the student learn into the future (Boud and Associates 2010). Together, this leads us to ask: what elements of human intelligences should we be assessing and why?

The meta-intelligence elements seem like a good place to start. These elements are the conscious direction of our own selves. They allow us to project from the past and the present to think or act differently into the future. We think two elements in particular are very useful: meta-knowing and perceived self-efficacy.

5.4.1 Meta Knowing

Meta-knowing is essentially what is called 'personal epistemology' or thinking and beliefs about "knowledge and knowing" (Hofer 2001, p. 355). Perry's 1968 seminal work on intellectual development in liberal arts college students was one of the first studies to explore how personal epistemologies develop. He described intellectual shifts; most students went from thinking about knowledge as absolute and authoritative to assume a more nuanced understanding that knowing is complex, relativistic and ambiguous. This general development from a more naïve to a more sophisticated understanding of knowledge has been marked by many others (Luckin 2018). A sophisticated personal epistemology is considered important to our intellectual advancement. We need to understand what knowledge is in order to know what it means to know something. We need to recognise what evidence means – with all its flaws and limitations – and how to make judgements based on that evidence in a complex and ambiguous situation in order to construct our knowledge. A personal epistemology is also very human – unique, contextualised, based on past experiences and full of personal meaning. It enables us to make decisions about what we wish to know and how we come to know it. It is hard to imagine an AI with a personal epistemology.

Interestingly, Perry (1968) also noted that the *assessment practices themselves* had exponentially changed in the first half of the twentieth century, from involving one frame of reference to requiring consideration of two or more frames of reference. This illustrates two things: firstly, personal epistemology development and assessment are strongly inter-related; and secondly, that what we require from intellect and assessment alike can shift radically across time.

5.4.2 Perceived Self-Efficacy and Evaluative Judgement

Luckin (2018, p. 56) suggests that perceived self-efficacy may be the most important element contributing to human intelligence. She notes that perceived self-efficacy: "requires an accurate evidence based judgement about our knowledge and understanding, our emotions and motivations and our personal context. We need to know our ability to succeed in a specific situation and to accomplish tasks both alone and with others. …" Luckin's formulation of perceived self efficacy has strong parallels to the idea of 'evaluative judgement'. Evaluative judgement is "the capability to identify the quality of work of both self and others" (Tai et al. 2018, p. 467). In other words it's how we assess work, determine whether a piece of work is 'good' or 'bad' and why. It is increasingly seen as a purpose for assessment: that is, doing assessment should develop students' evaluative judgement so that they can continue to do the work into the future.

Evaluative judgement differs slightly from perceived self-efficacy. In particular, it links understanding quality in our own work with the notion of understanding

quality in general. That is to say, we can't assess how well we are doing unless we can judge what we are striving for. Overall, we propose that a high level of perceived self-efficacy is required to make good evaluative judgements. Conversely, an ability to make evaluative judgements reflects a sophisticated perceived self-efficacy.

Higher education by its very nature builds students' evaluative judgement capabilities. For example, at the end of a science degree, a student can recognise what constitutes rigour in a scientific experiment. They can critique their own work and that of their peers; identify deficiencies and remedies. Through doing their degree – critically reading papers, doing simulations and conducting experiments – they may come to learn what quality means for a scientist and how this quality is represented in their own work. However, 'evaluative judgement' is not necessarily about the 'best possible' – it is about interrogating what constitutes 'best' within a particular situation. Both perceived self-efficacy and evaluative judgment are discipline or context dependent (Luckin 2018; Bearman 2018): that is, they are not 'generic' capacities. Possessing strong perceived self-efficacy or evaluative judgement in art history does not imply that there is any link to an equivalent intelligence or capability in carpentry.

5.4.3 Assessments That Build Personal Epistemologies and Evaluative Judgement

Personal epistemology and evaluative judgement are strongly linked. In order to determine what constitutes 'good' work, a student has to possess a sophisticated series of beliefs and thoughts about what knowledge is and how it can be applied. These are uniquely human approaches, and produce uniquely human outputs. It shifts the focus from being successful (yes or no), to coming to understand how success is constituted. By focussing on both developing personal epistemologies and also evaluative judgement, we can graduate students who will understand how to identify and qualify 'success'. In other words, our assessments can more overtly interrogate the links between quality, knowledge and the work of themselves and others. Our students will come to manage subjectivity, variance and diversity.

Assessments that build evaluative judgement through judging the work of self, peers and experts are described in the literature (Boud et al. 2018). Assessments that measure epistemic development are also reported (Hofer 2001). What we are proposing is that, in a time of AI, a proportion of our assessments should both assess and develop personal epistemology and evaluative judgement. We will explore what this explicit approach might look like through the example of a critical appraisal assessment in an undergraduate health sciences degree.

Critical appraisal assessments and similar are provided in many courses but are often seen as peripheral rather than core foundational knowledge. We suggest this does not reflect the needs of the digital age, when the exams or essays which test

academic knowledge should become increasingly less important in higher education. Critical appraisal develops personal epistemology (coming to understand what a scientific knowledge is) and builds evaluative judgement (capability to identify quality of scientific trials). Critical appraisal is not only a scientific phenomenon, similarly, students could be appraising or critiquing a short story or software code or a legal argument or an architectural drawing. However, we propose that critical appraisal is not a 'one off' assessment, but a series of tasks over time, which develop in complexity as the student develops their capabilities.

5.5 Example: Critically Appraising Scientific Papers in the Health Sciences

Assessing if a scientific study is high or low quality is not just about rules. It is about more complex notions; it is possible for a study to meet all the rules (be well designed, randomised, etc) but still be fundamentally meaningless (Bearman 2016). Therefore, our assessments should also draw on these more complex ways of engaging with knowledge. In order to do this, students need to develop their personal epistemologies. When students start a science degree, they may initially hold all scientific papers as authoritative. Over time, as they become more expert, they come to realise that not all studies are good studies, and further, what constitutes scientific knowledge is contestable. However, teachers may make assumptions that this development in thinking will be osmotically absorbed through the discourses, words and environment. Assessment – coupled with teaching of course – can provide deliberate opportunities to extend personal epistemologies.

In many ways, assessments are arbiters of what the student sees as important. We suggest that assessments can and should prompt students to extend their meta-intelligences, particularly meta-knowing. Students can always be asked to articulate: what type of knowledge is held within this task? What meaning does it have for others? What meaning does it have for me? We also suggest that these ideas should be accompanied by clear rationales as to why these types of reflections are important. This will assist students to see both: (1) value developing their personal epistemologies and evaluative judgements; and (2) simultaneously building their understanding of their future in a digital age. In other words, students should be made aware that part of what these reflections are doing is articulating what parts of the task are better completed by machines and what is better done by humans.

In our example, which is set in an undergraduate health sciences degree, we propose three linked critical appraisal tasks across three years of study (see Fig. 5.1). At first year, the expected level of personal epistemology is lower and we wouldn't anticipate a good understanding of what constitutes a high quality scientific study. The first year student is supplied with a publication, say a randomised controlled trial of a medication, and a standardised checklist of features that indicate study design quality (such as the CONSORT statement (CONSORT 2010)).

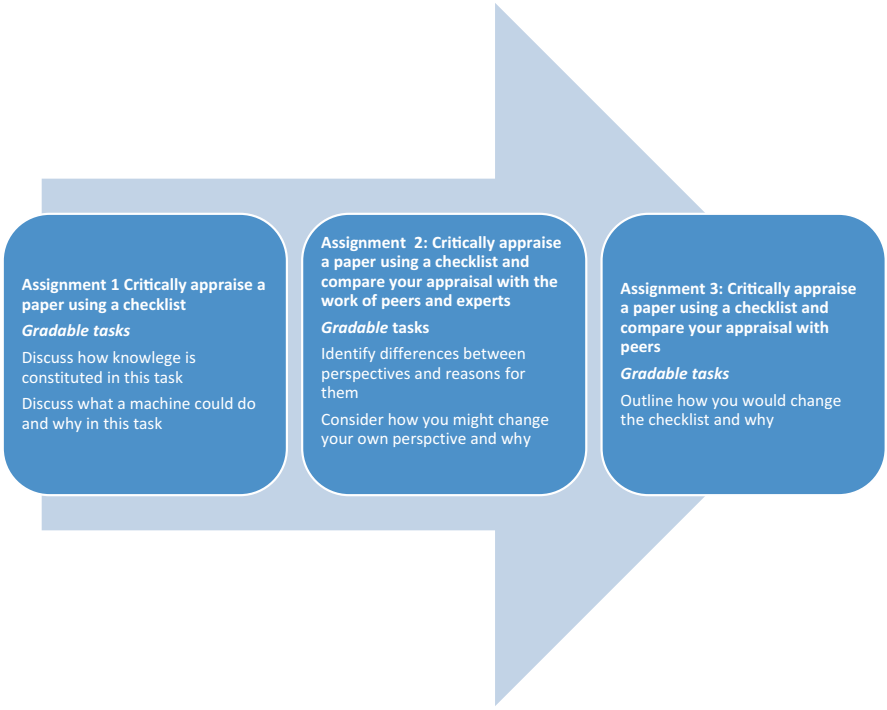


Fig. 5.1 Three linked critical appraisal tasks across 3 years of study

Examples of checklist items might be: is the trial randomised? if so, is it 'double blinded'? and so on. The student may work through the checklist and see if they can see these features. In doing so, they are learning to identify scientific knowledge and its markers of study design quality. However, this assignment doesn't ask a student yet to *do anything an AI can't do*. This first part of the assignment is about rules-based processing. However, the second part of the assignment targets building personal epistemologies and evaluative judgement.

In the second half of the assignment, we propose adding a reflective task. The reflection should be framed by a primer which outlines the differences between following the rules and understanding the content. This brief should also outline that, while graduates need a degree of competency with respect to rules, following rules and recognising patterns are areas where AIs should be able to assist. On the other hand, the human intelligence contribution is to be found within a series of reflective prompts on completion of the checklist itself. These prompts could commence with formative reflections: What do I learn from this paper and how? What do I learn from this task and how? These might feed into two gradable reflections: What parts of this task are based on rules? What sort of interpretation is required and why? Another possibility is to ask students: what part of the task you have just done might be performed by machines and why?

In second year, students are provided with a similar task. As before, students have a rules-based checklist, and a personal reflection, exploring how they made meaning of the task. This time the task is increased to take account of social intelligence. The students are asked to compare their checklist and their reflections with that of peers and experts. The gradable tasks might be: (a) identifying differences between self, peer and expert work; (b) articulating reasons for these differences; and (c) considering how they would rework their checklist and reflections and why. This type of assessment deliberately builds evaluative judgement, which necessarily develops a deeper understanding of both the content and of the foundations of disciplinary knowledge. It is also possible to include AIs as a reference point; for example, in the future, the students could compare their checklist with those of an AI as well as peers and experts.

In third year the task might look quite different again. Students might be given a range of papers, and a checklist. Tasks now should ask students to critique and compare but then subsequently to generate work. For example, students may be asked to critique the checklist itself; to propose alternative items and any doubts or ambiguities associated with these changes. In this way, they come both to understand what the standards are, and also to think about the concept of standards in general.

5.6 Pedagogical Features That Focus on Human Intelligence

This example has three pedagogical features that can be replicated in other assessments, which explicitly seek to build human intelligence. Firstly, *teachers provide rationales* for the assessment through a primer. This models a very human attribute – providing reasons for actions. Part of this rationale is a deliberate focus on distinguishing what students, as humans, can contribute that AI, as machine intelligence, cannot. Next, students are asked to *make – and later defend – their evaluative judgements* about the quality of their own work and others, with a view to understanding their role in developing criteria for success. This is not a concern for being 'correct', rather it is about coming to understand what constitutes quality in a particular domain. Again, this is a human endeavour: while algorithms can alter themselves, and hence derive new and innovative combinations, machines as yet can't meaningfully set parameters for success or quality. Finally, students are asked to describe *how knowledge is constituted in the task at hand*. The focus on knowledge in the particular task is important: it avoids students tokenistically repeating a theory of knowledge without considering how it applies to a particular domain. This is real intellectual work, and it may require us to think about grading differently. It is likely that quality work will show itself in divergence rather than convergence.

These three pedagogical features – teacher rationale, make and defend evaluative judgements, reflection on how knowledge is constituted in the task at hand – are already part of the assessment landscape. We are suggesting three particular innovations here. Firstly, that these three features jointly present an assessment landscape that maximises human endeavour. Secondly, that these types of assessments tend to

be secondary (10% reflections, 10% critical appraisals), while we are suggesting that they need to be primary. Finally, we propose that at every step, students are directed to distinguish what they can contribute and what machines can do. In this way, we can offer preparation for an employment landscape that may look radically different.

This type of approach is not a recipe for assessment in a time of AI. Other assessments may focus more on, say, social intelligence or meta-cognition, or the other forms of intentional 'meta' intelligences outlined earlier. These may be best assessed through creative or embodied tasks, which produce outputs that are qualitatively distinctive rather than quantitatively equivalent. If there is one thing that our comparison between AI and human intelligence reveals, it is that AI can follow recipes better.

5.7 Conclusions

In this chapter we have compared AI and human intelligence. AI works at scale and at speed. However, human intelligence is complex and multi-faceted. It includes our physicalities, experiences and relationships as well as our individual capacity for reason. Additionally, the ability to direct and regulate ourselves is key to human intelligence. However, many assessments focus mainly on our academic intelligence. This is not a new critique but one that is becoming more urgent, given the shift in technology and employment opportunities.

The advent of AI therefore offers as many opportunities as it does challenges. It may be the catalyst that forces us to focus on what really counts. Perry's (1968) work demonstrates how liberal arts assessments changed between 1900 and 1960, to take account of multiple perspectives. This paralleled – or possibly enabled – a shift in society towards embracing multiplicity of perspectives. We have an opportunity to develop assessments that create a future workforce that not only are employable, but also can contribute to broader shifts of living in a digital world.

References

Anthony, S. (2017, June 6). UK police arrest man via automatic face-recognition tech. *Ars Technica*. Retrieved from https://arstechnica.com/tech-policy/2017/06/police-automatic-face-recognition/.

Bearman, M. (2016). Quality and literature reviews: Beyond reporting standards. *Medical Education, 50*(4), 382–384. https://doi.org/10.1111/medu.12984.

Bearman, M. (2018). Prefigurement, identities and agency: The disciplinary nature of evaluative judgement. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work* (pp. 147–155). London: Routledge.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (Eds.). (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. Abingdon: Routledge.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues, 8*(2), 5–21.

CONSORT. (2010). *The CONSORT statement*. Retrieved from http://www.consort-statement.org/.

Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511973000.

Foltz, P. W., & Rosenstein, M. (2015). *Analysis of a large-scale formative writing assessment system with automated feedback*. Paper presented at the second (2015) ACM conference on learning @ scale (pp. 339–342), Vancouver, Canada.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational Psychology Review, 13*(4), 353–383. https://doi.org/10.1023/A:1011965830686.

IBM. (2018). *H&R Block's 70,000 tax pros are working with Watson to help clients maximize their refunds*. Retrieved from https://www.ibm.com/watson/stories/hrblock/.

Kovanović, V., Joksimović, S., Gašević, D., Hatala, M., & Siemens, G. (2015). Content analytics: The definition, scope, and an overview of published research. In G. Siemens, C. Lang, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 77–92). Edmonton: Society for Learning Analytics Research.

Kucheriavy, A. (2018, February 26). Artificial intelligence will take your job: What you can do today to protect it tomorrow. *Forbes Community Voice.* Retrieved from https://www.forbes.com/sites/forbestechcouncil/2018/02/26/artificial-intelligence-will-take-your-job-what-you-can-do-today-to-protect-it-tomorrow/#702da93b4f27.

Landauer, T. K. (2003). Automatic essay assessment. *Assessment in Education: Principles, Policy & Practice, 10*(3), 295–308. https://doi.org/10.1080/0969594032000148154.

Lawson, B. (2006). *How designers think*. London: Routledge.

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. London: UCL IOE Press.

Newcomb, A. (2016, March 15). Go Grandmaster Lee Sedol reflects on losing series to google's computer. *ABC News*. Retrieved from https://abcnews.go.com/Technology/grandmaster-lee-sedol-reflects-losing-series-googles-computer/story?id=37656509.

Perry, W. G., Jr. (1968). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.

Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach, global edition*. Harlow: Pearson.

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology, 29*(2), 353–398.

Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist, 52*(10), 1030–1037. https://doi.org/10.1037/0003-066X.52.10.1030.

Swanson, D. B., Norman, G. R., & Linn, R. L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher, 24*(5), 5–11. https://doi.org/10.3102/0013189X024005005.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 49*, 433–460.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396.

Wikipedia. (2018). *Artificial intelligence*. Retrieved from https://en.wikipedia.org/wiki/Artificial_intelligence.

Wind, S. A., Wolfe, E. W., Engelhard, G., Jr., Foltz, P., & Rosenstein, M. (2018). The influence of rater effects in training sets on the psychometric quality of automated scoring for writing assessments. *International Journal of Testing, 18*(1), 27–49. https://doi.org/10.1080/15305058.2017.1361426.

Margaret Bearman is a Research Professor within the Centre for Research in Assessment and Digital Learning, Deakin University. Margaret's research focusses on learning and teaching in higher and professional education. She is particularly interested in ways of improving feedback, the sociomateriality of standards, and what it means to be an online learner in a digital world.

Rosemary Luckin is a Professor of Learner Centred Design at the University College London Knowledge Lab. Her research involves blending theories from the learning sciences and techniques from Artificial Intelligence (AI) to visualize the detail of a learner's progress intellectually, emotionally and socially. Rose is also a Director of EDUCATE: a London hub for Educational Technology StartUps, researchers and educators to work together on the development of evidence-informed Educational Technology.

Chapter 6
Repositioning Assessment-as-Portrayal: What Can We Learn from Celebrity and Persona Studies?

Rola Ajjawi, David Boud, and David Marshall

Abstract Graduates must learn to present a version of themselves that aligns with the expectations and norms of their discipline or profession, organisations in which they might work and the public at large; and to be able to adapt as they change careers and workplaces. Persona studies offers a lens for reimagining more authentic forms of assessment design, ones where students can strategically construct their persona in different media and for different audiences through meaning making in relation to the worlds of work and study. We do this by mapping new possible pathways of assessment as portrayal across dimensions of persona offering specific examples of design.

The digital permeates what we do at work, altering fundamental aspects such as where work is done, by whom and how. Whereas in the past individuals might enact their professional identities mainly in the confined spaces of the workplace and the profession, the membrane between professional and private life has become more porous (e.g., flexible work patterns, accessing emails anywhere, etc.). The landscape of work is also rapidly changing to become less stable and more transient, with short-term contracts, requiring individuals to assume responsibility for their own career management and employability (Sullivan and Baruch 2009). Careers have become more fluid with individuals changing direction, not only out of necessity, but also to accommodate new interests or to seek better opportunities. Together these have resulted in a greater need to construct public identities visible wherever

R. Ajjawi (✉) · D. Boud
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: rola.ajjawi@deakin.edu.au; david.boud@deakin.edu.au

D. Marshall
Faculty of Arts and Education, Deakin University, Melbourne, VIC, Australia
e-mail: david.marshall@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_6

they operate. Workers increasingly create online profiles and/or portfolios to present themselves and their achievements to potential employers.

Students' ability to strategically navigate the personal-public permeable spaces and construct public identities is now fundamental to being career-ready. Assessment plays a critical role in career-preparedness through ensuring that not only do students meet the course outcomes but that graduates can meet the expectations of various stakeholders (including future employers). Graduates must learn to present a version of themselves that aligns with the expectations and norms of their discipline or profession, organisations in which they might work and the public at large; and to be able to adapt as they change careers and workplaces (Oliver 2015).

Universities have responded to these challenges through developing alternative forms of assessment that bring university and industry into closer alignment, for example, through the use of authentic tasks, work-integrated learning (WIL) assessment, portfolio-based assessment and digital credentialing. These advances in assessment tend to provide opportunities for students to portray their achievements in distinctive and more diverse ways. Therefore, we are already seeing a shift towards what we have termed 'assessment-as-portrayal' (Boud and Ajjawi 2019). Assessment-as-portrayal, by design, propels a more strategic form of portrayal of self for the digital world. These types of assessment tasks provide the possibility for students to construct a representation of their personal and professional selves, their achievements, their image, and how these may be conveyed into different potential work settings.

In this chapter, we seek to revisit assessment-as-portrayal using the lens of persona studies. Persona studies has evolved from studying how celebrities construct their strategic identities to a wider exploration of how we all navigate the public world (Marshall 2016). We propose here that the concept of persona can provide a productive way to understand the contemporary configuration of identity in digital assessment tasks. Through using key elements of persona studies, we wish to start a conversation about how existing assessment portrayals can be reimagined by drawing on the insights that have come from studies in quite different domains. By examining assessment portrayals we seek to understand how students can strategically construct their persona through different media and for different audiences through meaning making in relation to the worlds of work and study. Further, we borrow from research into persona studies in order to rethink how we might judge the quality of portrayals for future work. In the process, we draw critically on concepts from persona studies both to raise useful ideas about how we might think about portrayal and to avoid falling into the trap of legitimising strategies, used by celebrities and others, to deliberately create what might be argued to be 'false' personas.

We start by discussing the changing forms of portrayal students develop through schooling to university. We explore some of the contemporary drivers of assessment-as-portrayal and then reposition assessment-as-portrayal using the lens of persona studies to integrate personal and professional identities into assessment and assessment products, which hitherto have typically valued conformity over distinctiveness. Helping students to construct their public personas in relatively safe and scaffolded environments ahead of graduation better prepares them to negotiate the

communities they seek to join. Finally, we highlight areas of tension, affordance and future development that may assist in better preparing students for their future trajectories.

6.1 Changing Forms of Portrayal for Students

One of the transitions that students make as they progress in their education, is in terms of how and to whom they present their work. In primary school, their focus is on their teacher; as they proceed through high school they present themselves and their work to multiple teachers and finally to an external assessment which ratifies their achievement and enables them to progress to higher education. There they are confronted with producing work for multiple lecturers and tutors and increasingly to an external audience in professional placements and internships. In some courses, they are expected to produce a portfolio, and this may be curated by them as a new form of presentation of self. Interestingly, assessment as portrayal is becoming increasingly popular in schooling as students are encouraged to use portfolios and blogs to speak with peers, parents and their teachers. This sequence represents what we might call the official pathway of (re)presentation: at each level different parties view the work of the student and on occasions, pass judgement on it in the form of marks or grades which are at various stages inscribed in an assessment transcript.

Alongside this official pathway, students learn to present themselves to the world in many other ways. Initially to their peers and then through a range of organisations – sporting clubs, community organisations, etc. – and importantly, through social media. They create a view of themselves to others in the informal lifeworld. A characteristic of these presentations is their diversity: the students are not all in the same course, or working on the same assignments or even doing the same things: they have their own interests, their networks and contacts. They may present themselves quite differently to different audiences. As they move beyond their family and school into the wider world, these contacts and sites of activity proliferate.

Over time, as students get closer to the world of work, some take the opportunity to use other forms of presentation, through twitter, blogs or personal websites. Sometimes these are oriented towards hobbies or professional work, sometimes to the informal world. They communicate to others and they are in turn influenced by them. Problems can arise of course when the activities of these two worlds become comingled: access to digital media has seen a conjunction of the formal and informal worlds. Students want to present themselves to the world of employment, to differentiate themselves from the many hundreds of graduates who have the same degree and have met the same learning outcomes. They want to show what they can do and portray themselves as an attractive proposition in the world of work through a combination of curricular and co-curricular activities and artefacts.

We foresee a dilemma in gaps between presentations of self. The formal portrayal validated through grades and transcripts is an austere presentation, which is

poor at differentiating between individuals, and poor at providing descriptions for people other than educators. The informal portrayal through social media has an audience of friends and acquaintances and uses colloquial forms of discourse, which continues its original purpose of connecting friends and family. The common form of presentation of social media does not communicate much of interest to employers, nor does it use language and images that they identify as professional, or even appropriate to work. Furthermore, the traditional transcript does not portray students' achievements for each of the stipulated competencies or graduate learning outcomes, omitting specific achievements that might appeal to different employers. The limitations of traditional spaces creates opportunities for new possibilities: forms of presentation which draw on digital media familiar from the informal and public spaces alongside verified information presented in ways that show what a student/graduate can do. The use of such space can draw on the model of the learning portfolio in which products are assessed and validated – either conventionally or through digital credentials – but located in a user-friendly form curated by the student for particular purposes. It goes beyond the typical professional curriculum vitae or portfolio (collection of evidence of achievement) through allowing and incorporating more expressive and unverified representations of the self that can paint a picture of an actor in the world who has achieved recognised outcomes but who is also a more rounded individual engaged beyond the course with aspirations to do more. Thus, the presentation can be a hybrid between the formal – ratified by authority – and the informal, unratified but curated to suit a professional representation, e.g., blog entries, multimedia created resources, professional network maps. While the technology for this should already exist in courses that extensively use portfolios; what needs to be developed are various assessment protocols and practices that can be adopted to suit the varying ends of the producer.

6.2 Assessment, Portrayal and Persona

As assessment design faces the challenges of portrayal, it is also grappling with more basic concerns. In 2014, Vu and Dall'Alba foreshadowed an ontological turn for assessment design that integrates knowing, acting and being. They argue that assessment design must be considered in more sophisticated ways than 'do the tasks have "real-life value"?' if we want them to prepare students for the future. Drawing on Heidegger's work, they propose that authenticity should not only be an attribute of tasks – which is how the term is often used in assessment – but a quality of educational processes that engages students in *becoming* more authentic. To be considered authentic, assessment processes should engage the whole person; to integrate what students know, how they act and who they are. *Becoming* can be thought of as how we are in the world, the self in relation to others and things; it has personal and public dimensions (Vu and Dall'Alba 2014). Hager and Hodkinson (2009) used the metaphor of 'learning as becoming' to take account of both the individual (learning

as acquisition) and the relational (learning as participation) aspects of learning and the interdependence of both when designing educational courses.

Arising from a different set of concerns, persona studies evolved through the study of celebrities and their construction of a public persona. It now focuses more broadly on how we all develop an external persona. Persona can be considered a strategic identity: "a fabricated reconstruction of the individual that is used to play a role that both helps the individual navigate their presence and interactions with others and helps the collective to position the role of the individual in the social" (Marshall and Henderson 2016, p. 1). In this post-structural sense, each of us has multiple, intersecting and constructed identities, e.g., private and public, personal and professional. It is the melding of the private and public selves for a public audience that is of interest here. This is a strategic construction (conscious and active) that shifts and changes depending on the audience, the purpose, and the form of the media. Persona is a social construct that requires individual agency in the negotiation and construction of a strategic identity into a collective world (Marshall and Henderson 2016).

The drivers for making public the self in contemporary culture are not unique to celebrities. Indeed, in our own lifetimes we have seen radical changes to the presentation of work personas through portfolios and CVs made available to the public online through professional websites such as *LinkedIn*™. These encourage the strategic presentation of self as a brand to future employers – referred to as the 'branded' self (Marshall 2014). Other media outlets include *Twitter*™ where professional conversations can be tracked through hashtags such as #meded or #highereducation. For researchers, engagement with research outputs and dissemination through online portals are now part of performance indicators. Research suggests that *altmetrics* (alternative metrics) coming from mentions in social media, blogs and news outlets have a stronger correlation with citations compared to journal citations scores (Costas et al. 2015) and thus legitimizing social media and online blogging as forms of research engagement and knowledge exchange. Academic work is just one example of portfolio culture where we seek to (re)present ourselves to the outside world in mediatized forms. Transformation of forms of social connection and networking via new technologies enables individuals in this 'gig' economy to showcase themselves and their distinctiveness as well as adapt and repackage their skills and experiences (and selves) to fit the changing workforce. It is becoming an essential feature and process for increasing types of work in contemporary society.

In recent persona studies research, several elements of persona construction have been advanced. Along with the dimensions we deal with in greater detail below, it is useful to understand that the qualities of the construction of our online public self will involve a number of forms of communication – from text and image to moving images, emojis, and sometimes shared sophisticated professional quality sound and video. This online environment is a site of what persona studies identifies as "intercommunication" (Marshall 2016) where we blend the most interpersonal types of relating with the most highly mediated. And to negotiate this world, students need to become adept at the way that social media platforms construct an environment for us: in other words, students need to have well-developed digital and media literacy

as they work to produce their online and professionally-linked persona. The curriculum, including its assessment structures, can provide students with the opportunity to try on or perform particular personas with intercommunicative adeptness for the profession they seek to join or the kind of employment to which they aspire.

6.3 Reimagining Assessment Using Persona Studies

New assessment task designs are emerging in response to the employability agenda and the drivers outlined above (see also Jorre de St Jorre, Chap. 19, this volume). These enable students to portray their achievements in more public ways that communicate with future employers or clients directly and that are controlled by the student. Portfolio assessment and validated digital credentials are two prominent examples of this shift to assessment-as-portrayal. We define portfolios as the systematic collection and presentation of digital artefacts and reflections that are curated and managed by the learner as evidence of their achievements, as well as a representation of learners' personal and professional identities (Holt et al. 2016). Digital credentials (also known as digital badges) allow evidence of achievement to be more detailed and contextualised and to be shared more broadly than is possible through grades and transcripts (Miller et al. 2017). These encourage and engage students in curating evidence for employability, including workplace experiences and guidance on employer expectations (Jorre de St Jorre and Oliver 2018).

In the next part of this section, we consider these methods of portrayal (portfolios and credentialing) and ask how persona studies might help us to extend the scope of what is included in assessment. Specifically, we do this across five dimensions of persona studies – (1) public, (2) mediatized, (3) performative, (4) collective, and (5) intentional value. We describe these in turn, discussing how each dimension can provide a pathway for rethinking assessment.

1. *Public*

The *public dimension* refers to the "extended social network of the individual that includes personal friends, professional associates, plus their networks, and the systems and platforms that connect them all" (Moore et al. 2017, pp. 2-3). This spectrum of publicness might range from a tight group of friends to a massive and global audience. Our starting point in relation to assessment, is to bring students to an acceptance of their need to construct this public dimension and how they can think about what is needed when any artefact is presented publicly. If in the past, summative assessment was understood to be about the representation of knowledge achievements, the purposes of these new forms of assessment portrayals also need to be made explicit.

Possible directions in terms of assessment for students include building an online persona archive: students could chart their different public identities related to the various social media and other sites where they have a presence. Students then

justify ways that they have organised their identities related to their particular areas and disciplines and their associated profession. For example, one piece of evidence for a portfolio might include a reflective report using different media in which students discuss the different forms of social media they are using and how their use might be differentiated in relation to their personas as student and/or becoming professional. Students might be encouraged to evidence the various networks developed and the types of discourses and engagements promoted in each and how these might be viewed and adapted to different employers. Presentation decisions might include the type of blending of personal and private selves into this public identity. There are strategic decisions to be made here about an individualised and recognisable public identity. Some effort may be needed to ensure that the particular persona constructed is not blended with too much personal information that is not related to the nature of this particular persona. This emphasis might be particularly valuable as, for example, students have been found to include compromising images of themselves on *Facebook* without sufficient privacy controls (MacDonald et al. 2010). Students might be progressively scaffolded to 'expose/construct' their portfolios starting with their cohort of students in their discipline and then with the key figures in the work community of relevance as they test out different forms of their developing professional identity. This is the building of what could be termed a 'micro-public' – a particular grouping of people online that are connected to a shared interest or practice (Marshall 2014, pp. 163–164).

2. *Mediatized*

The second dimension relates to the mediatized quality of online persona. A "mediatised persona operates under the modulation of commercial interest and corporately governed structures that individuals have become acclimatised to, many in the hope of sharing in the benefits of a widely proliferating self-image that was once only the province of individuals in film, television, print, and radio" (Moore et al. 2017, p. 3). Over time, mediatized identities of online persona are accumulated through digital fragments.

There is no question that individuals and the platforms used construct multiple identities via different uses of online social media and other online locations. These different personas are built from the array of related text, images and shared media that an individual surrounds themselves with in these online spaces. Indeed, students construct texts about themselves online that contribute to their identities well before they join university. These formations of persona may or may not match the way in which a particular profession or form of work uses media to express themselves individually or collectively. What can be encouraged is developing the skills to curate these non-professional personas and ensure their separation from the developing professional persona.

In order to foster the development of paralleling and linking with professional personas beyond their personal and non-professional personas, we might consider the forms of media and forms of address that are commonplace and accepted within particular disciplines, professions or domains of work that help legitimate a practitioner. These general protocols of mediatized behaviour then dictate the forms and

level of sophistication developed in relation to textual communication and information sharing. Certain professions privilege particular forms of display and exchange: students would have to facilitate themselves with this reconfiguration of their online identity for these different paths. In addition, sources of information within a profession and its online rearticulation are varied and need regular exploration and student-related interpretation (Marshall et al. 2019). A nascent form of assessment fostering portrayal includes an initiative titled 'me in a minute' this digitally aided video capture encourages students to pitch their knowledge, capabilities and experience to prospective employers (Jorre de St Jorre et al. 2017). The video as an artefact can then be embedded and broadcast by students in different forms of social media attracting comments from family, friends and potential employers.

3. *Performativity*

The third dimension relates to performativity where the public performance of the self should not be considered entirely 'real' nor entirely 'fictional', but rather a collection of multiple staged performances that require agency but may also become ritualised (Moore et al. 2017). The public performativity of a persona in our context takes into account previous, university and future worlds of work but oriented to the public world requiring interpretation of what you do and who you are for an authentic audience. Here we recognise that students are always in the middle, with previous and current identities that matter and future identities they are working towards. Assessment-as-portrayal might seek to leverage different identities for different purposes. For example, more students now work during their degree (Baik et al. 2017) so they will have personal, student and worker identities amongst others. These concurrent identities may be leveraged to promote engagement in the idea of persona and performativity, whilst testing out their future personas. Students need to learn how to blend their different identities for different audiences in appropriate ways whilst remaining authentic. This allows for a powerful transition for students from scholastic life to real-world scenarios. To begin with, the learning environment could be developed as a public simulation (e.g., an online portfolio or blog site that is password protected and only accessed within a unit of study); it could then be extended beyond the classroom cohort into the public world in the latter years of study through, for example, publicly accessible sites. Of course, students themselves will perhaps take these identities, their professional student personas, and launch them publicly in various forms at earlier stages. This prompts a question about how to manage public construction in sanctioned and controlled ways. Who can be trusted? How do we ensure safety for the students when engaging in public debates and discussions within their emerging professional networks? This requires supporting students to use the language, rhetoric and discourses of the profession to build credibility within it.

4. *Collective*

The fourth dimension of persona is one that "works to produce, seek out, and move between connections, resulting in a collective" (Moore et al. 2017, p. 5). The use of the term 'collective' avoids normative notions of community and realises

there are patterned relations within a particular and related group. It offers a valuable notion of how collective identity is a key feature of online culture.

By thinking about the collective in assessment, students might consider what types of collectives they might like to be involved with through their portfolio. The conversation here is around raising awareness about the norms and expectations of these particular collectives outside of the academy. These collectives play a powerful role in professional/disciplinary socialisation; however, students need to also question what is taken for granted in these communities in developing their own authentic selves (Vu and Dall'Alba 2014). Alternatively, we might engage different industry stakeholders around considering criteria for judging digital credentials in relation to networks and group-related exchanges built through their online public identity. Students might engage in curating hashtags in relation to their future disciplines or professions – these might belong to different communities. They might reflect on how they use different hashtags in communicating with these collectives and what aspects of self are advanced and why. Beginning the process of networking is a key element of employability where educational practices may be designed to begin this process in a safe and scaffolded way. It is related to micro-publics, as mentioned earlier, and their specific connection to professions. Students have to develop an expertise about what kinds of micro-publics are necessary for their future careers and what remain part of their personal and private experience.

5. *Value*

This dimension recognises that personas are created with a desire for constructing value and reputation: they are intentional and strategic in garnering some form of attention, but also respect within a particular collective. If our objective is to have students develop workable but also valuable identities within a work and professional culture, it would be useful to integrate into our earlier digital credential example ways in which students, academics and the related industry work together to decide on criteria by which their work on building their portrayal and persona identity is judged. A constant through this journey is the student, and in previous work we have argued that assessment should work towards developing students' evaluative judgements. In other words, students need to learn to make effective judgements about the quality work in order to improve their own (Tai et al. 2018). However, students' work needs to relate to accepted notions of what makes for 'good' or 'good enough', whether it's in the academic community or the workplace.

Disciplinary and professional notions of quality are not necessarily stable; they are dynamic and shifting. To further complicate this, when the public community is evoked, the range of views of quality that students might need to judge, integrate and calibrate become even more diverse. What kind of assessment of relevant impact can be used to make sense of the emerging work on a professional reputation – do we use the online ones of impressions/shares/followers – or do we look for the strategic reading of these kinds of reputation? In the case of the portfolio of social media activity (and related sites) these are updated regularly as to their value and influence for the personal direction of developing a useful and valued persona

that connects to their future career. A potential assessment task is where students are encouraged to follow a leading online identity related to the profession and discipline, mapping their impact across various criteria and numerical readings. Another example is where students chart their own impact through the circulation of particular posts and determine connections with the discipline. As students invest further in their specialisation and their potential future profession, they can develop thresholds of connections that are more discerning as they progress in their studies.

6.4 Tensions, Affordances and Future Development

6.4.1 Tensions

No conversation about assessment is complete without paying attention to the dominant discourse of grading, reliability and validity (Orr 2007). Adopting a poststructural perspective on identity that privileges strategic and situated construction is in opposition to commonly held psychometric discourses in assessment. Through persona studies we seek to recognise the dynamic and constructed nature of identities while preserving context. Certification, validation, peer review – online culture operates under different forms of validation. Summative assessment opens up even more risk and tension. What aspects of portrayal can and should be summatively judged, and under what circumstances? In certain professions, such as Medicine, there is already a call for assessing professional identity formation (Cruess et al. 2016). What do you give marks to in relation to identity? Attempts to measure identity seem misguided and slippery as they imply both a static possessed identity and one that can be reduced to a decontextualized artefact.

As a form of assessment for learning, this approach enables feedback conversations in relation to *becoming*, drawing on a wider audience base. For example, we may seek to explore the social networks that students subscribe to publicly and to provide feedback on relevant professional networks within social media (Marshall et al. 2015). When feedback comments are offered by staff as part of the online university community there is a level of trust inherent to and sanctioned within that community and within educational relationships. However, there are different structures of trust within the sharing economy (Sundararajan 2016). Therefore, how do we establish an effective relationship to a public world? And how are those relations monitored by teaching staff? These are complex but vitally important dimensions of activity that will need to be negotiated and developed in advancing at least a partially accessible public portrayal approach to student work and assessment in, ultimately, safe ways.

Moving student work and community into the public domain or wider community requires judgements about readiness and audience and the types of scaffolding needed. These go from satisfying the teacher, to portrayal for a wider range of teachers, to even wider domains beyond the course. Expectations related to

portrayal would need also to be adjusted and recalibrated as a student advances through their course of study. Connections to related work and professional networks are not necessarily stable and they transform over time: the approach to integrating this public portrayal work into student assessment will require innovation in course design and content and some related transformation in the role of the university in shepherding this student development.

Managing risk for students and the institution when students' work is public is important to consider. This requires developing clear policy directions for technology, including identifying different forms of media that may be leveraged. These policy statements may then give rise to conversations with students about the ethics of portrayal of personas.

6.4.2 Affordances

The current environment of assessment in higher education offers us particular affordances such as increasing technologies of portrayal. As mentioned, there is a shift that is already occurring where students do need to portray themselves to the world of work and want to distinguish themselves from others in their class. Authors such as Clarke and Boud (2016) position portfolios as curation of evidence for feedback becoming more sophisticated as students progress through their degree, affording more public portrayal of competencies and learning. We add a focus on portrayal of self through persona to this agenda; one which promotes learners' questioning of who they are and what they stand for.

The availability of technologies for portrayal enable the crowdsourcing of feedback towards publicly displayed and sanctioned credentials. An example of this is establishing networks with future employers and receiving feedback from them. Another example, is one from a digital media subject where students shared their work publicly as it was being created, rather than privately when completed, garnering comments from peers, community members and industry practitioners in order to improve their work (assessment for learning case study http://newmediaresearch.educ.monash.edu.au/feedback/case-studies-of-effective-feedback/case-study-4/). Students in this unit may reflect on the comments they receive and any resultant learning. In a further example, students are encouraged to contribute to Wikipedia entries as part of their assessment task thus developing the ability to analyse source material critically and to advance their research, writing and digital literacy (Cole 2009). A reflective wrapper may be added considering the self in relation to the community they seek to join and how the knowledge and skills they are learning can become significant and meaningful to them into the future. Assessment can therefore "increase students' awareness of who they are becoming, as well as about what is required of them to act in defining their stand" (Vu and Dall'Alba 2014, p. 789).

Another affordance of the digital environment is that at the same time as students produce their identities, they can lurk to observe others' constructions and how they do it. They can see exemplars of the private/public construction of the same person

on different media over time. Students may be encouraged to look closely at a number of profiles within the discipline. As they try on different personas they come to understand their relationship to the profession in a strategic way, including perhaps that this may not be an authentic pathway for them to take in life.

6.4.3 Future Development

Throughout this chapter we have attempted to sketch some possibilities for assessment-as-portrayal through the lens of persona studies. Future research should seek to explore where and how such opportunities already appear within curricula, seeking to unpick the affordances of such portrayal at individual and structural levels. We might study how students relate their online identity to their student identity and their developing professional/disciplinary identity as they progress through their degree. Further, we may ask how students conceptualise their online personas relative to their future profession and how these shift, develop and change as students progress through their education. Practical work is needed to understand how criteria and quality descriptors might be sufficiently loose to allow for creation and creativity, enabling diversity rather than conformity of presentation.

6.5 Conclusion

In this chapter, we have sought to reimagine assessment through the lens of persona studies. This is a much larger endeavour than the space a chapter affords. Yet we hope to have started a conversation, and many more need to occur to build a diversification of purposes into assessment practices in higher education that incorporate becoming rather than acquisition or participation (Hager and Hodkinson 2009). This requires a reconfiguration of the university around different ways of knowing. It seems to us this shift in assessment practices to incorporate more contextual portrayals of achievements is already occurring. To capitalise on this shift, we propose a repositioning of these assessment practices such that portrayal is considered to be one of private and public selves strategically negotiated for particular audiences. It would be useful to see the quality of what could be deemed a professional persona emerging through the process of university education and assessment. What this means is developing ways of fostering and evaluating the building of a coherent public identity that is seen as identifiable and externally valuable to a profession/area of work and activity. Such construction requires student agency, judgements about what is appropriate in the social world they wish to join where they 'try on' different personas, and building of social networks prior to graduation. This lens enables us to engage students' multiple identities in the educational process whilst creating broader networks, therefore, building better bridges between the relatively safe higher education communities and the workplace.

References

Baik, C., Naylor, R., Arkoudis, S., & Dabrowski, A. (2017). Examining the experiences of first-year students with low tertiary admission scores in Australian universities. *Studies in Higher Education, 44*(1), 1–13. https://doi.org/10.1080/03075079.2017.1383376.

Boud, D., & Ajjawi, R. (2019). The place of student assessment in pursuing employability. In J. Higgs, W. Letts, & G. Crisp (Eds.), *Education for employability (volume 2): Learning for future possibilities* (pp. 167–178). Rotterdam: Brill Sense.

Clarke, J. L., & Boud, D. (2016). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International, 55*(4), 479–486. https://doi.org/10.1080/14703297.2016.1250664.

Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers & Education, 52*(1), 141–146. https://doi.org/10.1016/j.compedu.2008.07.003.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology, 66*(10), 2003–2019. https://doi.org/10.1002/asi.23309.

Cruess, R. L., Cruess, S. R., & Steinert, Y. (2016). Amending Miller's pyramid to include professional identity formation. *Academic Medicine, 91*(2), 180–185. https://doi.org/10.1097/ACM.0000000000000913.

Hager, P., & Hodkinson, P. (2009). Moving beyond the metaphor of transfer of learning. *British Educational Research Journal, 35*(4), 619–638. https://doi.org/10.1080/01411920802642371.

Holt, D., McGuigan, N., Kavanagh, M., Leitch, S., Ngo, L., Salzman, S., Watty, K., & McKay, J. (2016). Academic leaders' perspectives on adopting ePortfolios for developing and assessing professional capabilities in Australian business education. *Australasian Journal of Educational Technology, 32*(5), 1–18. https://doi.org/10.14742/ajet.2666.

Jorre de St Jorre, T., & Oliver, B. (2018). Want students to engage? Contextualise graduate learning outcomes and assess for employability. *Higher Education Research and Development, 37*(1), 44–57. https://doi.org/10.1080/07294360.2017.1339183.

Jorre de St Jorre, T., Johnson, L., & O'Dea, G. (2017). *Me in a minute: A simple strategy for developing and showcasing personal employability.* Paper presented at the Me, Us, IT! ASCILITE2017: 34th international conference on innovation, practice and research in the use of educational technologies in tertiary education, Toowoomba, QLD. Retrieved from http://2017conference.ascilite.org/wp-content/uploads/2017/11/Concise-JORRE-de_ST-Jorre.pdf.

MacDonald, J., Sohn, S., & Ellis, P. (2010). Privacy, professionalism and Facebook: A dilemma for young doctors. *Medical Education, 44*(8), 805–813. https://doi.org/10.1111/j.1365-2923.2010.03720.x.

Marshall, P. D. (2014). Persona studies: Mapping the proliferation of the public self. *Journalism, 15*(2), 153–170.

Marshall, P. D. (2016). *The celebrity persona pandemic*. Minneapolis: University of Minnesota Press.

Marshall, P. D., & Henderson, N. (2016). Political persona 2016 – An introduction. *Persona Studies, 2*(2), 1–18. https://doi.org/10.21153/ps2016vol2no2art628.

Marshall, P. D., Moore, C., & Barbour, K. (2015). Persona as method: Exploring celebrity and the public self through persona studies. *Celebrity Studies, 6*(3), 288–305. https://doi.org/10.1080/19392397.2015.1062649.

Marshall, P. D., Moore, C. W., & Barbour, K. (2019). *Persona studies: An introduction*. Malden: Blackwell-Wiley.

Miller, K. K., Jorre de St Jorre, T., West, J. M., & Johnson, E. D. (2017). The potential of digital credentials to engage students with capabilities of importance to scholars and citizens. *Active Learning in Higher Education, 2017*, 1–12. https://doi.org/10.1177/1469787417742021.

Moore, C., Barbour, K., & Lee, K. (2017). Five dimensions of online persona. *Persona Studies, 3*(1), 1–11. https://doi.org/10.21153/ps2017vol3no1art658.

Oliver, B. (2015). Redefining graduate employability and work-integrated learning: Proposals for effective higher education in disrupted economies. *Journal of Teaching and Learning for Graduate Employability, 6*(1), 56–65. https://doi.org/10.21153/jtlge2015vol6no1art573.

Orr, S. (2007). Assessment moderation: Constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education, 32*(6), 645–656. https://doi.org/10.1080/02602930601117068.

Sullivan, S. E., & Baruch, Y. (2009). Advances in career theory and research: A critical review and agenda for future exploration. *Journal of Management, 35*(6), 1542–1571. https://doi.org/10.1177/0149206309350082.

Sundararajan, A. (2016). *The sharing economy: The end of employment and the rise of crowd-based capitalism*. Cambridge, MA: MIT press.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Vu, T. T., & Dall'Alba, G. (2014). Authentic assessment for student learning: An ontological conceptualisation. *Educational Philosophy and Theory, 46*(7), 778–791. https://doi.org/10.1080/00131857.2013.795110.

Rola Ajjawi is Associate Professor in Educational Research at the Centre for Research in Assessment and Digital Learning, Deakin University where she leads a program of research into learning and feedback in workplace settings. She is Deputy Editor of the journal *Medical Education*.

David Boud is Professor and Foundation Director of the Centre for Research in Assessment and Digital Learning, Deakin University, Professor of Work and Learning at Middlesex University and Emeritus Professor at the University of Technology Sydney. He has been a pioneer in developing learning-centred approaches to assessment across the disciplines, particularly in student self-assessment, building assessment skills for long-term learning and new approaches to feedback.

David Marshall is Research Professor and holds a Personal Chair in New Media, Communication and Cultural Studies at Deakin University. Professor Marshall has published widely in two areas: the public personality/celebrity and new media culture.

Chapter 7
Towards Technology Enhanced Dialogic Feedback

Edd Pitt and Naomi Winstone

> *The essence of feedback processes with technology should be …dialogic, focused on student engagement with feedback, and with the aim of facilitating student self-monitoring of their own progress.*
> Carless (2015, p. 199)

Abstract In this chapter, we view technology-enhanced feedback (specifically audio, video and screencast feedback) through the lens of Carless's (High Educ 69(6):963–976, 2015) 'old' versus 'new' paradigm postulates. We explore the potential factors that may facilitate audio, video and screencast feedback promoting dialogue in a new paradigm approach to feedback. We identify three 'dilemmas' that may need to be addressed to facilitate such dialogue. In exploring this territory, we draw upon examples from the research literature and how this may inform our developing understanding of feedback dialogue within the digital world.

7.1 Framing Feedback in Higher Education

Prior to the massification of higher education, dialogue was central to the feedback process (Nicol 2010). Written comments on students' drafts would often be accompanied by synchronous discussion in tutorials, following which students could enact feedback information prior to submitting the assignment for grading. With the

E. Pitt (✉)
Centre for the Study of Higher Education, University of Kent, Canterbury, UK
e-mail: E.Pitt@kent.ac.uk

N. Winstone
Department of Higher Education, University of Surrey, Guildford, UK
e-mail: n.winstone@surrey.ac.uk

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_7

increase in cohort sizes and modularisation of units of study, came a shift away from this approach, towards one where "written feedback, which is essentially a one-way communication, often has to carry almost all the burden of teacher–student interaction." (Nicol 2010, p. 501).

Nicol's paper served as a 'call to arms', following which many prominent scholars in the area of assessment and feedback have promoted a shift in practice, away from a focus on the transmission of comments, towards a focus on student engagement with feedback, and the impact of feedback on students' learning. David Carless terms the transmission-focused approach as being the 'old paradigm' of feedback; in contrast, the more dialogic approach represents what can be seen as a 'new paradigm' (Carless 2015). The importance of the new paradigm lies in the recognition that the monologic transmission of feedback comments does not on its own facilitate learning (Boud and Molloy 2013); instead, "it is what the students can do with feedback rather than how the teacher provides it which is crucial" (Carless 2015, p. 28), because "information becomes feedback only when it is used productively" (Carless 2015, p. 192). This emphasis on student engagement and action represents 'student uptake' of feedback as conceptualised by Carless and Boud (2018).

Central to the new paradigm is the importance of ongoing dialogue in the feedback process. Dialogue in this sense gives primacy to the role of the *student* in the process, whereby they are encouraged to seek and discuss information (from multiple sources) that enables them to develop their understanding of what constitutes quality, and empowers them to take action on feedback. In this regard, Carless argues that:

> Dialogic feedback involves iterative processes in which interpretations are shared, meanings negotiated, and expectations clarified in order to promote student uptake of feedback. (Carless 2015, p. 196)

In this chapter, we view technology-enhanced feedback (specifically audio, video and screencast feedback) through the lens of Carless's (2015) 'old' versus 'new' paradigm postulates. In particular, we explore how the technology-enhanced feedback literature has conceived of dialogue between lecturers and students. Central to our discussion is an exploration of the discourse and conclusions drawn by researchers in this field and how this may inform our developing understanding of feedback dialogue within the digital world.

7.2 The Role of Technology in Feedback

A search in the Scopus database for publications on the use of technology in assessment and feedback in higher education shows that the number of publications per year increased by 700% between the years 2000 and 2017. The proliferation in the number of publications on this topic represents a growing interest in the potential for technology to enhance assessment and feedback processes, alongside increasing

availability of freely available technological tools that can be used to assess and provide feedback to learners (Yuan and Kim 2015).

The use of technology has been suggested for the provision of formative feedback (Hennessey and Forrester 2014), and the facilitation of peer-to-peer feedback (Van der Pol et al. 2008). Learning analytics have also been used to facilitate the generation of individualised feedback for learners (Pardo et al. 2017). Hepplestone et al. (2011, p. 123), reported that "a growing number of studies support the hypothesis that technology has the potential to enhance student engagement with feedback". Similarly, in their systematic review of the literature on students' engagement with feedback, Winstone et al. (2017a) found that the primary purpose of technological feedback interventions was to enable students to become more motivated to engage with feedback. Further, it is claimed that the use of technology can facilitate personalisation of feedback and foster relationships between student and teacher (Yang and Carless 2013), which seems crucial if we consider that feedback might not be most effective if provided on an anonymous basis (Pitt and Winstone 2018).

The most common uses of technology involve digital delivery of feedback information, using audio, video, and screencast technology. Uploading audio files of comments on students' work is thought to be beneficial because more detailed comments can be provided than might be possible through the more traditional written medium (Merry and Orsmond 2008). Furthermore, students may perceive audio feedback as more personalised than written feedback (Gould and Day 2013), find it easier to comprehend (Merry and Orsmond 2008), and consider it to be more supportive in tone than written feedback (Ice et al. 2007). Perhaps because of these affordances, student engagement with audio feedback is often described as superior to engagement with written feedback; for example, 78% of students in Mayhew's (2017) study reported that they would be more likely to take action or revisit audio feedback in comparison to written feedback.

Students often interpret audio feedback as a form of dialogue (Nicol 2010), perhaps because non-verbal cues such as prosody, emphasis, and tone can all be communicated through audio feedback in ways that are simply not possible with written feedback (Mahoney et al. 2018). Indeed, Mayhew (2017) argues that social interactions are not always afforded in written feedback. Video feedback however, affords greater individualisation and personalisation than written feedback (Henderson and Phillips 2015), and screencast feedback (where the markers verbal comments are accompanied by an annotated visual display of the student's work) has the further benefit of markers being able to pinpoint the locus of their comments (Mayhew 2017), and demonstrate how to correct errors (Stannard 2007). As discussed later in this chapter, factors such as these can lead students to feel more motivated to improve their work (Henderson and Phillips 2015).

So how might these technologies afford dialogic approaches to feedback? In this chapter, we explore the potential factors that may facilitate audio, video and screencast feedback (hereafter referred to collectively as 'technology-enhanced' feedback where appropriate) promoting dialogue in a new paradigm approach to feedback. We identify three 'dilemmas' (see Fig. 7.1) that may need to be addressed to facilitate such dialogue. We recognise that technology-enhanced feedback could be used

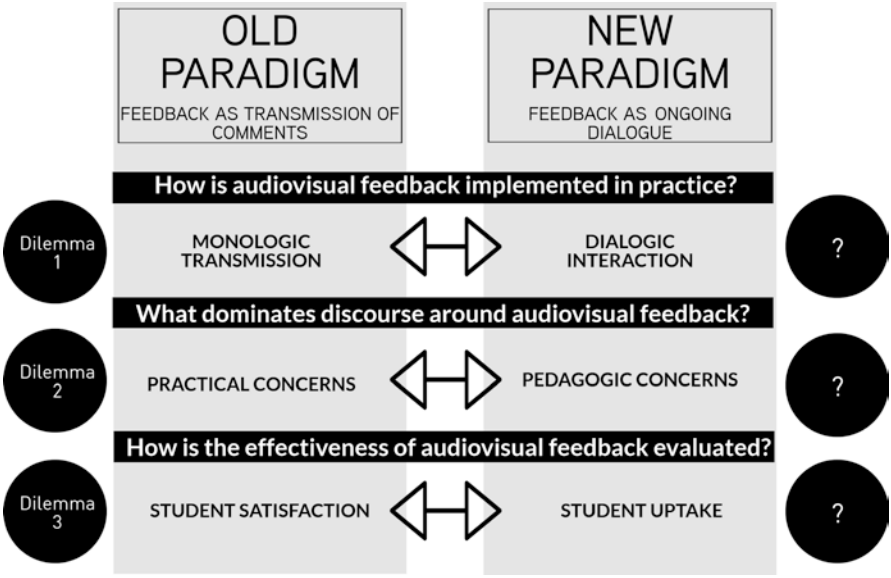


Fig. 7.1 Positioning technology-enhanced feedback within old and new paradigm approaches to feedback

in ways that merely replicate old paradigm principles within a digital environment; however, this is not the intent of the next section.

In exploring this territory, we draw upon examples from the research literature and qualitative research data. We did not undertake a systematic literature review; rather, we utilised a snowballing technique involving backward and forward citation chasing from a number of highly cited papers on technology-enhanced feedback. The qualitative data consisted of a series of semi-structured interviews with 28 academic staff (17 female) from universities across the UK that we thematically analysed. The participants represented a broad range of disciplines (Social Sciences n = 7; STEM n = 12; Health/Medicine n = 4; Arts/Humanities n = 5). Participants were recruited via opportunity sampling, and interviews lasted 30 minutes on average. Within these interviews, participants were encouraged to reflect upon their current feedback practices, and to discuss their perceptions of potentially using technology-enhanced feedback methods in the future. The purpose of these questions was to surface general perceptions about these methods across a broad spectrum of academics, such that our exploration of the use of technology-enhanced feedback, and recommendations for dialogic use of these methods, might be informed by perceptions 'on the ground'. In the section that follows, we explore each of these dilemmas in turn.

7.3 Dilemma 1: Monologic Transmission or Dialogic Interaction?

Although technology-enhanced feedback methods give precedence to the spoken rather than written word, this does not automatically make them dialogic. Whilst this may appear obvious, we begin with this point of clarification in recognition of that fact that "a major challenge for the development of effective feedback processes is the predominance of a view of feedback as the transmission of information, often in the form of a monologue" (Carless 2015, p. 192). In order for technology-enhanced feedback to facilitate dialogic interaction, the practice needs to be used in ways that move beyond the transmission of feedback comments, towards student uptake of feedback. Many discussions of the role of technology in feedback espouse that technology has facilitated advances in practice (Marriott and Teoh 2012), even initiating a paradigm shift in the giving of feedback information (Fish and Lumadue 2010). Our initial exploration of the dialogic use of technology in feedback practice led us to consider the models of feedback being communicated via the use of language. If such practices are being adopted to facilitate dialogic interactions, then we might expect to see these practices being framed in 'new paradigm' terms; that is, in ways that emphasise the engagement and active contribution of students.

Evident was an emphasis on 'conversation' or the 'conversational nature' of technology-enhanced feedback (Cranny 2016). Whilst 'conversation' can be seen as a synonym for 'dialogue', it is important to question whether this reflects dialogic approaches to feedback according to the definition we have adopted. Central to the argument promoting technology-enhanced feedback to increase conversation between lecturers and students is that it facilitates a stronger lecturer and student relationship. All students, but particularly those who consistently find study difficult, may benefit from strong relationships with their educators (Pitt et al. 2019). Further, relational elements of the feedback process increase personalisation, emphasising pertinent points through nuances in tone of voice, and by providing a more informal and engaging delivery (Carruthers et al. 2015).

Students often report that technology-enhanced feedback 'feels' like a real conversation, as emphasised by a student in Lunt and Curran's (2010) study on screencast feedback: "[The feedback] was a personal address to me and my coursework, quite like sitting in John's office and getting him to explain what I need to do" (p. 764). This narrative reflects what has been described as the 'enhanced presence' of the lecturer (Hennessy and Forrester 2014, p. 784). As far back as 2007, Stannard argued that screencast feedback synthesised the processes of returning work with detailed annotations, and actually meeting the student to discuss their work. Stannard argued the screencast feedback process felt more 'human', but was the interaction between lecturer and student dialogic or monologic? In Lunt and Curran's (2010) study, the student received a more personalised form of feedback in terms of tone and contextual content, but in an asynchronous, non-discursive manner.

Within the literature that we reviewed, there was some evidence of a focus on dialogic framing of technology-enhanced feedback, where it was recognised that such practices can enhance students' engagement with and uptake of feedback. For example, Crook et al. (2012) suggested that:

> An appropriate technological application has the potential to encourage staff to reflect on their current feedback practices so that they can provide more detailed, comprehensible and engaging feedback. Technologies may also provide the innovative edge that can help students engage more effectively with their feedback. (p. 2)

The nature of engagement as presented here requires careful scrutiny. In some cases, students are described as showing stronger engagement with technology-enhanced feedback simply because of its novelty (Crook et al. 2012). Enhanced motivation to engage with the contents of feedback may be a useful end in and of itself, but may not necessarily lead to uptake of the feedback. Nevertheless, given that a commonly-reported barrier to students' use of feedback is the difficulty they can experience 'decoding' the language contained within feedback comments (Pitt and Norton 2016; Winstone et al. 2017b), the framing of technology-enhanced feedback as being easier to understand could be seen to acknowledge the importance of student engagement. For example, both audio and video feedback are framed as being easier for students to understand, often because spoken language is typically more informal in tone than written feedback (West and Turner 2016). However, without evidence of the uptake of the feedback, a focus on the language used within feedback is more closely aligned with the old rather than the new paradigm.

Despite some evidence of the alignment of technology-enhanced feedback practices with a new paradigm model in the literature that we reviewed, an old paradigm approach still dominates. For example, many studies on the use of technology-enhanced feedback use transmission-focused terminology, such as: *transmitted* (Mathisen 2012), *provide/provision* (Crook et al. 2012; Mathieson 2012; Mathisen 2012; West and Turner 2016), *receive* (Mathieson 2012; Mathisen 2012), *give* (Mathisen 2012), *delivered* (Crook et al. 2012), *sent to* (McCarthy 2015) and *convey* (West and Turner 2016). The use of such terminology does not necessarily imply that the authors' focus is purely transmission-based; rather it reflects the way in which feedback is more commonly framed, not necessarily the way in which it is practiced. However, we have to be aware of positioning students as passive receivers of feedback through the use of transmission-focused language (Winstone and Pitt 2017). In our interviews, several lecturers recognised that if technology-enhanced approaches to feedback simply replicate the transmission of written feedback without taking advantage of opportunities for dialogic interaction, then there is limited advancement of practice:

> I don't, um, see a discussion. I mean that's very one way isn't it?… I think it's very one way…If I was to audio record something and then send it to a student, then …they don't have any way of replying do they? (Participant 9)

> I don't know how that's gonna improve on the written feedback …Because it's just transmission of information again isn't it? Um, and either way – auditory or through a video – you're just transmitting information … you're not promoting a dialogue, which is what I think we need. (Participant 20)

Further evidence of a transmission-focused paradigm through digital media could be seen through ways in which our participants spoke about the potential enactment of technology-enhanced feedback in their practice. Some lecturers suggested they would produce written feedback, convert this into an audio recording, or write a script for video feedback; essentially, verbalising written comments but increasing the workload associated marking itself:

> Audio feedback, it is a minor hassle but then it would actually take about two minutes per student to transfer their written feedback into an audio one. (Participant 1)
>
> There's an element of maybe rehearsal, cos again you can't edit what you have said. You'd have to maybe write a script first and then give the feedback to be sure that you're accurate in what you wanted to impart to the student. (Participant 24)

These quotations illustrate that common perceptions of technology-enhanced feedback are more closely aligned with an old paradigm, than a new more dialogic paradigm. This has been recognised within the literature; for example, Mathieson (2012) explains that, in her study, video feedback did not provide an opportunity for students to have a meaningful dialogic interaction with the lecturer to clarify points raised or ask further questions. Such assertions were similarly expressed by one student participant:

> Isn't 'interaction' supposed to be reciprocal?… there was really no interaction going on. (Mathieson 2012, p. 149)

Some researchers conclude their arguments by suggesting that we need to explore how more interactive dialogue can be facilitated using technology (West and Turner 2016).

7.4 Dilemma 2: Practical or Pedagogic Concerns?

Discussion of the pragmatic elements of technology-enhanced feedback practices are a ubiquitous feature of the literature (cost, time-efficiency). As portrayed in Fig. 7.1, the use of technology-enhanced feedback in the old paradigm leads to discourse surrounding practical elements of the transmission process; in contrast, new paradigm discourse surrounding the pedagogic elements of the process might predominate. Many authors identify practical advantages of technology-enhanced feedback practices, being low-cost and time-efficient means by which feedback can be provided to students (Ice et al. 2007). Lunt and Curran (2010) argue that audio feedback is more time-effective for the lecturer, as it is possible to speak in 1 min that which would take 6 min to write. The efficiency of technology-enhanced feedback in terms of facilitating more detailed comments was discussed in several papers (Mathieson 2012; Mathisen 2012; Turner and West 2013).

Counter to the time-saving argument, Morris and Chikwa (2016) suggest that by the time the lecturer has recorded audio feedback, saved it locally, uploaded it to the VLE and notified students of its return, it will have taken longer in total than the process of writing comments. Similarly, King et al. (2008) claim to have seen no evidence that technology-enhanced feedback saves time. Nevertheless, the potential impact on the workload associated with providing feedback to large cohorts is a commonly discussed issue in the literature. Fawcett and Oldfield (2016) argued that "audio feedback may be an effective way of managing high marking loads and the student need for timely feedback" (p. 81). When faced with increasing demands from other areas of their roles, it is easy to see why arguments relating to time-efficiency and workload reduction might appeal to lecturers and mangers alike.

Practical issues were also dominant in discussion of concerns and challenges regarding technology-enhanced feedback (Fawcett and Oldfield 2016). According to Marriott and Teoh (2012), screencast feedback is typically 2–3 min in length, resulting in upwards of 60 MB of data. As such, delivery of the file to students, and storage on University servers, were concerns. The need for a quiet space in which to record technology-enhanced feedback was also identified as a challenge (Hennessy and Forrester 2014). Arguably, focusing on such pragmatic issues risks trivialising the learning-focused element of the feedback process, directing attention towards reducing file size through limiting the length or using compression software.

This focus on the pragmatic elements was also evident in our interview data. Many respondents explained that they were often reticent to engage with technological advances because they perceived them to be overly complicated, whilst offering limited time-saving benefits:

> You have to save an audio file and then upload the audio file, so rather than making things simpler; it makes them more complicated (Participant 8)
>
> I feel a lot more positive about audio feedback, but I still don't do it because to me it feels that that's going to take so much more time to do. (Participant 27)

Similarly, respondents expressed scepticism about the use of video feedback, perceiving it would require a lot of effort to produce, even citing logistical concerns about their appearance on video:

> For the effort it would entail, I can't see that I'd want to do it. (Participant 15)
>
> I tend to do my marking at home when I don't have to get ready to go out. I feel like I'd have to get ready for work to then be presented on video. (Participant 2)

Whilst practical issues are important in affording effective use, a more fundamental concern should be the impact on learning and student uptake of feedback. Within the literature that we reviewed, we saw evidence of students adopting a transmission-focused mind-set in response to technology-enhanced feedback, as illustrated in this example from a student in Marriott and Teoh's (2012) focus groups following the release of screencast feedback:

> To me to get the feedback from my teacher like face-to-face, I'm going to have to make an appointment and I go and see him and he doesn't have time, he's just going to have to rush through it sort of thing, but now I have it all the time and can watch it whenever I want, you know that's much easier for me, I prefer doing it that way to be honest. (p. 594)

This student interprets screencast feedback as facilitating engagement, and removing the need for them to go and see their lecturer (where further dialogue potentially could take place). The student also identifies that a key benefit of screencast feedback is that they can revisit it whenever they wish. For example, Crook et al. (2012) and Cranny (2016) argue that digital feedback is permanent; it can be paused, rewound and played multiple times. However, students are afforded the same agency when receiving written feedback. We would argue that moving towards dialogic use of technology-enhanced feedback requires a stronger focus on the students' volition to engage with and utilise the feedback information in subsequent work and not the implied convenience of where it is stored or accessed.

7.5 Dilemma 3: Satisfaction or Student Uptake?

We argue that the adoption of any learning technology should be driven by a sound rationale, but what should the focus of this rationale be? The literature appears to contain many instances whereby the efficacy of technology-enhanced feedback is related to students liking or preferring it (Crook et al. 2012; Marriott and Teoh 2012; West and Turner 2016).

One explanation for students' preference for technology-enhanced feedback is relational; students often report that face-to-face dialogue with a lecturer can be uncomfortable and troublesome (Henderson and Phillips 2015). This is often believed to be the case because meeting in person with a lecturer can threaten a student's self-esteem, as the power imbalance in that setting makes students particularly aware of the limitations of their own understanding (Carless 2006; Sambell 2013). In this sense, then, it follows that it is "more congenial and less ominous to embrace oral formative feedback" in a technology-enhanced feedback environment (Hennessy and Forrester 2014, p. 783). This is particularly true for students new to higher education, who "may feel uncomfortable, shy, and/or insecure in approaching a tutor for help…we have found that audio feedback goes a long way towards resolving these kinds of issues" (Hennessy and Forrester 2014, p. 783).

We would argue that technology-enhanced feedback only 'resolves' these relational issues if students feel more comfortable approaching lecturers for further dialogue following technology-enhanced feedback, in comparison to written feedback. There is some evidence that audio feedback makes lecturers appear more approachable (Jackson 2012), and that lecturers appear to be more supportive through the medium of video feedback (Henderson and Phillips 2015). Both Orsmond et al. (2013) and Fawcett and Oldfield (2016) recognise the potential for technology-enhanced feedback to serve as a starting point for further one-to-one dialogue. It is possible that these reasons, alongside the perception of greater detail in technology-enhanced feedback, go some way to explaining students' preference.

Whilst important, student satisfaction should not be the primary motive for practitioners to adopt technology-enhanced feedback; rather, emphasis should be placed upon the effect that feedback medium has upon student uptake of the feedback.

Whilst breaking down relational barriers between markers and students could in principle facilitate greater uptake of feedback, the primary focus is not placed on what the student does with the feedback next, nor dialogic interaction with the student. In our interviews, the lecturers were clear in demonstrating scepticism surrounding the learning benefits of technology-enhanced feedback. They wanted evidence of its effectiveness upon student learning in comparison to written feedback:

> I'm not sure what the benefits might be for students on the whole, if there's any evidence to show if that's more beneficial than a written piece. (Participant 5)
>
> Is there evidence that makes students listen to it or read it or take it in or act on it? I'd love to see that before we bring in yet another method of giving feedback that students may or may not engage with really… Whether we think that by putting it as a podcast or videos or you know, does that make any difference? I'd love to know; cos my personal view is […] probably not. (Participant 10)

Our exploration of the literature did lead us to identify some evidence of the impact on student uptake of feedback, albeit largely based on self-report measures. At the basic level of engagement, Lunt and Curran (2010) reported that students were ten times more likely to open a file containing audio feedback, than to collect hard-copy written feedback. In Merry and Orsmond's (2008) small scale study (n = 15), students found the audio feedback easier to understand and implement; furthermore, they revisited the feedback, and sought further clarification from lecturers, indicating a stimulation to engage in face-to-face dialogue.

Students often report that they would be more likely to take action based on technology-enhanced feedback. For example, 69% of students in the study reported by Carruthers et al. (2015) agreed that they would later revisit the audio feedback when preparing their next assignment. After experiencing video feedback, 95% of Turner and West's (2013) participants reported a preference for using this form of feedback in future assessments, rather than written comments. Whilst these examples might on the surface align with a new paradigm focus on student uptake of feedback, no data were collected to confirm whether students' perceptions translated into actual behaviour. Hennessy and Forrester (2014, p. 784) reported that following formative audio feedback, final year students discussed a positive impact on their behaviour: "'Hearing' the spoken words of their tutor often prompted [students] to read more widely in order to address deficiencies in their work or to consider certain areas of their work more critically, as recommended". The impact of these actions on students' learning is not evident, and in one study, students reported that the impact of technology-enhanced feedback on their learning had been 'moderate' (Gould and Day 2013).

Whilst student satisfaction is likely to be important in facilitating engagement with feedback, emphasis should be placed on evidence of the efficacy of the approach in facilitating uptake of feedback. We found minimal evidence of the behavioural impact of technology-enhanced feedback; instead, self-reported intentions to engage with feedback formed the dominant source of evidence.

7.6 Facilitating Dialogic Use of Digital Tools in Feedback Practice

We began this chapter by emphasising the importance of dialogue in the feedback process, through the conceptual lens of a 'new paradigm' approach to feedback (Carless 2015), where student uptake of feedback, and the impact of feedback on learning, are of primary focus. We proposed three 'dilemmas', representing differential framing of technology-enhanced feedback within an 'old paradigm' and 'new paradigm' model of feedback. We drew upon our exploration of the literature on technology-enhanced feedback practice, alongside data from interviews with 28 UK academics. We end by proposing some recommendations for facilitating dialogic use of technology-enhanced feedback.

7.6.1 A Design Stance on Technology-Enhanced Feedback

If technology-enhanced feedback is to operate within a new paradigm model, then the design of the module or programme needs to create opportunities for the students to use feedback in subsequent assessments. Whilst it is clear that more detailed recommendations for development can be provided through technology-enhanced feedback (West and Turner 2016), the value is limited if students do not have the opportunity to implement these in subsequent assessments.

In designing assessment processes where students have the opportunity to enact technology-enhanced feedback, Espasa et al. (2018) present a useful tool which can be used to evaluate the extent to which a particular design is dialogic in nature. For example, Espasa et al. (2018) suggest that highly personalised feedback is likely to facilitate greater dialogue, as is feedback that includes suggestions, comments and explanations. Whilst technology-enhanced feedback can be used to provide such feedback, perhaps timing is more important in terms of creating a dialogic feedback environment; formative opportunities, where feedback occurs prior to summative assessment, may be more likely to encourage student uptake (Boud and Molloy 2013). There is greater dialogic potential in assessment designs where feedback is provided prior to submission, and where the student has the opportunity to resubmit work following enactment of such feedback (Espasa et al. 2018). There is potential for technology-enhanced feedback to facilitate this process; for example, students could submit drafts of an assignment, for which lecturers or even peers could provide brief audio or screencast feedback. Feedback is likely to be of greater use at this point than after the submission of an assignment at the end of the module (Carless 2015). Thus, rather than adding to lecturers' workload, this approach might represent 'redeployment' of feedback resource to a more impactful stage in the process. It is perhaps worth reflecting on the example from Hennessey and Forrester's (2014) study discussed under dilemma three; when receiving formative audio feedback, students "read more widely in order to address deficiencies in their

work or to consider certain areas of their work more critically" (p. 784). This example of uptake of feedback is firmly in line with a new paradigm approach. Furthermore, if students discuss their formative feedback with lecturers or peers, the dialogic potential is further enhanced.

7.6.2 *Seeking the 'Value-Added' of Technology-Enhanced Feedback*

Many of the participants in our interview study noted that technology-enhanced feedback could be implemented in a way that merely replicates the transmission of comments. To facilitate dialogue requires us to consider the unique potential afforded by audio visual feedback. Many of the proposed advantages (more detailed comments; opportunity for students to revisit comments at a later date) would also apply to traditional handwritten feedback, or electronic annotation feedback. Hennessy and Forrester (2014) clearly demonstrate an approach to seeking the 'value-added' potential of technology-enhanced feedback:

> More detailed feedback on a written piece of work could, arguably, be delivered through track changes or other functions of electronic feedback. However, what enhances the value of audio feedback for students is the level of appreciation they experienced by being 'spoken to' (Hennessy and Forrester 2014, p. 782)

Thus, the pedagogic potential should be an important part of the decision to adopt technology-enhanced feedback. If it is used in a way that replicates the transmission of written comments, just through a different medium, then the rationale for its use appears to be lacking. Our exploration of the literature did reveal areas where technology-enhanced feedback can provide this 'value-added'; for example, through screencast technology, students can see tutors actively demonstrating how to correct one example of an error in their work (in situ), leaving them to apply this feedback to the remainder of their work. This may well enhance students' agency to successfully implement feedback (Pitt 2017; Pitt et al. 2019; Winstone et al. 2017b). Similarly, where lecturers can appear to be more supportive and approachable through technology-enhanced feedback, this could break down relational barriers to student engagement in one-to-one dialogues with lecturers.

7.6.3 *Technology-Enhanced Feedback as a Stimulus, Not a Replacement, for Face-to-Face Dialogue*

Discussion of the literature and our own data showed that both lecturers and students recognise the fact that the 'dialogue' within technology-enhanced feedback environments is often asynchronous. It is important that technology-enhanced feedback is not seen as a replacement for face-to-face dialogue, rather as a stimulus for

dialogic interactions. For example, Carless (2015) proposes that dialogue can be facilitated by a lecturer asking questions to students through their feedback.

There is clear potential for technology-enhanced feedback to remove some of the barriers that might inhibit face-to-face dialogue between students and lecturers, leading to uptake of feedback. However, we would argue that in order for dialogic technology-enhanced feedback to become mainstream practice in higher education, we need to build a stronger evidence base not just of students' perceptions of the medium and of the efficiency and practicalities of this form of feedback. Instead, as clearly identified by the lecturers in our interviews, we need evidence of the impact of technology-enhanced feedback on students' behaviour and their learning outcomes. This evidence is much more challenging to obtain than that from self-report measures, which perhaps goes some way to explaining why "empirical research of actual feedback dialogue and its effects is limited" (Ajjawi and Boud 2018, p. 2). In their systematic review, Winstone et al. (2017a) also noted the predominance of self-report measures of behaviour and engagement with feedback, rather than behavioural outcome measures. Citing Macgregor et al. (2011), Hennessy and Forrester (2014, p. 778) state that "researchers rarely attempt to understand more comprehensively audio feedback efficacy or measure resultant student learning". The impact of technology-enhanced feedback on students' learning is central to a new paradigm focus, and stronger evidence of the behavioural and cognitive effects will enhance our understanding of how technology-enhanced feedback can 'add value' to dialogic feedback environments.

Acknowledgements: The interviews reported in this chapter were supported by a grant to the second author from the Society for Research into Higher Education (RA1648).

References

Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment & Evaluation in Higher Education, 43*(7), 1106–1119. https://doi.org/10.1080/02602938.2018.1434128.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233. https://doi.org/10.1080/03075070600572132.

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education, 69*(6), 963–976.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.

Carruthers, C., McCarron, B., Bolan, P., Devine, A., McMahon-Beattie, U., & Burns, A. (2015). I like the sound of that – An evaluation of providing audio feedback via the virtual learning environment for summative assessment. *Assessment & Evaluation in Higher Education, 40*(3), 352–370.

Cranny, D. (2016). Screencasting, a tool to facilitate engagement with formative feedback? *(AISHE-J) The All Ireland Journal of Teaching and Learning in Higher Education, 8*(3), 2911–2927.

Crook, A., Mauchline, A., Maw, S., Lawson, C., Drinkwater, R., Lundqvist, K., et al. (2012). The use of video technology for providing feedback to students: Can it enhance the feedback experience for staff and students? *Computers & Education, 58*(1), 386–396.

Espasa, A., Guasch, T., Mayordomo, R. M., Martínez-Melo, M., & Carless, D. (2018). A dialogic feedback index measuring key aspects of feedback processes in online learning environments. *Higher Education Research and Development, 37*(3), 499–513. https://doi.org/10.1080/07294360.2018.1430125.

Fawcett, H., & Oldfield, J. (2016). Investigating expectations and experiences of audio and written assignment feedback in first-year undergraduate students. *Teaching in Higher Education, 21*(1), 79–93.

Fish, W., & Lumadue, R. (2010). A technologically based approach to providing quality feedback to students: A paradigm shift for the 21st century. *Academic Leadership Live: The Online Journal, 8*(1), Article 5. Available at: https://scholars.fhsu.edu/alj/vol8/iss1/5.

Gould, J., & Day, P. (2013). Hearing you loud and clear: Student perspectives of audio feedback in higher education. *Assessment & Evaluation in Higher Education, 38*(5), 554–566.

Henderson, M., & Philips, M. (2015). Video-based feedback on student assessment: Scarily personal. *Australasian Journal of Educational Technology, 31*(1), 51–66.

Hennessy, C., & Forrester, G. (2014). Developing a framework for effective audio feedback: A case study. *Assessment & Evaluation in Higher Education, 39*(7), 777–789.

Hepplestone, S., Holden, G., Irwin, B., Parkin, H., & Thorpe, L. (2011). Using technology to encourage student engagement with feedback: A literature review. *Research in Learning Technology, 19*(2), 117–127.

Ice, P., Curtis, R., Phillips, P., & Wells, J. (2007). Using asynchronous audio feedback to enhance teaching presence and students' sense of community. *Journal of Asynchronous Learning Networks, 11*(2), 3–25.

Jackson, M. (2012). Improving the assessment feedback experience: A case study. *Enhancing Learning in the Social Sciences, 4*(3), 1–7.

King, D., McGugan, S., & Bunyan, N. (2008). Does it make a difference? Replacing text with audio feedback. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education, 3*(2), 145–163.

Lunt, T., & Curran, J. (2010). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education, 35*(7), 759–769.

Macgregor, G., Spiers, A., & Taylor, C. (2011). Exploratory evaluation of audio email technology in formative assessment feedback. *Research in Learning Technology, 19*(1), 39–59.

Mahoney, P., MacFarlane, S., & Ajjawi, R. (2018). A qualitative synthesis of video feedback in higher education. *Teaching in Higher Education, 24*(2), 157–179. https://doi.org/10.1080/13562517.2018.1471457.

Marriott, P., & Teoh, L. K. (2012). Using screencasts to enhance assessment feedback: Students' perceptions and preferences. *Accounting Education, 21*(6), 583–598.

Mathieson, K. (2012). Exploring student perceptions of Audiovisual feedback via Screencasting in online courses. *American Journal of Distance Education, 26*(3), 143–156.

Mathisen, P. (2012). Video feedback in higher education – A contribution to improving the quality of written feedback. *Nordic Journal of Digital Literacy, 7*(2), 97–116.

Mayhew, E. (2017). Playback feedback: The impact of screen-captured video feedback on student satisfaction, learning and attainment. *European Political Science: EPS, 16*(2), 179–192. https://doi.org/10.1057/eps.2015.102.

McCarthy, J. (2015). Evaluating written, audio and video feedback in higher education summative assessment tasks. *Issues in Educational Research, 25*(2), 153–169.

Merry, S., & Orsmond, P. (2008). Students' attitudes to and usage of academic feedback provided via audio files. *Bioscience Education eJournal, 11*, 1–11.

Morris, C., & Chikwa, G. (2016). Audio versus written feedback: Exploring learners' preference and the impact of feedback format on students' academic performance. *Active Learning in Higher Education, 17*(2), 125–137.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*(5), 501–517.

Orsmond, P., Maw, S., Park, J., Gomez, S., & Crook, A. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education, 38*(2), 240–252.

Pardo, A., Jovanovic, J., Dawson, S., Gasevic, D., & Mirriahi, N. (2017). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*, 128–138. https://doi.org/10.1111/bjet.12592.

Pitt, E. (2017). Students' utilisation of feedback: A cyclical model. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 145–158). Singapore: Springer.

Pitt, E., Bearman, M., & Esterhazy, R. (2019). The conundrum of low achievement and feedback for learning. *Assessment and Evaluation in Higher Education, 45*, 239–250. https://doi.org/10.1080/02602938.2019.1630363.

Pitt, E., & Norton, L. (2016). 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education, 42*(4), 499–516. https://doi.org/10.1080/02602938.2016.1142500.

Pitt, E., & Winstone, N. (2018). The impact of anonymous marking on students' perceptions of fairness, feedback and relationships with lecturers. *Assessment & Evaluation in Higher Education, 43*(7), 1183–1193. https://doi.org/10.1080/02602938.2018.1437594.

Sambell, K. (2013). Engaging students through assessment. In E. Dunne & D. Owen (Eds.), *The student engagement handbook: Practice in higher education* (pp. 379–396). Bingley: Emerald.

Stannard, R. (2007). *Using screen capture software in student feedback*. HEA English Subject Centre Commissioned Case Studies. Retrieved from http://english.heacademy.ac.uk/2016/01/16/using-screen-capture-software-in-student-feedback/

Turner, W., & West, J. (2013). Assessment for 'digital first language' speakers: Online video assessment and feedback in higher education. *International Journal of Teaching and Learning in Higher Education, 25*(2), 288–296.

Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804–1817.

West, J., & Turner, W. (2016). Enhancing the assessment experience: Improving student perceptions, engagement and understanding using online video feedback. *Innovations in Education and Teaching International, 53*(4), 400–410.

Winstone, N., & Pitt, E. (2017, September 14). Feedback is a two-way street. So why does the NSS only look one way? *Times Higher Education*. Retrieved from https://www.timeshighereducation.com/opinion/feedback-two-way-street-so-why-does-nss-only-look-one-way

Winstone, N., Nash, R., Parker, M., & Rowntree, J. (2017a). Supporting learners' engagement with feedback: A systematic review and a taxonomy of Recipience processes. *Educational Psychologist, 52*(1), 17–37.

Winstone, N., Nash, R., Rowntree, J., & Parker, M. (2017b). 'It'd be useful, but I wouldn't use it': barriers to University Students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026–2041.

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education, 18*(3), 285–297.

Yuan, J., & Kim, C. (2015). Effective feedback design using free technologies. *Journal of Educational Computing Research, 52*(3), 408–434.

Edd Pitt is a Senior Lecturer in Higher Education and Academic Practice and the Programme Director for the Post Graduate Certificate in Higher Education at the University of Kent, UK. He is an Honorary Fellow at the Centre for Research in Assessment and Digital Learning, Deakin University, Australia. His principle research field is Assessment and Feedback with a particular focus upon students' active role and emotional processing during feedback situations.

Naomi Winstone is a cognitive psychologist specialising in learning behaviour and engagement with education. Naomi's research focuses on the processing and implementation of feedback, educational transitions, and educational identities. Naomi is a Senior Fellow of the Higher Education Academy, and a National Teaching Fellow. Her work has been published in leading journals including *Educational Psychologist, Studies in Higher Education,* and *Assessment and Evaluation in Higher Education,* as well as the *Times Higher.* Naomi is Associate Editor of the *British Journal of Educational Psychology*, and Director of the Surrey Assessment and Learning Lab.

Chapter 8
Catering for Diversity in the Digital Age: Reconsidering Equity in Assessment Practices

Lois Ruth Harris and Joanne Dargusch

> *Assessment is used to provide a rationale and legitimacy for the social structures and power relations of modern day societies, and for one's place within these.*
> Leathwood (2005, p. 307–308)

Abstract While the affordances of the digital age certainly enable more diverse students to access higher education, higher education assessment is often underpinned by notions of equality rather than equity. Drawing together key literature and data from interviews with 53 first year undergraduate students from low socio-economic status backgrounds, this chapter identifies three potential causes of assessment inequity which appear to persist into the digital age: student assessment self-efficacy, prior preparation, and external pressures. It then identifies how the affordances of modern technology can be used to help combat these challenges.

8.1 Introduction

There is no question that the affordances of the digital age are making higher education accessible to populations who previously would not have been able to study for geographic, health, family, or financial reasons (Bearman et al., Chap. 3, this volume). Online learning platforms and video conferencing software, among other developments, have enabled universities to increase offerings at regional satellite

L. R. Harris (✉)
School of Education and the Arts, Central Queensland University, Rockhampton, QLD, Australia
e-mail: l.harris@cqu.edu.au

J. Dargusch
School of Education and the Arts, Central Queensland University, Noosa, QLD, Australia
e-mail: j.dargusch@cqu.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_8

campuses and via blended and fully online modes of study. Hence, these technological tools allow universities to more effectively reach learners at home and in their workplaces.

Governments worldwide have supported the growth in flexible study options as it aligns with national goals of increasing human capital and equity (Burke 2012; Leathwood 2005; Pitman et al. 2019). For example, within Australia, the Bradley Review (Bradley et al. 2008), subsequent higher education policy (e.g., demand driven student places), and available funding sources [e.g., Higher Education Participation and Partnerships Program (HEPPP), and the National Priorities Pool (NPP)] have all encouraged universities to increase non-traditional student enrolments, particularly those from low-socioeconomic status (low SES) backgrounds. Internationally, governments have gone about attempting to widen participation in different ways (Burke 2012) and issues of context should not be ignored given complex social and system differences (Younger et al. 2019). Pitman et al. (2019) argue that support for these kinds of policies appears based on the "general assumption that increased access and participation for disadvantaged students will lead, ipso facto, to consequential post-graduation benefits" (p. 46).

However, research suggests that growth in non-traditional student enrolment numbers does not automatically translate to more equitable higher education outcomes (e.g., Giani 2016; Pitman et al. 2019; Webb et al. 2017). As Altbach et al. (2019, p. viii) note "making higher education more inclusive requires not only moving historically underrepresented groups into higher education, but also meeting their unique needs". There is also a need to better understand "the enduring influence of social class and family influence on participation and outcomes" (Webb et al. 2017, p. 138). It is important to examine equity considerations across all phases of higher education: access (who gets to enter); survival (who gets to remain as long as they wish); output (who can achieve well on assessments); and, outcome (who gets improved income, employment, and political power as a result of their studies) (Farrel 1999), and assessment results (whether from high school assignments, entrance exams, or undergraduate course work) influence student outcomes during all phases. In research into equity, access and survival (i.e., admission and retention rates; see See et al. 2012 and Younger et al. 2019 for systematic reviews) are more frequently the focus than the outputs and longer term outcomes for students from diverse backgrounds (e.g., Giani 2016; Li and Carroll 2017; Pitman et al. 2019). This chapter seeks to contribute to the literature examining equity in relation to output.

However, creating equitable assessment outcomes is complicated. The notion that assessment systems are impartial and fairly reward those with ability and sufficient effort has contributed to discourses which often position students from non-traditional backgrounds as responsible for their own lack of assessment success due to insufficient effort, commitment, preparation, or ability (McKay and Devlin 2016; O'Shea et al. 2016). This chapter explores how assessment systems in the digital age might need to change if equity is a major goal of higher education. Drawing on student voice data from the HEPPP funded Supporting Students' Assessment Success (SSAS) project, it identifies potential sites of assessment inequity for low SES students which persist within the digital age, proposing possibilities which

may help diverse students more effectively and accurately demonstrate their learning.

8.2 Walking the Tightrope: Balancing Assessment Accountability and Equity

While assessment sits as part of a larger pedagogical system, there is no question that it often drives student learning. As Boud (2000, p. 155) notes, summative assessment "provides an authoritative statement of 'what counts' and directs student attention to those matters. It tells us what to learn". There is evidence that many students strategically prioritise their learning around assessment expectations (Harris et al. 2018). If students are to experience deep learning via their assessment tasks, long-standing calls for tasks to be authentic and sustainable, and to include opportunities for formative assessment and feedback seem well grounded (Boud 2000).

However, how do quality assessment practices simultaneously promote deep learning and student equity? While the terms equity and equality are often used almost synonymously within policy, they have different theoretical underpinnings and meanings which have significant implications for assessment. Espinoza (2007) explains:

> The 'equity' concept is associated with fairness or justice in the provision of education or other benefits and it takes individual circumstances into consideration, while 'equality' usually connotes sameness in treatment by asserting the fundamental or natural equality of all persons. (Espinosa 2007)

Within an assessment context, Tierney (2013) suggests that "equality is maintained with the same tasks and criteria for all learners, while equity involves differentiating according to students' needs" (p. 133).

The idea that equity might require consideration and differentiation based on individual difference is clearly at odds with the notion of standardisation which underpins most higher education assessment policies. Higher education course designers must consider accountability concerns, including, but not limited to: satisfying accountability and accreditation requirements from the institution, government bodies, and external industry stakeholders; ensuring tasks are resistant to academic dishonesty; and creating comparability across varying study modes. These powerful accountability systems are usually underpinned by the 'equality' notion that standardisation (i.e., everyone experiencing the same task under the same 'equal' conditions) is the most appropriate way to promote fairness. The notion that identical treatment is 'fair' runs deep. Studies indicate university staff and students struggle to see how assessment and its conditions can be altered for individuals without impacting its validity, even when adjustments are made for legitimate reasons (e.g., disability) (Meyer et al. 2010; O'Shea et al. 2016).

Equity, however, might require us to move beyond notions of standardisation as fairness. McArthur (2016) argues that it is important to consider the fundamental inequalities between students as they participate in university assessment practices

and how these can be redressed in ways which allow all to demonstrate learning. For example, when discussing assessment extensions, she points out that while illness is viewed as legitimate grounds for special consideration, negative impacts caused by poverty (e.g., having to undertake paid work or do without necessary books or computer access) are unrecognised. She argues that:

> … these supposedly 'fair' systems are themselves highly selective and based on socially constructed notions of what should and should not count, and these may deserve rethinking. (p. 973)

As Leathwood (2005) identifies:"'Standards', 'quality' and 'assessment' are not neutral and value-free" (p. 320). Given the many newly possible affordances of technology, it seems timely to reconsider these socially constructed meanings within the context of the digital age.

8.3 Gathering Data about Assessment Inequity

While there are many potential markers of disadvantage within higher education (e.g., First Nations background, rural or remote location, language status, women in non-traditional areas), this chapter will focus on those who are of low SES, given the Australian government's current attention to this group. However, it is important to note that many low SES students carry multiple potential markers of disadvantage. For example, in Term 1, 2017 of our study, of the 420 student participants categorised as low SES, 389 (92%) were also classed as regional or remote, 11 (3%) reported a disability, 19 (5%) identified as Aboriginal or Torres Strait Islanders, 18 (4%) were from non-English speaking backgrounds, and 154 (37%) were women studying in non-traditional areas [percentages rounded to the nearest whole number]. Hence, student categorisations are seldom clear cut or discrete.

To examine potential examples of assessment inequity persisting into the digital age, a review of literature from the past 18 years (2000–2018) was drawn upon when examining new empirical data from the HEPPP funded Supporting Students' Assessment Success (SSAS) Project. To gather literature, major search engines (e.g., Scopus, Google Scholar) were examined using combinations of key words such as "low SES", "assessment", "higher education", and "equity". Relevant papers were also obtained from the reference lists of papers reviewed.

Through the collection and analysis of empirical data, the SSAS project sought to explore first year students' experiences of assessment, with a focus on low SES students. These data informed interventions including changes to tasks and the creation of support materials with the goal of improving assessment equity at a multi-campus Australian university (for more results from this study, see: Dargusch et al. 2017a, b; Harris et al. 2018; Taylor et al. 2017). From the project's wider data set, this chapter draws primarily upon data gathered from n = 53 telephone interviews with low SES students to highlight the challenges which must be overcome to achieve equity within the digital age. Drawing on categorical analysis (Coffey and

Atkinson 1996), interview data were first read multiple times by the first author before codes were developed relating to potential assessment inequities these students faced. Particular attention was paid to interactions between assessment and student characteristics which might be related to low SES (e.g., income, employment, preparation). Themes generated from the empirical data were then matched with themes derived from the reviewed literature; three major types of potential inequity were identified in relation to assessment.

However, within our data and the greater literature, the heterogeneous nature of low SES students soon became apparent; hence, we encourage readers to use these categories as a starting point for considering common challenges such students may face. While Harvey et al. (2016) identify that "… the primary cause of underrepresentation of low SES students remains low school achievement" (p. 70), work by McKay and Devlin (2016) reminds us that many low SES students perform well and exhibit extraordinary resilience in the face of challenge. Interviewed students varied greatly in age (17 to 55), previous educational experiences (early school leavers to those with university degrees), industry knowledge and experience, and personal responsibilities (e.g., work, family, community involvement). Students also had different ideas about the purpose of assessment and their role as a student. Hence, the sites of potential assessment inequity we identify do not impact upon all low SES students equally (or even at all) and are unlikely to constitute an exhaustive list. Instead, they illustrate some of the most prevalent challenges which may disproportionately impact upon low SES students studying in the digital age, many of whom also share other markers of disadvantage.

8.4 Examining Potential Causes of Assessment Inequity

Across our own data and the wider literature, we identified three potential sites of assessment inequity for low SES students which persist within the digital age: student assessment self-efficacy, prior preparation, and external pressures. Drawing on our own qualitative data and examples sourced from the literature, we will illustrate the practical problems these can create for students as they attempt to complete university assessments, particularly for the first time.

8.4.1 Assessment Self-Efficacy

One major challenge low SES students may experience is around their self-efficacy in relation to assessment situations, particularly within their first year. Jury et al. (2017) report the presence of many inappropriate and negative stereotypes of low SES students. When staff and students internalise these, this can lead to lower student self-efficacy and/or possible discrimination from staff. For example, students

coming from lower income backgrounds may feel like they do not belong or deserve to be at university (often referred to as the imposter phenomenon, Parkman 2016).

Our data contained numerous examples of students reporting low self-efficacy in relation to assessment. For example, Ellen, an Engineering student, described her initial reactions to the assessment task saying "Like what have I done? What have I got myself into? There's no way I can pass, there's no way I can keep up." Within the interview, she appeared to display elements of the imposter phenomenon (Parkman 2016), clearly questioning if she belonged in higher education:

> I – I just didn't have the experience that my peers had…. I was like the mum in the group… I was quite intimidated at, you know, the level of experience that, you know, my team mates had.

For many students within our study (including Ellen, who ultimately achieved a Distinction in the unit), these negative beliefs were unfounded. However, when students lacked confidence, they were also less likely to seek necessary help, particularly if initial requests were not handled in a supportive manner ("I didn't want to go and cause any more issues [by asking more assessment questions]" Imogen, Business and Law).

As assessment experiences shape students' views of their own competence as learners (Christie et al. 2008; O'Shea 2014), it seems particularly important to support them through their first experiences of university assessment. As Leathwood and O'Connell (2003) identify:

> the impact of what are perceived to be poor assessment results on those with low self-esteem, who already feel that they can never be good enough or never get it right, can be profound. (p. 609)

Hence, when seeking equitable assessment outcomes for low SES students, it is necessary to consider the interaction between students' personal beliefs about themselves as learners (and their place within a university) and their assessment experiences and results.

Additionally, within our data, most interviewed low SES students did not challenge or criticise the assessment tasks and conditions, instead taking responsibility for their own lack of success if they had not done well. Student comments include:

> I clearly hadn't met the marking rubric in that particular area and that was very clearly my own folly and nothing to do with the assignment. (Noah, Engineering)
>
> I, myself, was not prepared I suppose and that's my fault. (Imogen, Business & Law)
>
> I hadn't realised that they [quizzes] had a cut-off date…. But that could be partially my own fault too for not particularly paying attention to where they say it may be cut off. I'm not too sure. (Eloise, Business & Law)

The only time assessment's validity or appropriateness was questioned was around peer-assessment results. Students voiced concerns about whether they were sufficiently competent and impartial to act as assessors, a concern which has been documented extensively elsewhere (Panadero 2016). Hence, a challenge to equity seems to exist around helping students from low SES backgrounds feel empowered and able to meet assessment expectations and avoid negative patterns of self-blame if initial results are not as hoped.

8.4.2 Uneven Preparation

An additional source of assessment inequity may stem from students' uneven preparation for study. For example, Leathwood and O'Connell (2003) found that many of the non-traditional students in their study felt they "had been left to sink or swim" (p. 610). In relation to assessment, work consistently highlights the need to help students develop assessment and feedback literacy (Carless and Boud 2018; Haggis 2006). Devlin and O'Shea (2012) found that low SES students particularly noted the importance of lecturers clearly explaining assessment tasks and providing exemplars to help them visualise expectations, with many recommending such scaffolding and support is necessary (e.g., Bearman and Ajjawi 2018; Broadbent et al. 2018; Dargusch et al. 2017b). However, as Haggis (2006) notes, there is sometimes lecturer resistance to providing scaffolding, with some describing it as 'spoon feeding' (p. 532).

SSAS project data highlighted that many low SES students had exaggerated strengths and weaknesses. For example, professional experience meant some had extensive prior knowledge of subject matter and skills ("We did the Clinical in the Res School and I breezed through that. That was things we do every day [in her aged care job]" Alice, Nursing). However, despite this extensive industry experience, Alice, who explained "I have a special needs child and a baby and I work full time", still had to travel over 1000 km to attend a mandatory residential school to demonstrate her 'every day' skills. Conversely, many were underprepared for some of the demands of higher education assessment. As Noah from Engineering explained "I sort of had a bit of a knowledge gap in some areas."

The students we interviewed often described their first assessment experiences as challenging and daunting as they struggled to unpack the assessment task and marking criteria. As Elsa, a Business and Law student, explained "it's your first semester so you need time to understand the expectations of what a university assignment requires." Students often described being unable to gauge whether their work met expectations: "I was pretty nervous… even when I had submitted my second one, I still didn't know whether it was a pass or a fail" (Ned, Education). As Ned explained:

> I was going in completely blind.… I had no idea if I was even answering the question right… I like simpler words so when I read the criteria sheet, I'm like "oh my god, this is overcomplicated. I don't know if this is right. I am answering the right questions?"

Ned is representative of many of the students we interviewed who experienced difficulty understanding what was expected of them, demonstrating the need to bring such students "into the guild of professionals" (Sadler 2009, p. 57) through the development of assessment and feedback literacy (Carless and Boud 2018). Given the potential gaps in experience, low SES students in particular may need additional support to understand common genres within higher education and be helped to unpack and visualise task expectations. However, it is also important to remember the heterogeneous nature of these low SES students in relation to preparedness and avoid stereotypes about areas of strength and weakness.

8.4.3 External Pressures

Data from both our study and the wider literature indicate that low SES students are likely to be balancing assessment requirements with other professional and personal obligations. Studies have identified that students with complex work and family responsibilities often have difficulty finding time to complete learning and assessment tasks, particularly when multiple pieces of work are due at the same time or when unexpected events occur such as illness, increased work demands, or family emergencies (Christie et al. 2008; Harris et al. 2018).

Within our study, work seemed to particularly impact upon assessment success. For example, Business and Law student Roxanne explained, "I submitted my assignment early because I had work commitments on and I probably, in hindsight, I should have looked over that, because I didn't get the mark that I wanted." Students working part-time jobs also experienced variations in hours which did not always work well with peak study times within the term. Yana, another Business and Law student explained:

> I work at a real estate agent and I fill-in for other staff so, at a moment's notice, it can become full-time. And in the middle of the course, I ended up having to do three weeks' full-time work as well as trying to do three weeks' full-time university and have assignments due in at the same time. So, I know the third week got pretty messy ….and I definitely did less – that was probably my lowest quiz.

Low SES Business and Law student Alyssa described how her unanticipated move from part- to full-time work in the middle of term made it impossible for her to submit her final assignment, dropping her unit grade from a High Distinction to a Pass:

> I was granted an extension only to the point that [the course coordinator] could grant it and it just became a bit – well, I've already passed it and Ps get degrees…. I would have loved to have submitted, but it just wasn't a realistic timeframe for me…. I don't feel that I've failed in any way, I just feel that if we're trying to be reflective of what life is really like,… perhaps would have been a little more leeway.

Alyssa was one of the only students within our entire study who questioned the appropriateness of current practices around assessment extensions, arguing for 'leeway'. Clearly, non- or under-performance on assessment like this has serious implications for students' survival within their courses of study as well as their potential for employment at the end. In Alyssa's case, the Pass she received for the unit clearly does not reflect her actual capabilities, yet it may impact upon her employment chances on graduation given grades can be seen as a proxy for capability.

Within our study, some staff recognised the challenges that such students were experiencing and made minor adjustments to deadlines. For example, Business and Law student Elsa reported:

> I emailed him [the course coordinator] once saying, 'Hey, I thought the quiz was today, so I've logged on now, but it actually finished yesterday. Can you reopen it?' And he was like, 'Yep, no worries. It's reopened as a once-off.' So I reached out at one point and I got some assistance that I required.

However, while this lecturer's decision was clearly in the best interests of this student, it does have implications for 'equality'; other students may have forgotten or been unable to complete the quiz, but did not have the courage to ask for extra time. It also potentially represents a breech in the university's assessment protocol, where any such adjustments should be in response to a formal extension request supported by appropriate documentation. These two examples (one where the lecturer intervened and overrode policy and one where the lecturer did not) illustrate the tensions between complying with policies and systems and supporting students when they need it.

8.5 Harnessing the Affordances of the Digital Age to Improve Assessment Equity

The data from our study and the larger research literature highlight common challenges low SES students may experience when navigating assessment, particularly in first year courses, in relation to self-efficacy, prior preparation, and external pressures. In this section, we propose ways the affordances of the digital age may be better harnessed to address these inequities via more nuanced recognition of prior learning opportunities before students begin, better mechanisms to help students identify and receive academic support during their first year, and more flexible timelines within units of study throughout their degree.

8.5.1 Using Technology to Make Recognition of Prior Learning more Nuanced

Our data showed that some low SES students are wasting valuable time and resources demonstrating knowledge and skills they already have. While universities may have mechanisms in place for recognition of prior learning (RPL), it is important to make sure students are aware of these opportunities and consider how digital technologies (e.g., ability to assess student participation in practical tasks, experiments, group work via digital assessment modes, Katz and Gorin 2017) could make them far more nuanced. For example, while a student may have excellent practical knowledge and skills, there may be gaps in theoretical knowledge, so exempting students from entire units or years of study may not be appropriate. If equity is the goal, perhaps assessment systems can be developed where evidence of prior learning and competency (or a set of pre-tests prior to the unit) can be used to exempt students with prior industry or educational experience from individual assessments tasks within particular units. Such mechanisms may enable time poor students to devote additional time to areas of need and potentially spare them from some costly on-campus assessments. It might also encourage students who may have low self-efficacy to formally

acknowledge their areas of strength. A more nuanced approach to RPL might also be supported via a move from more traditional grades and reporting to reporting by learning outcomes or by digital credentialing (see Jorre de St Jorre, Chap. 19, this volume) as this would potentially allow students to draw on pre-existing, as well as newly generated, evidence of learning to address criteria.

8.5.2 Using Technology to Help Students Identify and Address Academic Needs

While students may be aware of some gaps within their preparation and know about resources available via student support units (e.g., help with academic literacy, mathematics, computing, etc.), unless the student is referred to services due to assessment failure, most current approaches require the student to identify needs and seek help. It seems vital that students are supported to become aware of any potential gaps and have opportunities to learn 'assumed' knowledge and skills before submitting their first assessment; this could occur in different ways. There are already many different ways technology is being used within online and blended enabling programs, but as Rienties et al. (2012) identify, more work is needed to make sure use of technology is clearly linked to learning objectives and pedagogy, maximising the benefit these programs have for aspiring higher education students.

Opportunities for formative feedback could be present within the first few weeks of a unit via a low-stakes early assessment task (Shields 2015), ideally directing students to specific tutorials and resources to upskill them in the areas of need. For example, online quizzes can provide students with immediate feedback on their understandings of content, although clearly this is an inappropriate way to assess many important facets of knowledge and skills students must develop during their studies. Online peer-assessment systems can also help students gain useful feedback from peers via Mobile Response Technology (MRT) and software packages (Rotsaert et al. 2018).

Another possibility is via online diagnostic testing, like that currently being developed by Jason Bryman, Timothy Cleary, and Heidi Andrade in the United States (for more information, see http://daacs.net/). Their open access Diagnostic Assessment and Achievement of College Skills (DAACS) tool allows students to complete short diagnostic tests in self-regulated learning, reading, mathematics, and writing. These provide students with feedback about current competency in these domains and link students with resources to help them develop in specific areas of need, allowing them to improve via use of self-paced, open-access materials. It is vital that students with potentially low self-efficacy are empowered to feel that they can improve and tools like this can provide them with the pathways through which to do so. Early findings exploring DAACS' impacts on student university success indicate that having students simply take the test is not enough to improve outcomes (Bryer et al. 2019). Students and/or their advisors must meaningfully engage with

the results and invest the time needed to work with resources designed to help students improve in areas of need. Bryer et al. (2019) found that academically weaker students were less likely to engage meaningfully with DAACS feedback and resources, suggesting that such students may need additional forms of support.

In addition to support for students around global areas of strength and weakness in their prior knowledge and skills, there is also potential need to focus student attention on how to participate effectively in higher education assessment processes (Smith et al. 2013). Students from equity group backgrounds may require additional support to help induct them into "the guild" (Sadler 2009). During our project, students identified many forms of scaffolding which can help students self-assess their progress against goals including checklists, exemplars, and rubrics. Of particular value were short videos shared on the Learning Management System (LMS) which featured the lecturer unpacking the assessment task (Taylor et al. 2017); these on-demand videos allowed students to view them when and as many times as they wished as they sought to understand expectations. Likewise, many students reported that seeing an exemplar, particularly when a type of task is new to them, made it much easier to visualise assessment expectations, consistent with the literature (e.g., Bearman and Ajjawi 2018). Technological affordances can make exemplars far more interactive than they have been in the past (e.g., links within the document to lecturer video or audio comments about the work; on-demand videos walking students through an exemplar) and such possibilities should be further explored.

Students also benefit from formative feedback (with many new online modes being developed for such delivery, Gikandi et al. 2011; Henderson and Phillips 2015). However, they also need the time and space to make use of it (including that gained via their own self-assessment using tools like rubrics and exemplars or feedback from peers). Regardless of what tools and scaffolding are selected, students need to be supported to understand and use them in the ways intended.

8.5.3 Using Technology to Create Flexibility in Course Timelines

Finally, it is vital to move beyond the notion that standardization equals equity to a philosophy which recognises the importance of giving each student the conditions and time needed to demonstrate learning objectives. Particularly within distance education environments, which can be more self-paced as they are less reliant on weekly scheduled contact between teaching staff and students, there is no fundamental reason why assessment completion must be confined to the 10–12 week conventional university term. While giving students longer periods of time to complete unit assessment tasks may mean major changes to the bureaucracy, rules, and procedures around higher education assessment, it seems important to be realistic about the complex lives and pressures many equity group students (and others) experience and work with them in supportive ways to help them achieve the best possible results.

Some institutions are already experimenting with this model, but primarily at the post-graduate level within Australia (e.g., CQUniversity MBA (Leadership) program, Deakin University's Start Anytime units). However, Open Universities Australia has also adopted this model for its Unilearn bridging program. While such an approach may work best for more experienced and self-regulated learners, there are potential ways of applying the model at all levels which could be explored.

In Australia, we can also learn from some of the successes and challenges reported by American universities employing more flexible and self-paced undergraduate and post-graduate degree models, often referred to as Competency-Based Degrees (McDonald 2018). The flexibility and potentially lower costs offered by these models may be of great help to students who are juggling complex priorities. However, research indicates that this way of studying places high expectations on students (e.g., students need to be able to motivate themselves in the absence of external deadlines, self-assess their own learning competencies and needs, possess good time-management skills, McDonald 2018). Staff operating such programs may also face increased challenges (e.g., such programs may lack alignment with other university timelines and procedures, face additional regulatory hurdles, experience resistance from staff, Dragoo and Barrows 2016). Additionally, research identifies constant tensions reported between keeping institutional and student costs down, while simultaneously creating high quality student learning environments and experiences (Dragoo and Barrows 2016; McDonald 2018).

Clearly, while some changes to system rules may be relatively easy to address (e.g., having more points within the calendar year when grades are verified), more work is still needed to further explore and mitigate the many challenges noted by those trying to create more assessment flexibility within their degree programs. In the interim, it at least seems logical to carefully examine university assessment extension policies, making sure that these are sufficiently flexible so that lecturers and students can work together to create reasonable deadlines which allow students the opportunity to demonstrate their knowledge and skills.

We are only starting to scratch the surface of what may be possible via digital assessment (Katz and Gorin 2017) and how these kinds of experiences may allow more flexibility for distance learners particularly (e.g., assessing distance students on their completion of an experiment in a digital environment while on-campus students complete the task during a laboratory session), keeping focused on what learning objectives students are demonstrating rather than how this is being accomplished or under what conditions. At a minimum, we suggest following the recommendation of Luzeckyj et al. (2015) that assessment deadlines across programs are examined to make sure they are spread across the term and that submission times do not encourage students to act in ways which are detrimental to their health and well-being (e.g., a 9:00 am submission time may encourage students to stay up all night working on an assessment).

8.6 Conclusion

This chapter poses the question of how assessment might need to change if equity is a major goal of higher education. It identifies key considerations which are needed around student assessment preparation, task and scaffolding design, and assessment conditions. The affordances of the digital age are certainly helping break down barriers to study for those living in regional and remote areas and juggling complex personal and work commitments. However, there is clearly more which can be done to utilise technology in ways which will serve the interests of students from equity group backgrounds. Meyer et al. (2010) recommend creating policy at a university level, outlining how assessment practices will best promote equity, with a focus on examining how student needs are accommodated, reflecting on whether chosen genres are likely to privilege particular groups, and monitoring and troubleshooting around the progress of students from non-dominant groups. It is also worth reconsidering how grades are distributed; if assessment is conducted in reference to standards, all students have the opportunity to demonstrate mastery and experience success, potentially rendering norm-based approaches like grading on the bell-curve as obsolete.

Within higher education, there are competing ideas about equality on one hand (which many interpret as standardisation) and equity on the other. We believe that equity is about acknowledging the challenges students may be experiencing and making appropriate changes in our own assessment design, conditions, and implementation to allow each to achieve his or her best possible results. Technology has already allowed us to make great inroads into this process. We look forwards to seeing how it can help open up even more equitable assessment possibilities in the future, acknowledging that "To achieve equity—justice—may require structured inequalities, at least temporarily" (Samoff 1996, p. 267).

References

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2019). *Trends in global higher education: Tracking an academic revolution*. Taipei: Sense Publishing.

Bearman, M., & Ajjawi, R. (2018). From "seeing through" to "seeing with": Assessment criteria and the myths of transparency. *Frontiers in Education, 3*(96). https://doi.org/10.3389/feduc.2018.00096.

Bearman, M., Dawson, P., & Tai, J. (This volume). Chapter 3: Digitally mediated assessment in higher education: Ethical and social impacts. In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world*. Dordrecht: Springer.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian higher education*. Canberra: Australian Government.

Bryer, J. M., Lui, A. M., Andrade, H. L., Franklin, D. W., & Cleary, T. J. (2019). *Efficacy of the Diagnostic Assessment and Achievement of College Students on Multiple Success*

Indicators. Paper presented at the American Educational Research Association annual meeting, Toronto, Canada.

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing summative assessment with a formative flavour: A case study in a large class. *Assessment and Evaluation in Higher Education, 43*(2), 307–322. https://doi.org/10.1080/02602938.2017.1343455.

Burke, P. J. (2012). *The right to higher education: Beyond widening participation*. London: Routledge. https://doi.org/10.4324/9780203125571.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.

Christie, H., Tett, L., Cree, V. E., Hounsell, J., & McCune, V. (2008). 'A real rollercoaster of confidence and emotions': Learning to be a university student. *Studies in Higher Education, 33*(5), 567–581. https://doi.org/10.1080/03075070802373040.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. London: Sage.

Dargusch, J., Harris, L. R., Reid-Searl, K., & Taylor, B. A. (2017a). Creating first-year assessment support: Lecturer perspectives and student access. *Distance Education, 38*(1), 106–122. https://doi.org/10.1080/01587919.2017.1299566.

Dargusch, J., Harris, L R., De Vries, P., & Lawson, W. (2017b). *Why showing is better than telling: Using exemplars to improve 1st year students' assessment understanding and achievement.* Paper presented at Australian Association for Research in Education (AARE) 2017 Conference, Canberra.

Devlin, M., & O'Shea, H. (2012). Effective university teaching: Views of Australian university students from low socio-economic status backgrounds. *Teaching in Higher Education, 17*(4), 385–397. https://doi.org/10.1080/13562517.2011.641006.

Dragoo, A., & Barrows, R. (2016). Implementing competency-based education: Challenges, strategies, and a decision-making framework. *The Journal of Continuing Higher Education, 64*(2), 73–83. https://doi.org/10.1080/07377363.2016.1172193.

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: A new model for analysis of the educational process. *Educational Research, 49*(4), 343–363. https://doi.org/10.1080/00131880701717198.

Farrel, J. (1999). Changing conceptions of equality of education: Forty years of comparative evidence. In R. Arnover & C. Torres (Eds.), *Comparative education: The dialectic of the global and the local* (pp. 149–178). Rowan & Littlefield: New York.

Giani, M. S. (2016). Are all colleges equally equalizing?: How institutional selectivity impacts socioeconomic disparities in graduates' labor outcomes. *The Review of Higher Education, 39*(3), 431–461. https://doi.org/10.1353/rhe.2016.0013.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004.

Haggis, T. (2006). Pedagogies for diversity: Retaining critical challenge amidst fears of 'dumbing down'. *Studies in Higher Education, 31*(5), 521–535. https://doi.org/10.1080/03075070600922709.

Harris, L. R., Brown, G. T. L., & Dargusch, J. (2018). Not playing the game: Student assessment resistance as a form of agency. *The Australian Educational Researcher, 45*(1), 125–140. https://doi.org/10.1007/s13384-018-0264-0.

Harvey, A., Andrewartha, L., & Burnheim, C. (2016). Out of reach? University for people from low socio economic status backgrounds. In A. Harvey, C. Burnheim, & M. Brett (Eds.), *Student equity in Australian higher education: Twenty-five years of a fair chance for all* (pp. 69–85). Singapore: Springer.

Henderson, M., & Phillips, M. (2015). Video-based feedback on student assessment: Scarily personal. *Australasian Journal of Educational Technology, 31*(1). https://doi.org/10.14742/ajet.1878.

Jury, M., Smeding, A., Stephens, N. M., Nelson, J. E., Aelenei, C., & Darnon, C. (2017). The experience of low SES students in higher education: Psychological barriers to success and interventions to reduce social-class inequality. *Journal of Social Issues, 73*(1), 23–41. https://doi.org/10.1111/josi.12202.

Katz, I. R., & Gorin, J. S. (2017). Computerising assessment: Impacts on education stakeholders. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 472–489). New York: Routledge.

Leathwood, C. (2005). Assessment policy and practice in higher education: Purpose, standards and equity. *Assessment & Evaluation in Higher Education, 30*(3), 307–324. https://doi.org/10.1080/02602930500063876.

Leathwood, C., & O'Connell, P. (2003). 'It's a struggle': The construction of the 'new student' in higher education. *Journal of Education Policy, 18*(6), 597–615. https://doi.org/10.1080/0268093032000145863.

Li, I.W., & Carroll, D. (2017). *Factors influencing university student satisfaction, dropout and academic performance: An Australian higher education equity perspective*. Perth: National Centre for Student Equity in Higher Education (NCSEHE), Curtin University.

Luzeckyj, A., McCann, B., & Graham, C. (2015). *Exploring the experience of being first in family at university: A 2014 student equity in higher education research grants project*. Perth: National Centre for Student Equity in Higher Education, Curtin University.

McArthur, J. (2016). Assessment for social justice: The role of assessment in achieving social justice. *Assessment & Evaluation in Higher Education, 41*(7), 967–981. https://doi.org/10.1080/02602938.2015.1053429.

McDonald, N. A. (2018). A private, nonprofit university's experiences designing a competency-based degree for adult learners. *The Journal of Continuing Higher Education, 66*(1), 34–45. https://doi.org/10.1080/07377363.2018.1415632.

McKay, J., & Devlin, M. (2016). Low income doesn't mean stupid and destined for failure': Challenging the deficit discourse around students from low SES backgrounds in higher education. *International Journal of Inclusive Education, 20*(4), 347–363. https://doi.org/10.1080/13603116.2015.1079273.

Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., & Johnston, P. M. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly, 64*(3), 331–350. https://doi.org/10.1111/j.1468-2273.2010.00459.

O'Shea, S. (2014). Transitions and turning points: Exploring how first-in-family female students story their transition to university and student identity formation. *International Journal of Qualitative Studies in Education, 27*(2), 135–158. https://doi.org/10.1080/09518398.2013.771226.

O'Shea, S., Lysaght, P., Roberts, J., & Harwood, V. (2016). Shifting the blame in higher education – Social inclusion and deficit discourses. *Higher Education Research and Development, 35*(2), 322–336. https://doi.org/10.1080/07294360.2015.1087388.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247–266). New York: Routledge.

Parkman, A. (2016). The imposter phenomenon in higher education: Incidence and impact. *Journal of Higher Education Theory and Practice, 16*(1), 51–60.

Pitman, T., Roberts, L., Bennett, D., & Richardson, S. (2019). An Australian study of graduate outcomes for disadvantaged students. *Journal of Further and Higher Education, 43*(1), 45–57. https://doi.org/10.1080/0309877X.2017.1349895.

Rienties, B., Kaper, W., Struyven, K., Tempelaar, D., van Gastel, L., Vrancken, S., Jasińska, M., & Virgailaitė-Mečkauskaitė, E. (2012). A review of the role of information communication technology and course design in transitional education practices. *Interactive Learning Environments, 20*(6), 563–581. https://doi.org/10.1080/10494820.2010.542757.

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 45–63). Dordrecht: Springer.

Samoff, J. (1996). Which priorities and strategies for education? *International Journal of Educational Development, 16*(3), 249–271.

See, B. H., Gorard, S., & Torgerson, C. (2012). Promoting post-16 participation of ethnic minority students from disadvantaged backgrounds: A systematic review of the most promising interventions. *Research in Post-Compulsory Education, 17*(4), 409–422.

Shields, S. (2015). 'My work is bleeding'; exporing students' emotional responses to first-year assignment feedback. *Teaching in Higher Education, 20*(6), 614–624. https://doi.org/10.1080/13562517.2015.1052786.

Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education, 38*(1), 44–60. https://doi.org/10.1080/02602938.2011.598636.

Taylor, B., Harris, L. R., & Dargusch, J. (2017). Portfolio assessment in engineering: Student perspectives on effective implementation. *International Journal of Quality Assurance in Engineering and Technology Education, 6*(2), 1–21.

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125–144). Los Angeles: SAGE.

Webb, S., Burke, P. J., Nichols, S., Roberts, S., Stahl, G., Threadgold, S., & Wilkinson, J. (2017). Thinking with and beyond Bourdieu in widening higher education participation. *Studies in Continuing Education, 39*(2), 138–160.

Younger, K., Gascoine, L., Menzies, V., & Torgerson, C. (2019). A systematic review of evidence on the effectiveness of interventions and strategies for widening participation in higher education. *Journal of Further and Higher Education, 43*(6), 742–773. https://doi.org/10.1080/0309877X.2017.1404558.

Lois Harris works as a Senior Lecturer in the School of Education and the Arts at Central Queensland University, Rockhampton Campus and as a Senior Research Fellow at the Australian Catholic University's Institute for Learning Sciences and Teacher Education. Her research focuses on student engagement, classroom assessment, feedback, self-assessment, and tertiary assessment support, drawing on quantitative, qualitative, and mixed-method designs.

Joanne Dargusch is a Senior Lecturer in the School of Education and the Arts at CQUniversity, Noosa Campus. She is the Head of Course for the Graduate Certificate in Tertiary and Adult Education and also teaches undergraduate students in Bachelor of Education (Secondary) and supervises research higher degree candidates. Joanne's research examines assessment and pedagogy in both school-based and higher education contexts.

Chapter 9
Assessment as and of Digital Practice: Building Productive Digital Literacies

Marcus O'Donnell

Abstract Most modes of assessment in higher education, apart from paper-based exams, use some form of digital technology. However, assessments that scaffold, develop and assess integrated digital practices are much rarer. Even when using newer digital modes such as blogging, assessments often only replicate traditional textual practices with little attention paid to the specific multimodal, networked affordances of the medium. This paper will frame an approach to sustainable and productive digital assessment through integrated digital practice that emphasizes the reflexivity necessary for participatory networked communication. Such a framework seeks to move from the textual/compositional bias which governs much assessment practice in higher education towards genuine digital assessments which engage students as designers and assess their digital design literacies as well as their critical encounters with discipline knowledge.

Digital technologies are now a fundamental part of everyday life and most professional routines. In contemporary higher education all learning is in some sense mediated by technology: academics use digital devices to prepare and present their material. Students, in most cases, receive some of their learning resources via digital devices and most modes of assessment, apart from paper-based exams, use some form of digital technology, even if this is just word processing.

In spite of technologies' ubiquity, the most effective ways to use technology to enhance learning is still a live question, and as a number of scholars (Bayne 2015; Kirkwood and Price 2013a, 2014; Oliver 2013) have shown the concept of technology enhanced learning is still undertheorized and ill-defined. Kirkwood and Price have shown in a series of studies (2013a, b, 2014; Price and Kirkwood 2014) that technology enhanced learning is, more often than not, designed to increase efficiencies or add small supplements to traditional practices. In their meta-analysis of published studies only 30% of technology enhanced learning projects were designed to

M. O'Donnell (✉)
Deakin University, Melbourne, Australia
e-mail: marcus.odonnell@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_9

transform learning interactions through the affordances of technology. Sue Timmis and colleagues (Timmis et al. 2016) conducted a similar series of analyses of digital assessment practices and came to similar conclusions:

> Critiques in the literature suggest an over emphasis on technology for efficiency and the potential for standardising, grading and recording data. This appears to be limiting the development of more imaginative and creative possibilities. Where innovation is taking place, this is often in isolated pockets or subject to funding constraints on continuation and sustainability. Furthermore, policy discussions and decisions on assessment can contribute to an over simplified view of technology as a 'quick fix' or a means of replicating existing methods rather than seeing its potential to challenge and re-model existing practices and re-imagine the purpose of assessment and its relationship to learning. (p. 458)

So, what are the affordances of learning technologies that enable the types of transformative learning interactions that take us beyond instrumentalist content transactions? And how is this learning assessed in a congruent and meaningful way?

I have previously defined the key affordances of digital technologies as "the ability to generate, store, reuse, remix and share multimedia data in self-organising networks" (O'Donnell in press). This paper argues that conceptualising the digital in this way encourages us to focus on the iterative and collaborative possibilities for digital learning as a productive process that involves active participation and engagement within a network. I also argue that in a world where digital practices are transforming both our everyday life (Säljö 2010) and the world of work (Aoun 2017; Marsh 2018) assessment in higher education must specifically warrant both digital skills as well as disciplinary skills.

As a number of scholars have noted (Griffin and Care 2014; van Laar et al. 2017) there is a strong overlap between key generic attributes, defined as "21st century skills," and digital literacies. By focusing on how students learn in digital networks a range of other skills come into play such as collaboration, communication, ethical decision-making and citizenship. Even personal skillsets such as self-management are now highly digitally mediated through personal organization software, as is teamwork and collaboration. So, a learning focus on a broad set of digital literacies will also deliver on key graduate attributes which underpin the contemporary university's promise to students around transferable employability skills that govern lifelong learning. Equally other graduate attributes must now be taught within a digital context.

Authentic and sustainable assessment in this context must therefore be both authentically digital as well as authentically connected to contemporary practices of work and citizenship. I will argue that "design thinking" provides a framework for thinking through sustainable assessment (Boud and Soler 2016) in a digital context. Such a framework seeks to move from the textual/compositional bias which governs much assessment practice in higher education towards genuine digital assessments which engage students as designers and assess their digital design literacies as well as their critical encounters with authentic discipline knowledge.

Gunther Kress and colleagues have suggested that in a contemporary multimodal environment, "design" is now the key mode for negotiating texts/signs.

> The shift, conceptually, from composition to design mirrors a social shift from competence in a specific practice conceived in terms of understanding of and adherence to convention governing the use of a mode—writing, say—to a focus on the interest and agency of the designer in the making of signs-as-texts. Design is the practice where modes, media, frames, and sites of display on the one hand, and rhetorical purposes, the designer's interests, and the characteristics of the audience on the other are brought into coherence with each other. (Bezemer and Kress 2008: p. 174)

Although Kress and others rightly point to the emergence of multimodality as a defining element of new media environments and contemporary digital experience (and, as I noted at the beginning of this paper, all contemporary learning is in some way a digital experience), writing remains the dominant mode of assessment in higher education (Tuck 2018, p. 3). And composition rather than design remains its central metaphor.[1] Even in well-structured digital assessments that aim to develop collaborative elaboration of ideas, writing often remains the primary genre, and textual conventions often drive assessment criteria.

In a 2012 post, titled "A Better Blogging Assignment," on the popular *Chronicle of Higher Education* Prof Hacker blog, for example, Mark Sample, who describes himself as a blogging enthusiast who has used blogging extensively in his courses, outlines a range of ways that he structures blogging assignments to encourage lively debate and collaborative development of knowledge (Sample 2012). Yet in another post when he outlines his assessment criteria for blogging assignments the criteria are purely textual:

> *Exceptional.* The blog post is focused and coherently integrates examples with explanations or analysis. The post demonstrates awareness of its own limitations or implications, and it considers multiple perspectives when appropriate. The entry reflects in-depth engagement with the topic. (Sample 2010)

Similarly, a recent publication from the UNSW Teaching and Learning Centre (Teaching UNSW Sydney 2017) frames the benefits of blogging in terms of collaborative learning and building a community of practice but their sample rubric covers only logical sequence of posts, content knowledge, grammar, writing style, referencing and structure of ideas.

These are examples of assessments that are digitally *delivered* rather than digitally *designed*. They remain focused on the compositional practice of traditional writing even though this writing has been transferred to a digital context. They may assess the generic skills of critical thinking and perspective taking but they miss the opportunity to engage with an integrated approach to digital design literacies which contextualise these skills to a contemporary digital environment. So how do we move beyond this compositional bias and towards authentically digital assessments which engage students as designers and assess their digital design literacies as well as their critical encounters with discipline knowledge? If we take the imperatives of

[1] This is more true of the humanities and social sciences than of the sciences or design disciplines like engineering but even though the sciences may rely less on essay style texts they are often even more dependent on assessment as knowledge-based recall whether through textual or computational notation.

design as described by Kress – as a process that brings a focus on both the design agency of the learner and their encounter with an audience or culture – as central to digital experience and digital literacy, this has important implications for all aspects of the way that we design and assess digital learning.

9.1 Students as Learning Designers, Design Thinking and Sustainable Assessment

Design has been an influential paradigm in curriculum development – with learning design a well-established discipline (Laurillard 2013), but it has only rarely been applied to the student experience of learning (Levinsen and Sørensen 2013). Sørensen and Levinsen (2014, 2015; Levinsen & Sørensen 2013) in a recent school-based project in Denmark have shown that students as learning designers is a powerful way to understand the complexities of digital learning. Students are both digital producers and learning designers they argue:

> As digital producers, the students acquire knowledge about a subject, whereas as learning designers, they define (sub)goals, select content and organize their learning process in relation to producing learning objects for other students. The processes run parallel in an ongoing complementary interaction, where the students continuously position themselves as either producers or learning designers. (Sørensen and Levinsen 2014, p. 61)

They argue that this is an agentic process for both teachers and students, with teachers designing learning experiences that empower students to design their own learning. The challenge is, they write, "to design a dynamic frame where the teacher and students continuously interact, construct and stage teaching and learning aimed at the students' construction of subject or interdisciplinary knowledge." (Sørensen and Levinsen 2014, p. 64) They also stress (Sørensen and Levinsen 2015) the importance of underpinning this process through the development of students' evaluative judgement (Tai et al. 2018). This is supported through a variety of feedback and feedforward processes from both teachers and peers and that include a "focus on students as actors and address their self-reflection" (Sørensen and Levinsen 2015, p. 293).

This joint staging of learning with dialogic feedback is essential to any notion of sustainable learning and assessment. As Boud and Soler (2016) note, sustainable teaching and learning practice demands "repositioning the notion of feedback not as an act of information giving to students, but as a co-productive process in which both students and others have key roles to play. Learning cannot be sustainable in any sense if it requires continuing information from teachers on students' work." (p. 403) The ultimate aim of engaging in a dialogic feedback process is to ensure that students become confident and relatively accurate assessors of their own and their colleagues' work – an ongoing skill they will require as autonomous professionals and life-long learners. Boud and Soler have proposed that the most important elements for students in developing these skills are:

1. identifying oneself as an active learner;
2. identifying one's own level of knowledge and the gaps in this;
3. practising testing and judging;
4. developing these skills over time; and
5. embodying reflexivity and commitment. (Boud and Soler 2016, p. 402).

This type of iterative reflexive practice is very close to the paradigm that has come to be known as "design thinking" which arose out of studies of the way designers think and create (Cross 2006). It has now been adopted across multiple disciplines and has evolved and been defined in a set of different but complementary ways (Johansson-Sköldberg et al. 2013; Carlgren et al. 2016). Several common ideas emerge from these definitions which emphasize:

- an open-ended enquiry approach to problem solving
- beginning with detailed observation and analysis, which often includes reframing the problem
- adopting a user centric approach characterized by empathy and a diversity of perspectives
- iterating different solutions, through physical and cognitive prototyping and reframing
- engaging in evaluative reflection-in-action
- the materiality of the process and the importance of tacit feedback through building models, visualizing problems, thinking through doing and a spirit of creative playfulness.

Design thinking shares a number of characteristics with paradigms already familiar to educators such as Kolb's experiential learning cycles (Kolb 2014), action research (Laurillard 2008) and Schön's model of reflection-in-action (Schön 1987). Each of these models presents a cyclical model of problem solving which is iterative, developmental and reflective. However, design thinking's focus on user-centred, empathic design, drawing on the tacit feedback from the materiality of prototyping, brings a new emphasis to this discussion.

So how do the intersecting paradigms of design thinking, learners as designers, twenty-first century digital literacy and sustainable assessment enable us to think more clearly and creatively about developing genuine digital assessments and learning activities?

9.2 Productive Design Literacies and Sustainable Digital Assessments

Design thinking, learners as designers, critical digital literacy and sustainable assessment each focus on a process-oriented approach to collaborative problem-solving, they make critical and evaluative processes explicit as an essential part of a holistic, lifelong, sustainable learning process.

Moving from a *compositional* metaphor to a *design* metaphor as we conceptualize new digital learning and assessment activities produces a shift in thinking which highlights the dialogic practices at the heart of sustainable learning and assessment. Instead of conceptualizing standards in terms of a set of compositional conventions, design thinking invites us to focus on reflective and collaborative processes that underpin standards as a practice of iteration and evaluation.

There are many implications for assessment practice if we adopt the digital design paradigm. For example, in a design thinking assessment paradigm, students would *present* assessments not *submit* them, and this simple change in language immediately conceptualizes assessment as a collaborative one-to-many activity rather than as an isolated one-to-one activity. Several principles can be drawn from looking at digital learning and assessment through a digital design lens (see Table 9.1).

Although I am arguing that there is a need to move from a reliance on a compositional paradigm, which still governs a large proportion of assessment in higher education, I am not arguing that it needs to be completely abandoned. There will, for example, still be a need to teach and assess business students' ability to produce standard format business reports that follow an agreed set of explicit textual

Table 9.1 Digital learning and assessment through a digital design lens

A digital design focused assessment is produced with a specific audience/ user in mind	Digital platforms and design process are always audience/user focused and design based digital assessments will always be contextualized as having a particular audience and/or end user(s), this provides an authentic professional context or real world orientation for final presentation. This audience focus not only influences the style of communication, how can I explain this disciplinary concept to a general audience, for example; it can also prompt questions about rhetorical purpose. How can this be of use to my audience? How might this be of use to them in their lives?
A digital design focused assessment is collaborative and encourages empathic perspective taking	Digital assessment is collaborative because it is situated within networks, this does not mean that it is always group-work, but it is contextualized in the disciplinary community of practice, it is designed for an audience, it is reviewed, or open for review, by fellow scholars and this collaborative framework means it is done from an empathic perspective which entails thinking from within multiple alternative points of view.
A digital design focused assessment is multimodal	When assessments are presented rather than submitted they are produced with the most fitting combination of media, and the affordances of different types of media are explored to enable the practice of different types of knowledge engagement and presentation.
A digital design focused assessment is iterative and is negotiated with a range of stakeholders	Even where it involves the production and presentation of summative artefacts a digital design assessment remains focused on a clearly designed iterative process in which both the teachers and students are co-designers and are encouraged to map the multiple stages of this design process.
A digital design focused assessment is reflective	Any collaborative, iterative process demands multiple points of reflection-in-action and these are structured in a digital learning context to scaffold and develop a learner's evaluative judgement.

conventions and are authentic to professional practice. However, even in examples like this, the ability to more easily produce infographics through online generators and include hyperlinks in interactive PDFs can add new digital design dimensions to these traditional formats. Even within authentically digital assignments there will be important compositional elements and digital conventions that will form part of this practice. Good online writing, for example, follows a set of conventions governed by what we know about how people read on screen: they scan for key points. This makes headers, dot points and visual conventions that break up the text for the scanning-eye particularly important.

9.2.1 What Does This Mean in Practice?

Earlier in this paper I pointed to a number of examples of student assignments which, although they employed multimodal digital platforms such as blogging software, applied only textual or compositional criteria in scaffolding and assessing the task. In the next section I will look at blogging as a case study to explore how a digital design paradigm can be applied to produce an authentically digital assessment.

9.3 Blogging as a Design Oriented Sustainable Digital Assessment

Although blogging has been used extensively in higher education (Sim and Hew 2010) as a class management and interaction tool and as a mode of assessment, as a number of authors have pointed out (O'Donnell 2006; Kim 2008; Deng and Yuen 2011) there has been little attempt to conceptualise the particular affordances of blogging as a mode of sustainable assessment. Blogging assessments are most often defined at the postgraduate level and usually as a primarily written assessment rather than as an authentic design project which involves the creation of a sophisticated, multimodal, digital artefact.

I have used blogging as a learning and assessment platform across a number of different units in media theory, journalism (O'Donnell 2006) and business disciplines, and I have worked with other academics to implement blogging in a variety of other disciplines. As I will show in the exemplars below (see boxed assessment descriptions) blogging is easily adaptable to different professional contexts and different learning outcomes ranging from the production-oriented needs of journalism to the more identity focused needs of postgraduate early-career researchers in business studies.

Exemplar 1: Blogging Assignment for Postgraduate Business HDR Students

This blogging assignment was developed for a Research Dissemination unit taught as part of the compulsory coursework for a Doctorate of Business Administration program. Students establish a research blog and make weekly posts outlining aspects of their current research and aggregate news from their research area. The blog should also include general information about their expertise as a researcher and professional. Ten posts are completed over the course of the semester. Students were encouraged to use the multimedia affordances of blogging and this was linked to a second assignment on visual presentation skills.

Assessment criteria.

Students will demonstrate:

- an ability to effectively use the affordances of blogging technology (including use of hyperlinks, blogroll, categories, themes)
- an ability to communicate research ideas in an engaging, direct manner
- an ability to effectively aggregate a range of news from their research area
- an evolving sense of a personal brand as a researcher.

Students are encouraged to comment on each other's blogs and are given an opportunity to present their blog to class and discuss opportunities and challenges that they are currently having. This encourages a collaborative peer led development, without formally assessing this component.

As blogging pioneer Rebecca Blood (2000) has described, in an essay about the historic evolution of blogging, the "original weblogs were link-driven sites". They were driven by the blogger's personal sense of what was interesting in the wilds of the early web. It is both this *linking practice* and experimenting with *personal voice* that was critical in Blood's own experience of blogging and she describes how this had profound effects on her own self-awareness and confidence. However, as blogs have become more mainstream and are increasingly used for commercial marketing, mainstream media and institutional branding, as well as personal communication, it has become more difficult to define their distinctive place in the digital ecosystem. Garden (2012) in her analysis of the literature concluded that the proliferation of blogs makes any definition inherently problematic and underpinning this difficulty is the question: is blogging a communication genre or a technological platform?

At first this distinction between a platform – a technology with certain affordances – and a genre – a communication type with a particular context and audience – may seem like a critical distinction, especially in an academic setting. Indeed, in terms of scaffolding individual aspects of a blogging task for students, it will be necessary to engage with both the specific technological affordances of blogs as well as specific types of public disciplinary communication. However, in terms of

developing sustainable digital literacies, it is also necessary to explore the way genre and technology are necessarily imbricated in a digital context.

In an influential article which used a large content analysis of over 200 blogs, Herring et al. (2004) defined blogging as a "bridging genre" which brought online writing into contact with a range of associated historic writing practices – journaling for example. In this sense it is both a new genre and one which participates in a range of other familiar genres of academic and popular communication. Thinking explicitly about blogging as a way of bridging different types of communication opens up a set of rich possibilities to engage students in exercises that invite them to actively design and construct complex digital communication artefacts as well as reflect on that process of production.

Exemplar 2: Blogging for Introductory Journalism Students
In an introduction to journalism course we used blogs as a portfolio space that allowed students to explore a developing sense of themselves as both a student and a journalist. It was used across a series of linked assignments that combined the development of the blog with the use of mobile technology. Together these assessments aimed to introduce students to:

- basic journalistic tasks such as interviewing, filming, and editing, and
- their new environment as university students and emerging journalists
- the creation of project-based web publications through blogging.

There were two separately assessed assignments each of which was carefully scaffolded in stages with in-class activities and online resources.

Task 1: Students establish a blog and complete a number of blog posts which must include:

- 3 Individual classmate photo portraits with 100 word captions;
- 2 images of university life with 100 word captions;
- 2 video vox-pop posts interviewing students from across the university on a topic;
- 1 two minute edited video interview with a classmate.

Task 1 Assessment criteria: Students will demonstrate an ability to

- Use the basics of blogging technology as a storytelling medium;
- Take engaging photographs which convey information and emotion;
- Capture engaging content through video interviews;
- Complement visual storytelling with simple, well expressed written content.

(continued)

Task 2: Students produce a series of new posts on their blog which must include:

- 2 × 500 word profiles of classmates with a photographic portrait one of the interviews focusing on interests outside of media and journalism;
- 2 aggregated/curated news posts dealing with issues in journalism and storytelling (300 words each);
- 1 × 500 word news story which draws on original interviews (at least 4) with classmates on career aspirations and issues in journalism today the post must be accompanied by an original image and appropriate caption.

Task 2 Assessment criteria: Students will demonstrate an ability to:

- Use blogging platforms effectively as a storytelling medium;
- Create engaging and informative content through aggregation and curation of secondary sources;
- Create engaging and informative content through original interviews;
- Write effectively, following standard grammatical and journalism style conventions.

9.4 Blogging as a Technologically Scaffolded Genre

Deng and Yuen (2011) present one of the few studies which attempt to bring the task focus of their assessments together with an explicit focus on the technological affordances of blogging. Their model links six affordances of educational blogging with three types of blogging behaviour: writing enables self-expression and self-reflection; reading generates social connection and triggered reflection and commenting generates social interaction and reflective dialogue.

> In more concrete terms, writing blogs allows for individual-centered expressions of personal feelings and thoughts, and fosters self-reflection. From the perspective of the blog readers, the sense of connections and social rapport within a community can be cemented and the real stories captured in blogs can serve as food for reflection or a source of inspiration. On the community dimension, peer interactions through commenting provide an alternative venue for the exchange of the social support. (Deng and Yuen 2011, p. 450)

This is an important step forward in the literature because it begins to provide a framework for the how and why of blogging as a genre of academic communication. It focuses on blogging as an integrated practice of reading, writing and commenting and as the production of a digital artefact rather than merely a series of posts. In this sense it begins to provide a framework for the way particular forms of – for example – reflective writing might emerge as a distinctive genre in a blogging format. Even though it begins with a focus on writing it moves from a purely compositional focus to a focus on the way writing is delivered and received as a communication in a digital ecosystem. However, I would suggest that further emphasis on some of the technological specificity of blogging would add further

scaffolding, necessary for effective sustainable assessment. Being explicit about these affordances would help reveal to students a set of transferable design elements of blogging. These technical elements of blogging would include:

- hyperlinking – which encourages curatorship and source acknowledgement and embeds scholarly writing within a disciplinary community of practice
- categories and tagging – which introduces taxonomic organization of information and allows for the emergence of connected thinking
- reverse chronological posting – encourages regularity of posting and the iterative development of ideas
- comments – facilitates feedback and peer interaction
- blogrolls of associated or admired bloggers/websites – facilitates the development of a network
- integration with other social media services – increases sharability
- multimedia embedding with captioning and sourcing – develops multimodal literacies.

Taken together these mechanisms begin to describe and scaffold particular types of writing (and other types of multimodal communication), reading and commenting. They could be used as both aids and triggers for particular types of reflection and scholarly communication, and they scaffold a particular type of knowledge management, organization and exchange.

Scaffolding Blog Post Types

In working with students on blogging assignments I introduce blogging as a flexible medium which allows for a number of different styles of academic communication through this simple typology of posts:

- **Comment post**: is a post which begins with a link to a media report, video or another blog and then provides comment and analysis on the issue.
- **Comparative post:** is a longer version of the comment post which links together several reports around a similar theme and provides some commentary.
- **Bookmark post:** is a short post with a link and block quote to a report or other blog entry with little or no commentary. These are a way of adding to threads that you have already begun or a way of beginning something which you can come back to in more detail when you have more time.
- **Reflective post:** sums up some of your own thinking about a particular issue and will probably be used to bring together some ideas that have been explored in previous posts. In this case the links may primarily be to your own previous posts.
- **Descriptive post:** a short description of something you have noticed or an event that has happened to you that has relevance to the topic being covered. This could be enhanced by an image or video.

This rich form of multilayered communication can only be realized if blogging assignments explicitly engage with, and scaffold, the full set of generic affordances that are available through the digital platform. But as Peter Goodyear and colleagues remind us "an affordance of a thing for a person depends on qualities of the thing relative to capabilities of the person (skills, perceptual acuity, etc)" (Goodyear et al. 2014, p. 138). So, looking at particular technology affordances in the abstract is not enough to design a productive, sustainable, authentically digitally assessment. Design thinking assists us in providing a process to scaffold and harness these affordances by providing a framework for students and teaching teams to actively design learning opportunities drawing on these affordances, but doing so in a capabilities building framework.

In the examples given above I invite students to solve an integrated set of design problems:

- a specific communication problem – how to tell the stories of first year university students and their encounters with the institution or how to build a personal brand as an emerging business studies researcher in a particular area;
- how to relate this to their disciplinary context;
- how to work with specific digital processes: linking, digital photography, tagging, commenting etc. to solve these design problems and make these contextual links;
- how to bring all of this together into a cohesive digital artefact.

This is done in the context of an active learning classroom which is task-discovery focused. Students are supplied with a range of resources prior to class and they are then asked to create artefacts, often in groups, during class time. The iterative emergent nature of design thinking is emphasised through a discovery/task/presentation model that builds collaborative learning. Importantly students are explicitly introduced to the assessment criteria that openly engage the digital as well as disciplinary context of the tasks. In the journalism assignment described above, for example, a collectively developed rubric and a set of checklists were developed by the class[2] which scaffold the production of the student's blog. This enables them to engage in self-evaluation of the multiple components of the design/communication process.

An authentically digital, effectively scaffolded blogging assignment requires students to engage in design at the macro and micro level. At a micro level developing a blog is built up through a series of related tasks – both the combination of different individual posts as well as the deployment of the design and taxonomic elements of blogging. This exposes students to an iterative process that encourages them to actively make connections and design their own learning. At a macro or project level developing a blog is an exercise in design thinking which addresses a particular design problem: what combination of elements effectively communicates about a selected topic to a nominated target audience in the most effective way. The typology of posts and other assignment scaffolds emphasize the multiple modes and iterative development of this type of networked scholarly communication instead of

[2] This class was often 150+ students divided over 5–6 seminar groups so the negotiation of this collective rubric was another digitally mediated process done with simple Google Docs technology.

prescribing a single set of conventions that defines a compositional approach to writing. This encourages students to actively think through their design choices and design their learning experience.

The assessments described above are designed to explicitly scaffold each of the digital design principles outlined earlier in this paper:

Audience/user focus: Both the journalism assignment and the business studies assignment ask students to explicitly address real audiences.

Collaborative/empathic: The journalism assignment asks students as the designers to put themselves in another's place, to communicate the voice of another. The business studies assignment emphasises the linked nature of blogging and also asks students as designers to design their communication in the context of the wider communicative ecosystem: linking to others who write in similar areas and commenting on fellow bloggers work. Feedback and interactions from fellow class members is built into the structure of the assessments.

Multimodal: Students are explicitly asked to address blogging as a multimodal medium and as visual design problem. This invites students as designers to consciously think about communication as a multimodal process and reflect on the ways that different elements of communication work and interact.

Iterative/negotiated: the reverse chronological structure of blog posts which slowly add to the emergence of a rich digital artefact, creates a naturally iterative approach to designing blog-based communication.

Reflective: while blogging is not inherently reflective, the type of processes I have just described do offer multiple prompts for reflection and encourage an iterative and self-consciously constructed approach to communication.

Situating blogging as an integrated design problem which addresses both communication design – what, to whom, and in what way – as well as artefact design – how can the form best address these communication needs – invites students to address holistically the critical and the productive aspects of digital design literacy.

9.5 Conclusion

The research on assessment tells us that more emphasis needs to be placed on the process rather than just the outcome of assessment tasks. Sustainable assessment is dependent on a dialogic process that encourages confident self and peer assessment. Design thinking gives us a framework for thinking through these processes and digital ways of working provide a set of tools. Drawing on frameworks such as this, tools like blogging platforms can be used in higher education to transform assessment. They can enable particular types of personal, interactive, multimodal, reflective communication that is presented in an organized coherent way as part of the development and presentation of a student's digital identity. In this way digital assignments can move from merely replicating traditional textual compositional practices to a productive assessment *of* and *as* digital practice.

References

Aoun, J. E. (2017). *Robot-proof: Higher education in the age of artificial intelligence*. Cambridge, MA: MIT Press.

Bayne, S. (2015). What's the matter with 'technology-enhanced learning'? *Learning. Media and Technology, 40*(1), 5–20.

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195.

Blood, R. (2000, September 7). Weblogs: A history and perspective. *Rebecca's Pocket*. Retrieved from http://www.rebeccablood.net/essays/weblog_history.html.

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413.

Carlgren, L., Rauth, I., & Elmquist, M. (2016). Framing design thinking: The concept in idea and enactment. *Creativity and Innovation Management, 25*(1), 38–57.

Cross, N. (2006). *Designerly ways of knowing*. London: Springer.

Deng, L., & Yuen, A. H. K. (2011). Towards a framework for educational affordances of blogs. *Computers & Education, 56*(2), 441–451.

Garden, M. (2012). Defining blog: A fool's errand or a necessary undertaking. *Journalism, 13*(4), 483–499.

Goodyear, P., Carvalho, L., & Dohn, N. B. (2014). Design for networked learning: Framing relations between participants' activities and the physical setting. In S. Bayne, C. Jones, M. de Laat, T. Ryberg, & C. Sinclair (Eds.), *Proceedings of the 9th international conference on networked learning 2014* (pp. 137–144). Edinburgh: Networked Learning Conference Committee.

Griffin, P., & Care, E. (2014). Project method. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st century skills: Methods and approach*. Dordrecht: Springer.

Herring, S. C., Schiedt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th annual Hawaii international conference on system sciences, 2004*. Hawaii: IEEE.

Johansson-Sköldberg, U., Woodilla, J., & Çetinkaya, M. (2013). Design thinking: Past, present and possible futures. *Creativity and Innovation Management, 22*(2), 121–146.

Kim, H. N. (2008). The phenomenon of blogs and theoretical model of blog use in educational contexts. *Computers & Education, 51*(3), 1342–1352.

Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: What is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology, 39*(1), 6–36.

Kirkwood, A., & Price, L. (2013a). Missing: Evidence of a scholarly approach to teaching and learning with technology in higher education. *Teaching in Higher Education, 18*(3), 327–337.

Kirkwood, A., & Price, L. (2013b). Examining some assumptions and limitations of research on the effects of emerging technologies for teaching and learning in higher education. *British Journal of Educational Technology, 44*(4), 536–543.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. London: FT Press.

Laurillard, D. (2008). The teacher as action researcher: Using technology to capture pedagogic form. *Studies in Higher Education, 33*(2), 139–154.

Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Hoboken: Routledge.

Levinsen, K., & Sørensen, B. H. (2013). Students as learning designers. In R. McBride & M. Searson (Eds.), *Proceedings of the Society for Information Technology & Teacher Education International Conference 2013* (pp. 5142–5149). Chesapeake: Association for the Advancement of Computing in Education.

Marsh, E. (2018). Understanding the effect of digital literacy on employees' digital workplace continuance intentions and individual performance. *International Journal of Digital Literacy and Digital Competence, 9*(2), 15–33.

O'Donnell, M. (2006). Blogging as pedagogic practice: Artefact and ecology. *Asia Pacific Media Educator, 1*(17), 5–19.
O'Donnell, M. (in press). Transformative digital pedagogy: Reimagining connection, collaboration and design, In R. Lawson, M. O'Donnell & E. Leinonen (Eds.), *Developing undergraduate curriculum*. London: Springer.
Oliver, M. (2013). Learning technology: Theorising the tools we study. *British Journal of Educational Technology, 44*(1), 31–43.
Price, L., & Kirkwood, A. (2014). Using technology for teaching and learning in higher education: A critical review of the role of evidence in informing practice. *Higher Education Research & Development, 33*(3), 549–564.
Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: Technologies, social memory and the performative nature of learning. *Journal of Computer Assisted Learning, 26*(1), 53–64.
Sample, M. (2012, July 3). *A Better Blogging Assignment [Blog post]*. Retrieved from https://www.chronicle.com/blogs/profhacker/a-better-blogging-assignment/41127.
Sample, M. (2010, August 14). *Pedagogy and the Class Blog [Blog post]*. Retrieved from http://www.samplereality.com/2009/08/14/pedagogy-and-the-class-blog/.
Schön, D. A. (1987). *Jossey-Bass higher education series. Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.
Sim, J. W. S., & Hew, K. F. (2010). The use of weblogs in higher education settings: A review of empirical research. *Educational Research Review, 5*(2), 151–163.
Sørensen, B. H., & Levinsen, K. T. (2014). Digital production and students as learning designers. *Designs for Learning, 7*(1), 54–73.
Sørensen, B. H., & Levinsen, K. T. (2015). Powerful practices in digital learning processes. *Electronic Journal of E-learning, 13*(4), 291–301.
Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481.
Teaching UNSW Sydney. (2017). *Assessment Toolkit: Assessing with Blogs*. Retrieved from https://teaching.unsw.edu.au/assessing-blogs
Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476.
Tuck, J. (2018). *Academics engaging with student writing: Working at the higher education Textface*. London: Routledge.
van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577–588.

Marcus O'Donnell is Director, Cloud Learning Futures at Deakin University in Melbourne Australia. He was academic lead for the Degrees at FutureLearn project a world first initiative to put a suite of degrees on an open social learning platform. His background is as a journalist, editor and journalism academic and his research interests include the intersections of contemporary transmedia storytelling, digital learning and learning design. He is also the lead educator on the FutureLearn course, Professional Resilience, and he has an ongoing research interest in empowering student resilience through contemplative approaches to higher education.

Part III
The Role of Big Data in Reimagining Assessment

Margaret Bearman

Big data and learning analytics are changing the face of higher education. This section of the book highlights their current and future roles in assessment and feedback. These four chapters describe a mix of 'what is happening now' with extrapolations of 'what might be possible'. The previous section explored how to expand the pedagogic potential of assessment in a digital world, while this section looks at the potential of the unfolding power of data and analytics for assessment. The authors discuss how big data – the quantified traces of processes that are an integral part of digitally mediated education – can change assessment and feedback. While assessment is mostly characterised as reflecting the status quo – these chapters also provide an indication of how big data by its very nature can change assessment practices.

There are two core concerns that are highlighted by this section and are discussed in further detail below. Firstly, the chapters analyse how 'big data' can be transformed into meaningful information that can inform assessment and assessment practices. Secondly, they provide some suggested ways in which educators might work with this data derived information to enhance assessment.

Transforming Trace Data into Meaningful Information

Big data contains many different kinds of information such as click data, time on task, grades, demographics and enrolment data. In some instances, this data is very easy to interpret; in other instances, it is very difficult to assign meaning. This presents a range of challenges. Milligan's chapter (Chap. 13) on metrolytic standards investigates how we might validly measure constructs using big data. However, as suggested by Pardo and Reimann (Chap. 12), in order to generate rich data, we need Technology Rich Environments (TREs) that might include data such as affective states. Yet another alternative means of interpretation is presented in Rogaten et al., whereby meaning is created a priori through assignment to categories; in this way meaning is assigned through the initial classification (Chap. 11).

How Data-Derived Information Influences Assessment and Feedback

At some point data-derived information, refined by algorithms or classification, must interface with educators and/or students as part of assessment and feedback processes. What value can this data-derived information bring? Rogaten et al. (Chap. 11) ably demonstrate how providing information about the interaction between different assessment types and student success can inform institutional understandings of assessment designs. Additionally, as Pardo and Reimann (Chap. 12) and Knight (Chap. 10) point out, analytics could help make assessment processes more efficient. However, one of the biggest areas of promise is in providing educators/students with information about students that is currently unavailable to them. This is speculative work, with authors suggesting that analytics will be valuable to trace 'soft skills' and 'affective states'. One outstanding question is the degree of automation underpinning this provision of information. Knight argues that in the short term, we should conceptualise this information as augmenting educator's feedback (and hence assessment) processes rather than providing a direct conduit to students (Chap. 10). This provides a pragmatic way forward in a field where making sense of data is one of the largest challenges.

The work in this section represents a subset of the overlap between assessment and educational analytics. Taken together, all four chapters represent both big data's promises and the challenges with respect to assessment in higher education.

Chapter 10
Augmenting Assessment with Learning Analytics

Simon Knight

Abstract Learning analytics as currently deployed has tended to consist of large-scale analyses of available learning process data to provide descriptive or predictive insight into behaviours. What is sometimes missing in this analysis is a connection to human-interpretable, actionable, diagnostic information. To gain traction, learning analytics researchers should work within existing good practice particularly in assessment, where high quality assessments are designed to provide both student and educator with diagnostic or formative feedback. Such a model keeps the human in the analytics design and implementation loop, by supporting student, peer, tutor, and instructor sense-making of assessment data, while adding value from computational analyses.

10.1 Introduction to Learning Analytics

Rising use of digital technology across education has heralded an increased focus on the potential of data to inform our understanding of student learning. Increasing attention to learning data has come from a number of sub-disciplines in education, most recently the emergence of 'Learning Analytics', a field with specific focus on the use of data derived from student learning to inform that learning. As Ferguson's (2012) "The State of Learning Analytics in 2012: A Review and Future Challenges" charts, the developmental trajectory of learning analytics has been driven by an interest in applying business intelligence techniques to the increasing amounts of data made available through virtual learning environments and other learning systems. This interest has shifted from primary concerns around accountability and efficiencies, to an increasing focus on pedagogic concerns.

The field of learning analytics has seen keen interest in understanding how to effectively implement novel techniques, which support learning. Across educational stakeholders there is an increasing desire to use data effectively to inform and

S. Knight (✉)
Faculty of Transdisciplinary Innovation, University of Technology Sydney, Ultimo, NSW, Australia
e-mail: Simon.Knight@uts.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_10

understand practice; in particular, there is a strong desire to achieve impact through implementing and supporting effective learning strategies. Learning analytics can contribute to this, but to do so, we need to develop a deeper understanding of the kinds of problems it can tackle, and how to integrate analytics into practical pedagogic contexts.

There is a growing expectation that educators use data as a form of evidence of student learning (and course evaluation). This shift in focus onto learning analytics as a form of assessment data highlights two intertwined concerns. First, as others have noted, educators must have a degree of *data literacy* to be able to navigate the, often quantitative, information that they are provided with (see, for examples, Mandinach 2012; Mandinach and Gummer 2016). Second, this data literacy must have a focus on how data is used *in context* to make decisions. Such a perspective prompts a move beyond simply supporting educators in navigating large-scale assessment data or predictive models based on Learning Management System (LMS) log data, to supporting educators in making practical decisions about the data they collect and make decisions on.

This chapter focuses in on a particular component of those challenges, suggesting that a key path for the impact of learning analytics is a focus on 'intelligence augmentation' (IA) over artificial intelligence (AI).[1] That is, rather than focusing on how artificial intelligence might be deployed for automation of tasks, I shift focus to explore how artificial intelligence can augment assessment, in particular by amplifying the impact of high quality assessment with learning analytics derived feedback. Such an approach has dual benefits, in 'bringing along' educators in the implementation of learning analytics, though supporting their existing practice rather than requiring wholesale changes, and provides sites for learning analytics in which there is clear pedagogic potential through understanding that existing practice.

In this chapter, I introduce a broad model of assessment as a fundamental design context for educators. I then present three ways in which learning analytics might intersect with this model, arguing that 'augmenting assessment with learning analytics' provides for particular benefits. I will then illustrate this approach using a particular assessment context (peer assessment), and various sites for augmentation in that context, drawing on examples from the literature to do so.

10.1.1 Understanding Assessment as Design

In this chapter I frame assessment building on the social constructivist model of assessment processes in Rust et al. (2005). I take it that when educators set about designing assessments, their fundamental question is "How do we get knowledge of

[1] For an introduction to this idea, see https://theconversation.com/artificial-intelligence-holds-great-potential-for-both-students-and-teachers-but-only-if-used-wisely-81024

what the students know?" (Committee on the Foundations of Assessment 2001). A key part of this assessment process is the bringing to alignment of student and educator expectations around what is to be learnt, with active engagement with criteria development and feedback from both educators and students (Rust et al. 2005).

In a simplified model based on this perspective (Fig. 10.1), educators across educational levels engage in assessment design as a component of their course design. In so doing, they design assessment tasks and criteria by which to assess the completion of those tasks (the top box in Fig. 10.1). In order to undertake this design work, educators should have a rich understanding of the learning context alongside the assessment literacy to design and deploy tasks to probe learning. They must, therefore, understand how the assessment is related to the students learning, how assessments are constructed (by themselves or others) as measurement tools (e.g., validity, feedback, etc.), and knowledge of the range of assessment types and responses to them (Price et al. 2012). At most basic, educators might consider whether the assessment is summative in nature (sometimes called 'assessment *of* learning') – end of unit examinations, for example – or is intended to provide formative feedback towards further learning (sometimes called 'assessment *for* learning').

Influencing this process are characteristics such as the educator's epistemic cognition (how they think about the student's knowledge and the influence of that on their instruction) (Barnes et al. 2017; Fives et al. 2017). That is, when educators create assessments they are making judgements about what they would like their students to attain, and must consider the question "what will I learn about my students from the formative assessment event?" (Fives et al. 2017, p. 3). Even in a simplified model such as the one below, we hope that educators will also engage in a reflective process of using the students' outcomes in their assessments to (a) drive instruction, and (b) revise the assessment tasks for future iterations. That is, that the assignments students submit are a learning opportunity to develop courses and modules, as well as to develop the students' individual learning.

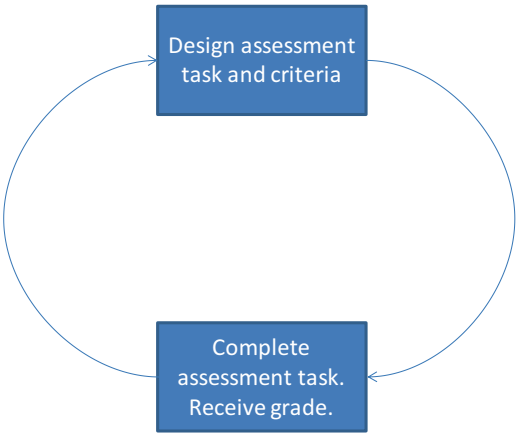


Fig. 10.1 A simplified assessment model, in which educators design assessments that students complete

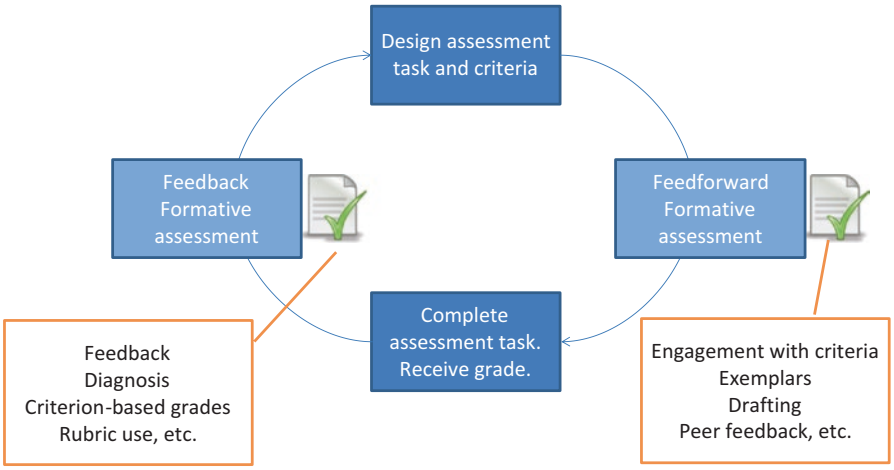


Fig. 10.2 Assessment with feedforward and feedback

As a part of this process of assessment – as reflected in the lower box in Fig. 10.1 – students of course must complete the assessment tasks, and receive marks on them against the criteria. Again, here, in addition to the content knowledge, students must engage a degree of assessment literacy, to understand what is being asked of them and how best to display this (Price et al. 2012). On receipt of their feedback (perhaps in written form alongside grades or rubrics), they should also engage in a reflective/metacognitive cycle, again influenced by individual differences such as achievement goals, and epistemic cognition (Muis and Franco 2009; O'Donovan 2017).

To extend this simplified model, increased attention has been paid to the implementation of feedback and feedforward cycles in assessment processes. Feedforward is formative feedback that supports students in understanding the assessments they will complete and their criteria, through engagement with criteria (perhaps even writing their own) and exemplars, and so on (Wimshurst and Manning 2013). Feedback, then, is formative feedback provided post-assessment activity to support students in understanding why they received the mark that they did, and how they might improve in the future. Thus, Fig. 10.2 indicates a cycle of assessment with feedback and feedforward; breaking down possible sites for augmentation with learning analytics.

10.1.2 Models for Transforming Assessment with Learning Analytics

One way of conceiving of learning analytics and its role in teaching and learning is as a tool for assessment (Knight et al. 2014). In this model, through the analysis of process-based trace data and artefacts created in learning tasks, researchers and

other stakeholders aim to make claims about that learning, at any one of the four components above. For example, we might provide feedback on a written essay (artefact), or the writing process (process), and this feedback might come before or after the final submission, provided to the student, or to support the educator in evaluating and revising their assessment design. In this understanding of learning analytics, we can further conceive of two broad approaches.

One potential is for learning analytics to facilitate a shift away from the summative assessment of artefacts produced. Instead, learning analytics might facilitate more process-based assessments such as choice based assessments, that explore the meaningful choices that students make in completing tasks (Schwartz and Arena 2013), and the processes undertaken in 'performance assessments' of authentic tasks (Benjamin et al. 2009; Darling-Hammond and Adamson 2010; Linn et al. 1991; Pellegrino 2013; Stecher 2010). New forms of feedback can thus make use of the trace data in "data-rich" learning (Pardo 2017). However, work adopting this approach requires the collection of new types of data, potentially using new technologies to collect that data, to be represented back to learners and educators in ways that also require research. The development of assessments based on novel process-based data is challenging, as work on intelligent tutoring systems has shown.

Thus, this development is likely to be time-consuming, expensive, and require systemic changes. As Roll and Wylie (2016) note in the context of the 25th anniversary of the Journal of Artificial Intelligence in Education, in order to achieve impact in education, researchers should undertake two strands of research: evolutionary, with a focus on existing classroom practice; and revolutionary, with a focus on more wholesale change of systems. This former, *evolutionary* approach, then, looks at how learning analytics can augment existing systems. In contrast the latter might involve development of new types of assessment, and the automation of existing assessments to create structural change in systems. These three approaches will be discussed, followed by a focus on the potential of *augmenting* in the rest of this chapter.

10.1.2.1 Approach 1: Learning Analytics for New Types of Assessment

Learning Analytics for New Types of Assessment First, learning analytics can be developed to assess constructs that were not readily assessable using traditional methods. This is including approaches such as intelligent tutoring systems, or choice based and performance assessments, that follow procedures such as evidence centred design (Mislevy et al. 2012) which provide a logic for mapping behaviours to target constructs (such as self-regulation, collaboration, etc.). In this model (see Fig. 10.3), whole assessment structures can be redesigned, with criteria based on fine-grain analysis of the knowledge components in a domain, and feedback and task completion based around completing defined activities with practice algorithmically oriented to the target constructs.

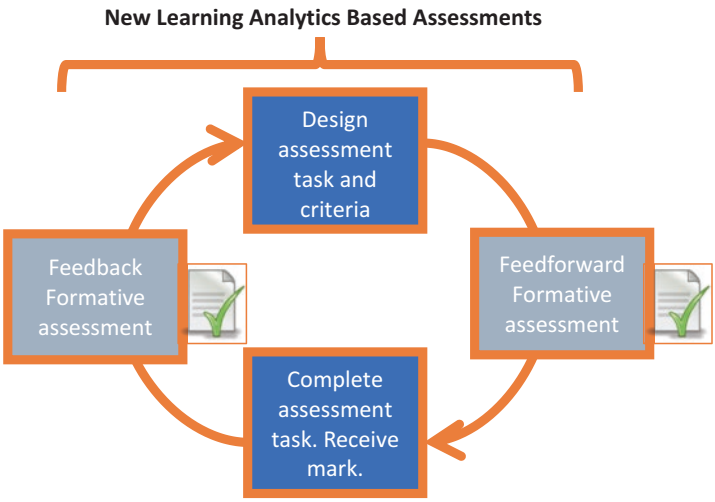


Fig. 10.3 New learning analytics based assessments

10.1.2.2 Approach 2: Learning Analytics to Automate Assessments

Learning Analytics to Automate Assessment Second, learning analytics can be used to automatize existing assessment structures. Thus, rather than tutors or instructors engaging in grading, a system is developed to automatically assess, for example through the automated essay scoring systems (Shermis and Burstein 2013). In this model (Fig. 10.4), substantial portions of the assessment structure may remain static, with analytics targeted either at the automation of existing work (e.g., automated essay scoring), or at scoring based on process data collected in the creation of artefacts aligned with existing assessments.

10.1.2.3 Approach 3: Learning Analytics to Augment Assessment

Learning Analytics to Augment Assessment Lastly – and the focus of this chapter – learning analytics can be used to augment assessment, to support the analysis and feedback on existing produced artefacts or process in the context of well-established effective e-assessment or technology enhanced assessment. That is, rather than focusing on developing new assessments or methods of completing assessment, instead focus on augmenting existing assessment structures through the augmentation of feedback and feedforward processes for effective assessment. In this model (Fig. 10.5), rather than seeking to automate grading (or enrich it through analysis of process data), the focus is on the formative feedback and the potential of learning analytics in that space. As highlighted below, this approach has some distinctive benefits.

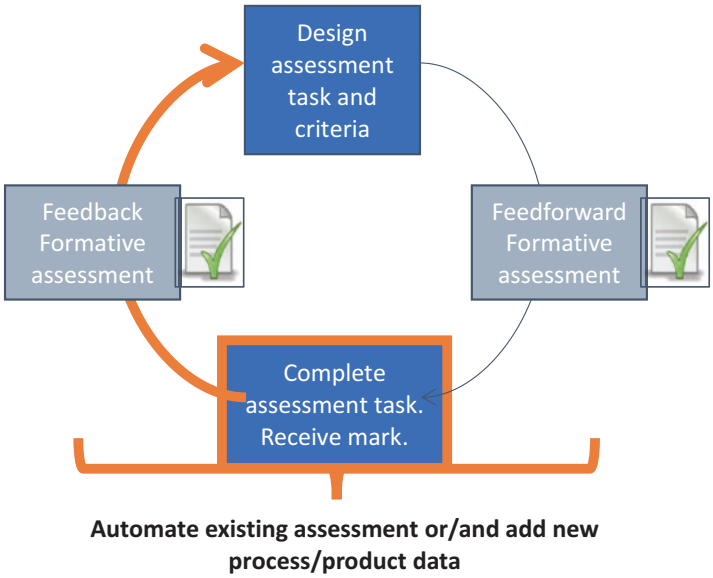


Fig. 10.4 Automate existing assessment or/and add new process/product data

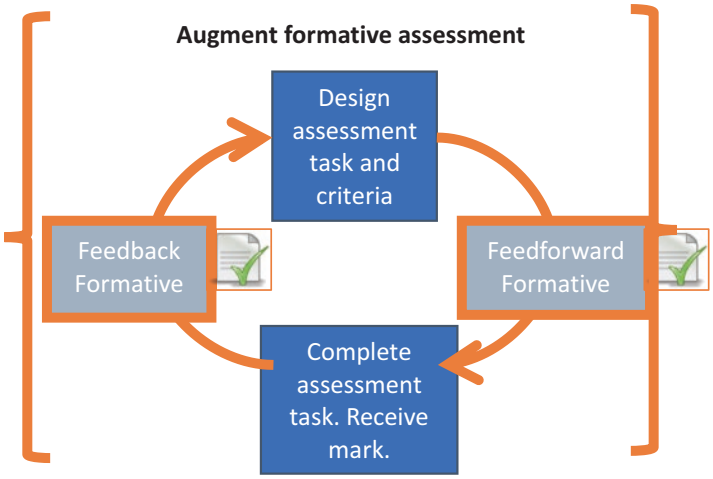


Fig. 10.5 Augment formative assessment

This chapter argues that to achieve maximum impact and adoption, educators and learning analytics researchers should work together to develop approaches to assessment augmented by learning analytics. Such augmentation has a dual effect of increasing adoption of learning analytic approaches, thus potentially opening the door to more revolutionary (than evolutionary) changes, while also increasing adoption of existing good practice through the support of that practice via learning analytics.

10.1.3 Benefits of Augmenting Assessment with Learning Analytics

Writing in the 25th anniversary issue of the International Journal of Artificial Intelligence in Education (IJAIED), Baker (2016) suggests that AI researchers working in the education space should shift focus. In that piece, Baker notes that adoption of AI technologies in education has not been widespread. He thus suggests that, instead of developing ever smarter intelligent tutoring systems, researchers might instead focus on amplifying human intelligence, supporting humans to make decisions through the use of intelligently designed systems. Such a proposal would, for example, shift attention away from auto-scoring systems, and towards using the same data to report to educators and students in intelligently, to support them in making decisions.

Baker notes a number of potential advantages to such an approach, including that it provides for more flexibility of response and intervention. This is because such an approach does not necessitate the assumption that students in the future will have the same behaviours as those modelled now, nor that the interventions remain the same. That is, a risk of automating existing processes is that it reifies existing practice in both assessment and outcome, where this may not be appropriate. An augmented approach, then, does not – necessarily – require the large-scale design and implementation of pedagogic agents, which is an expensive and time consuming process. Instead, the focus is on how technologies can be used flexibly to augment intelligence.

As Zhao et al. (2002) highlight, the distance of any innovation from existing: culture, practice, and technological resources, will impact its uptake by educators. That is, innovations will not be taken up that: are counter to existing cultural context; do not align well with practices of individual educators; or require significant change to available technology. In a similar vein, in their introduction to a panel discussion on overcoming barriers to achieve adoption of learning analytics, Ferguson et al. (2014) note the need to consider:

- institutional context and culture;
- buy-in from stakeholders; and
- understanding of specificity and user needs.

Distinct advantages of focusing on the potential of learning analytics to augment assessment are that such an approach should: align better with existing culture and practice; be less likely to require significant technological change; and keep decision making firmly within the autonomy of the educator. As such, this approach is a way to support and enhance existing practice, raising awareness of the potential of learning analytics, in turn increasing the potential for impact from those analytics in existing and novel applications. Although over a longer term a focus on systems change – and implementation of new assessment structures – is likely to be an

important component of educational improvement, I suggest that this is more likely – not less – with a key focus on augmentation of assessment with learning analytics. As such three key advantages of augmentation emerge:

1. Augmentation provides for easier integration
2. Augmentation is more flexible in use
3. Augmentation has lower upfront cost.

In proposing this approach, an explicitly *design* oriented perspective is taken. As Lockyer et al. (2013) highlight, because learning analytics provides new methods for data capture (in place of self-report methods), it can help educators in testing their assumptions regarding their learning design. Learning analytics, then, offers the potential to deliver on the promise of technology enabled assessment for learning (Crisp et al. 2016; Dawson and Henderson 2017), with feedback to students identified as one key target for such analytics (Timmis et al. 2016). By augmenting assessments with learning analytics, educators and students can receive a richer picture of what learning is taking place, and be supported in both developing the learning tasks and their responses towards those tasks.

10.1.4 Augmenting Assessment with Learning Analytics: A Worked Example

In this model, then, developing approaches in learning analytics would be used to provide feedback to students and educators for their educational decision making. This feedback might be diagnostic or formative in nature, providing information on specific areas of weakness and providing information on how particular features of their learning might be changed. Such augmentation might also simply provide additional information to stakeholders that they could use to support their decision making. For example, topic modelling might indicate that students wrote about two themes – a piece of non-normative or evaluative information – which could be used to target specific feedback content.

In the following sections some key advantages of this approach are drawn out through their exemplification in a particular set of design cases. These designs are based on a basic model of peer assessment, which may be used in both feedback and feedforward contexts. Each of these designs has in fact been implemented as indicated by citations throughout, although the combined set of design patterns has – to the best of my knowledge – not been brought together. In highlighting these designs, I wish to draw attention not only to the specific uses being developed, but also to the general approach to augmentation and its potential to improve the adoption and integration of data informed approaches in education.

10.1.4.1 Peer Assessment

In peer assessment models, students engage in assessing their peers' works for formative or summative purposes, while their own work is assessed by those same peers. The benefits of peer assessment for learning are well documented (for example, Topping 1998), with increased attention on specific design configurations to support its effective implementation (Strijbos and Sluijsmans 2010). In addition, peer assessment can produce reliable grade-feedback (for example, Cho et al. 2006b), and has clear potential in e-learning contexts (Whitelock 2010).

However, amongst both students and educators, there is often a focus on the potential of peer assessments to reduce the staff marking burden, and other features that relate to structural concerns around the fairness, efficacy, and quality of peer assessment. Students can have a perception that peer assessment is inaccurate or unreliable, thus reducing their motivation to participate, in addition to which students (like other assessors) do in fact disagree. To address these concerns, multiple designs can be adopted that augment a basic peer assessment design such as that shown in Design 10.1. These designs may be further augmented by learning analytics, as indicated in the following iterations.

Design 10.1: Peer Assessment

Problem: We want our students to develop their assessment literacy through applying the assessment criteria, and providing and receiving feedback on their work for formative purposes.

Task: Peer assessment involves students assessing each other's work, typically prior to submitting a final version of the same work.

Tools/materials and participant structures: Peer assessment is typically conducted individually, often with students asked to assess multiple assignments (although this requirement may be removed for formative purpose). Assessment is typically anonymous, and involves provision of both a score and comment.

Iterations and Augmentation: Peer assessment may be used as a stand-alone assessment design, or augmented by one of the designs below.

Developing the basic peer assessment design, we can add a number of *complementary* designs, that extend peer assessment, and may be augmented by learning analytics. For example, in Design 10.2, we see the addition of automated peer-allocation. This simple technique – based on prior assessment data, automated essay scoring, or topic modelling – can be used to allocate peers in a way to support all students learning. For example, prior research indicates that all students can benefit

Design 10.2: Peer Allocation

Problem: Peers may be assigned texts that are not aligned with their content knowledge, or that fail to support their learning (because they are misaligned in terms of quality).

Task: This design *complements* the peer assessment design (Design 10.1). In peer assessment, students can be assigned to assess either convergent or divergent work (i.e., work that is of a similar quality/topic as their own, or diverges from their own).

Tools/materials and participant structures: As in peer assessment.

Iterations and Augmentation: Learning analytics can be used to allocate peers automatically based on content or ability, for example, using prior assessment data, automated essay scoring, or topic modelling.

from feedback from lower achieving students, but that low ability students tend to benefit more from students who are similar to them (Cho et al. 2006a). So, it may be desirable in a number of contexts to assign peer assessment artefacts on the basis of prior knowledge, the content of the artefact, or some other features.

In addition, Design 10.3 provides a complementary pattern that introduces a 'calibration' or benchmarking task, in which students engage with assessing exemplar texts prior to their own task completion. Such tasks have, as with peer assessment more generally, often been discussed in the context of training for peer-assessment as a means to reduce the marking burden (for example, in calibrated peer review, see e.g., Balfour 2013). However, calibration tasks have a richer potential in engaging students in the application of assessment criteria to known exemplars, and the ensuing potential for diagnostic feedback and discussion about their judgements. In peer assessment, it is often the giving (not receiving) of feedback that students learn most from (Cho and MacArthur 2011; Cho and Cho 2011; Lundstrom and Baker 2009; Nicol et al. 2014). Similarly, in calibration tasks we would expect that by engaging students in giving feedback, they will learn. As such, supporting students in calibrating their judgement and feedback should support their assessment of their own work and address the concern that student's feedback can be misaligned with that of their tutors (McConlogue 2015; Patchan et al. 2009; Rollinson 2005; Watson and Ishiyama 2012). One potential means through which to provide this support is in the form of 'feedback on feedback'; that is, automated feedback on the quality of feedback that has been provided (typically in written form). This feedback might, for example, simply analyse student comments for phrases and words that are related to feedback categories (Whitelock et al. 2010, 2012), or investigate the kinds of feedback that are in fact acted on, and relate these to textual features of the feedback (Nelson and Schunn 2009; Nguyen et al. 2017; Yadav and Gehringer 2016).

Design 10.3: Calibration Tasks

Problem: Students should (1) critically apply the assessment criteria to artefacts of the form they will produce, (2) engage actively with exemplars, (3) calibrate their judgement of their own and other's work based on feedback (directly, and mediated via the diagnostic information provided to an instructor).

Task: Students are asked to assess exemplars, typically 3–5 artefacts of varying quality or on different topics. Students use the assessment criteria to assess these exemplars, providing a grade or point score, alongside written feedback. Once they have done this, students can be provided with feedback on the 'accuracy' of their assessments, as well as seeing both instructor and peer feedback provided to the same exemplars. This design *complements* Design 10.1, and can be used as a standalone task.

Tools/materials and participant structures: This task is typically conducted in individual then group structures, with students assessing individually (perhaps after a period of discussion of the criteria), and then discussing the collective feedback.

Iterations and Augmentation: Diagnostic feedback can be provided to students and instructors in a number of ways. First, student judgements provide insight into how aligned they (individually, and as a cohort) are with the instructor's assessments of the same artefacts. Second, students can be provided with support in giving high quality feedback, through the use of natural language processing on their written comments. This feedback has been shown to produce higher quality comments from students, which we would anticipate supporting their learning about the criteria and thus developing their own evaluative judgement.

Finally, peer assessment may also be augmented through providing students with targeting towards specific features of the artefacts they are asked to assess. This differs from Design 10.3, in that Design 10.3 is focused on specific features of the *feedback,* while Design 10.4 also draws on specific features of the *assessed artefact.* As such, Design 10.4 provides a mechanism to focus feedback. In some regards this design is similar to a *complementary design* (not expanded here) in which students are scaffolded in completing their peer assessment through the use of rubrics or specific questions (see, for example, Gielen and De Wever 2015), and indeed, such structuring may also be useful.

Learning analytics can augment this approach by specifically foregrounding target features in the artefacts to the assessors, who then choose to use this support in writing their feedback. That is, the peer feedback can be used to mediate automated feedback via the peer's elaboration and focusing. For example, in the Academic Writing Analytics (AWA) project (see, for example, Knight et al. 2017), natural language processing is conducted on student writing to detect the presence of 'rhetorical moves' (rhetorical structures in the text that provide structural markers to the

Design 10.4: Focused Feedback

Problem: We want students to draw on specific features of the artefacts that they assess in providing feedback.

Task: In asking students to provide feedback, we can foreground particular features of the artefact to them, through specific instructions and structured peer-assessment forms.

Tools/materials and participant structures: As in Design 10.1, with the possible additions of structured peer-assessment forms, and automated writing analytic tool access.

Iterations and Augmentation: Peers engage with artefacts that have been analysed and annotated, summarized or otherwise visualized to (a) support their own judgement of the work, and (b) mediate the automated feedback through a process of peer-sensemaking. This design complements Design 10.1.

reader). In preliminary work, peers have been asked to provide feedback to each other on their work, making use of the AWA tool to particularly foreground rhetorical structures in their comments (see, for example, Shibani 2017, 2018; Shibani et al. 2017).

10.2 Conclusion

Learning analytics can be used to support assessment processes, in the provision of feedback to both learners and educators regarding evidence of learning. This paper has discussed different ways in which learning analytics can be integrated into practical pedagogic contexts. This chapter has argued that one means through which learning analytics can achieve impact is through augmenting assessment design. Impact will be achieved by combining learning analytics with effective assessment through the kind of augmentation approach described here. I have argued that technology integration is a key consideration in taking this approach as a means to achieve impact with learning analytics. The flexibility of use of learning analytics in the context of designing assessments has been illustrated at a basic level through a set of peer assessment design patterns which illustrate how learning analytics can augment assessment. In each case, a core assessment design is highlighted, with the potential to *augment* that design with learning analytics tools demonstrated. By adopting such an approach, we can foster the testing and development of learning analytics with educators in the loop, making key decisions about how analytics relate to and impact their teaching contexts, and provide support for existing good practices in assessment design.

References

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education, 26*(2), 600–614. https://doi.org/10.1007/s40593-016-0105-0.

Balfour, S. P. (2013). Assessing writing in MOOCs: Automated essay scoring and calibrated peer review. *Research & Practice in Assessment, 8*(1), 40–48. Retrieved from http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF4.pdf.

Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65*, 107–116. https://doi.org/10.1016/j.tate.2017.02.017.

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., et al. (2009). *Returning to learning in an age of assessment: Introducing the rationale of the collegiate learning assessment*. New York: CAE.

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629–643. https://doi.org/10.1007/s11251-010-9146-1.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. https://doi.org/10.1037/a0021950.

Cho, K., Schunn, C. D., & Charney, D. (2006a). Commenting on writing typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*(3), 260–294. https://doi.org/10.1177/0741088306289261.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006b). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891.

Committee on the Foundations of Assessment. (2001). *Knowing what students know: The science and design of educational assessment*. https://doi.org/10.17226/10019.

Crisp, G., Guàrdia, L., & Hillier, M. (2016). Using e-assessment to enhance student learning and evidence learning outcomes. *International Journal of Educational Technology in Higher Education, 13*(1), 18. https://doi.org/10.1186/s41239-016-0020-3.

Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford: Stanford University, Stanford Center for Opportunity Policy in Education (SCOPE). Retrieved from https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-report_0.pdf.

Dawson, P., & Henderson, M. (2017). How does technology enable scaling up assessment for learning? In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education. The enabling power of assessment* (Vol. 5, pp. 209–222). Singapore: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-981-10-3045-1_14.

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges* (Technical report no. kmi-12-01). The Open University, UK: Knowledge Media Institute. Retrieved from http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf.

Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014). Setting learning analytics in context: Overcoming the barriers to large-scale adoption. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 251–253). New York: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2567592.

Fives, H., Barnes, N., Buehl, M. M., Mascadri, J., & Ziegler, N. (2017). Teachers' epistemic cognition in classroom assessment. *Educational Psychologist, 52*(4), 270–283. https://doi.org/10.1080/00461520.2017.1323218.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315–325. https://doi.org/10.1016/j.chb.2015.06.019.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics, 1*(2), 23–47. https://doi.org/10.18608/jla.2014.12.3.

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (2017). Academic writing analytics for civil law: Participatory design through academic and student engagement. *International Journal of Artificial Intelligence in Education, 28*(1), 1–28. https://doi.org/10.1007/s40593-016-0121-0.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21. https://doi.org/10.3102/0013189X020008015.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439–1459. https://doi.org/10.1177/0002764213479367.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064.

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011.

McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher Education, 40*(9), 1495–1506. https://doi.org/10.1080/03075079.2013.868878.

Mislevy, R., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centred design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*(1), 11–48. Retrieved from https://jedm.educationaldatamining.org/index.php/JEDM/article/view/22.

Muis, K. R., & Franco, G. M. (2009). Epistemic beliefs: Setting the standards for self-regulated learning. *Contemporary Educational Psychology, 34*(4), 306–318. https://doi.org/10.1016/j.cedpsych.2009.06.005.

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x.

Nguyen, H., Xiong, W., & Litman, D. (2017). Iterative design and classroom evaluation of automated formative feedback for improving peer feedback localization. *International Journal of Artificial Intelligence in Education, 27*(3), 582–622. https://doi.org/10.1007/s40593-016-0136-6.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518.

O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education, 74*(4), 617–633.

Pardo, A. (2017). A feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education, 43*(3), 428–438. https://doi.org/10.1080/02602938.2017.1356905.

Patchan, M. M., Charney, D., & Schunn, C. D. (2009). A validation study of students' end comments: Comparing comments by students, a writing instructor, and a content instructor. *Journal of Writing Research, 1*(2), 124–152. Retrieved from http://d-scholarship.pitt.edu/22893/1/JoWR_2009_vol1_nr2_Patchan_et_al.pdf.

Pellegrino, J. W. (2013). Measuring what matters in a digital age: Technology and the design of assessments for multisource comprehension. In D. G. Sampson, P. Isaias, D. Ifenthaler, & J. M. Spector (Eds.), *Ubiquitous and mobile learning in the digital age* (pp. 259–286). New York: Springer. https://doi.org/10.1007/978-1-4614-3329-3_17.

Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford: ASKe, Oxford Centre for Staff and Learning Development.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal, 59*(1), 23–30. Retrieved from http://eltj.oxfordjournals.org/content/59/1/23.short.

Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education, 30*(3), 231–240. https://doi.org/10.1080/02602930500063819.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice-based assessments for the digital age*. Boston: MIT Press.

Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions*. Abingdon: Routledge.

Shibani, A. (2017). Combining automated and peer feedback for effective learning design in writing practices. In Y. Yu, S. Murthy, B. Chang, M. Chang, G. Biswas, & J. Mason (Eds.), *Proceedings of the 25th international conference on computers in education*. Canterbury: Asia-Pacific Society for Computers in Education. Retrieved from http://icce2017.canterbury.ac.nz/proceedings.

Shibani, A. (2018). Developing a learning analytics intervention design and tool for writing instruction. In S. B. Shum, R. Ferguson, A. Merceron, & X. Ochoa (Eds.), *Companion proceedings 8th international conference on learning analytics & knowledge (LAK18)*. Sydney: Society for Learning Analytics Research (SoLAR). Retrieved from http://bit.ly/lak18-companion-proceedings.

Shibani, A., Knight, S., Buckingham Shum, S., & Ryan, P. (2017). Design and implementation of a pedagogic intervention using writing analytics. In W. Chen, J.-C. Yang, A. F. M. Ayub, S. L. Wong, & A. Mitrovic (Eds.), *25th international conference on computers in education* (pp. 306–315). Christchurch: Asia-Pacific Society for Computers in Education. Retrieved from http://icce2017.canterbury.ac.nz/sites/default/files/proceedings/main/C3/Design%20and%20Implementation%20of%20a%20Pedagogic%20Intervention%20Using%20Writing%20Analytics.pdf.

Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford: Stanford University, Stanford Center for Opportunity Policy in Education. Retrieved from http://edpolicy.stanford.edu/sites/default/files/publications/performance-assessment-era-standards-based-educational-accountability_0.pdf.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265–269. https://doi.org/10.1016/j.learninstruc.2009.08.002.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249.

Watson, W., & Ishiyama, J. (2012). *Using calibrated peer review to facilitate writing assignments in large political science classes*. Paper presented at APSA 2012 Teaching & Learning Conference, Washington DC. Abstract retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1997612.

Whitelock, D. (2010). Activating assessment for learning: Are we on the way with web 2.0? *Web, 2*, 319–342.

Whitelock, D., Pinto, R., & la Saez, M. (2010). Modelling the teachers' feedback process for the design of an electronic interactive science tool with automatic feedback. *International Journal of Continuing Engineering Education and Life Long Learning, 20*(2), 189–207. https://doi.org/10.1504/IJCEELL.2010.036815.

Whitelock, D., Gilbert, L., Hatzipanagos, S., Watt, S., Zhang, P., Gillary, P., & Recio, A. (2012). Addressing the challenges of assessment and feedback in higher education: A collaborative effort across three UK universities. In L. Gómez Chóva, A. Lopez Martínez, & I. Candel Torres (Eds.), *Proceedings of the 6th international technology, education and development conference 2012*. Valencia: IATED. Retrieved from http://eprints.soton.ac.uk/273207/.

Wimshurst, K., & Manning, M. (2013). Feed-forward assessment, exemplars and peer marking: Evidence of efficacy. *Assessment & Evaluation in Higher Education, 38*(4), 451–465. https://doi.org/10.1080/02602938.2011.646236.

Yadav, R. K., & Gehringer, E. F. (2016). Metrics for automated review classification: What review data show. In Y. Li et al. (Eds.), *State-of-the-art and future directions of smart learning* (pp. 333–340). Singapore: Springer. https://doi.org/10.1007/978-981-287-868-7_41.

Zhao, Y., Pugh, K., Sheldon, S., & Byers, J. L. (2002). Conditions for classroom technology innovations. *Teachers College Record, 104*(3), 482–515. Retrieved from http://crcsalon.pbworks.com/f/Conditions+for+Classroom+Technology+Innovations.pdf.

Simon Knight is a lecturer in the University of Technology Sydney, Faculty of Transdisciplinary Innovation. He is co-editor-in-chief of the Journal of Learning Analytics, and on the editorial board of the International Journal of Computer-Supported Collaborative Learning. With a background in psychology, philosophy, and education, Dr Knight's work focuses on how people think about evidence, and use evidence to inform that decisions, particularly seen through the lens of epistemic cognition. In the learning analytics and learning design context, this research focuses on how people design analytics, and what theoretical assumptions those designs embody.

Chapter 11
Are Assessment Practices Well Aligned Over Time? A Big Data Exploration

Jekaterina Rogaten, Doug Clow, Chris Edwards, Mark Gaved, and Bart Rienties

Abstract In the last 20 years a range of approaches have been adopted to facilitate assessment of learning as well as assessment for learning. With the increased interest in measuring learning gains using assessment data, it is important to recognise the potential limitations of using grades as proxies for learning. If there is a lack of alignment in terms of grade descriptors between modules within a qualification, students might perform really well on one module, and may underperform in a module that has relatively "harsh" grading policies. Using principles of big data, we explored whether students' grade trajectories followed a consistent pattern over time based upon their abilities, efforts, and engagement in two distinct studies. In Study 1, we explored a relatively large dataset of 13,966 students using multi-level modelling, while in a more fine-grained Study 2 we focussed on the pathways of students choosing their first two modules in six large qualifications. The findings indicated substantial misalignments in how students progressed over time in 12 large qualifications in Study 1. In Study 2, our analyses provided further evidence that students' grades did not seem to be well aligned. In all qualifications we found a highly significant effect of change over time depending on the achievement group. Based upon these findings, we provide clear recommendations how institutions might use similar insights into big data, and how they may improve the longitudinal alignment of grading trajectories by using consistent grading policies.

J. Rogaten · C. Edwards · M. Gaved · B. Rienties (✉)
Open University UK, Milton Keynes, UK
e-mail: jekaterina.rogaten@open.ac.uk; chris.edwards@open.ac.uk; mark.gaved@open.ac.uk; bart.rienties@open.ac.uk

D. Clow
Independent Consultant, Milton Keynes, UK
https://dougclow.org/

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_11

11.1 Introduction

Assessment is a key driver for learning (Bearman et al. 2016; Boud 2017; Coates 2016). In the last 20 years a range of approaches have been adopted to facilitate assessment of learning (Boud 2017; Boud and Falchikov 2006; Coates 2016), as well as assessment for learning (Bearman et al. 2016; Carless 2007; Carless et al. 2011). With the introduction of the Teaching Excellence Framework (TEF) in the UK, there is increased interest in measuring learning gains (Johnson 2015; McGrath et al. 2015). The broad assumption of the TEF is that universities that provide students with excellent teaching and learning opportunities will lead to high learning gains and value added, which will be financially rewarded.[1]

One approach that is currently developed across a number of Office for Students projects[2] is to use students' academic performance as a proxy for estimating learning gains. This approach capitalises on the large quantities of student data routinely gathered by every university and may provide preliminary data-driven big data comparisons between different subjects, or even across different universities. Over the years, researchers and practitioners have tested a range of measurement approaches aiming to capture relative improvements in student learning (e.g., Anderson 2006; Hake 1998).[3] Furthermore, using students' academic performance as a measure of learning progress has other advantages; firstly, it is widely recognized as a common proxy for mastery and learning. Secondly, grades are relatively free from self-reported biases, and thirdly, using academic performance allows a direct comparison of research finding with the results from other studies (Bowman 2010; Gonyea 2005; Rogaten et al. 2017). If a qualification (i.e., a set of modules and courses that build towards a certificate or degree) is well designed and assignments are aligned according to well-defined grade descriptors and/or rubrics (Bell et al. 2013; Dawson 2017; O'Donovan et al. 2004), it would be reasonable to assume that as the level of difficulty increases, the grading over time will be adjusted.

Although the use of grades as proxies for learning gains sounds attractive, it is important to recognise the potential limitations of using grades as proxies for learning (Boud 2018; Rogaten et al. 2017). A number of factors might explain why learning gains in a qualification, and grades in particular, might go up or down over time. Within the TEF framework, an assumption is that as students develop knowledge and skills in a qualification, students will strengthen their abilities to interlink concepts, to master key skills, and to be able to solve increasingly complex

[1] Although the specific details of the proposed measurements and metrics to be used for learning gains still have to be determined by the UK Government, future government funding might become related to students' learning gains as part of the teaching excellence narrative (Ashwin 2017; Johnson 2015).

[2] https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/

[3] The most common way of assessing learning gains is through use of pre-post testing (e.g., Dimitrov and Rumrill 2003). Although pre-post testing is considered as a standard and favourable approach for assessing learning gains, it can be a costly process, especially when measured across different modules.

problems (Higher Education Commission 2016; Johnson 2015; McGrath et al. 2015). However, as argued by a recent opinion piece by Boud (2018) without clear learning outcomes that are embedded in a framework of explicit standards, it might be potentially inappropriate to compare assessment grades across modules and over time. At the same time, with the Open University UK (OU) the practice of awarding marks relative to the standard expected of students at that stage of the qualification, rather than expecting marks to increase substantially over the course of the qualification, seems widespread.

If there is a lack of alignment in terms of grade descriptors between modules within a qualification, students might perform really well on one module, and may underperform in a module that has relatively harsh grading policies. It is well known from the several studies carried out into the reliability of assessment (e.g., Meadows and Billington 2005; Moxley and Eubanks 2015), that there can be substantial disparities and inconsistencies between (and even within) human markers, and there is evidence to suggest this may be a particular problem in distance education (Rienties et al. 2017a, b). Furthermore, as argued by Boud (2018) grades of a module may not be indicators of each of its learning outcomes, and when grades are aggregated across different learning outcomes, it may become impossible to determine what the respective outcomes each represent.

A particular concern relevant for our big data study is that Boud (2018, p. 5) argued that "[p]ass marks are determined within a disciplinary culture in relation to the internal features of the course unit and its tasks; no calibration of pass marks in relation to agreed standards is typically undertaken across units. Fifty per cent, say, is assumed to have a universal shared meaning and referent across different subject matter and different levels". Indeed, if we find large variations across modules across a qualification, this may imply that we may need to look at the potential alignments or misalignments between assessments within and across modules within a qualification.

Therefore, we used principles of big data (Ferguson et al. 2016; Rienties and Toetenel 2016) in an online distance learning context to explore whether students' grade trajectories followed a consistent pattern over time based upon their abilities, efforts, and engagement in two distinct studies. This chapter presents these studies both in terms of the assessment practices they reveal but also as an illustration of how big data can inform institutional assessment practices. While we acknowledge the vast body of research on assessment practice, as also highlighted in this book, relatively few studies have used principles of big data to explore whether students' grade trajectories follow logical patterns (or not), and whether individual student characteristics might mediate these relations. Rather than hypothesis testing, in this chapter we primarily use an explorative study in one specific big data context, namely the OU.

Our exploration of students' grade trajectories from a large number of online students over time may help researchers to reflect upon whether the efforts of those who design assessments actually led to consistent assessment practices over time, and in particular whether those who graded students on their work did so in a consistent, intertemporal manner, ideally across a range of qualifications and

disciplines. We specifically chose to conduct our study at the largest university in Europe, the OU, because great care and attention is provided towards designing and implementing modules and qualifications (Rienties and Toetenel 2016; Toetenel and Rienties 2016), and extensive quality assurances and practices are in place given the complex, large scale of educational provision amongst thousands of teachers and instructors at a distance (Richardson 2013; Richardson et al. 2015).

In Study 1, we used a relatively large dataset of 13,966 students to explore grade trajectories over time using multi-level modelling, whereby we acknowledge a hierarchical structure of the dataset through nesting data within three levels: module level and its related characteristics (e.g., module structure, workload, complexity of assessments, alignment of assessments with previous and follow-up modules); students level with its related characteristics (e.g., ability, socio-demographics); qualification level and its related characteristics (e.g., composition and sequence of modules to obtain a qualification). Therefore, our first research question is: To what extent are grade trajectories of students over time consistently aligned from one module to another, and how are these grade trajectories influenced by students' characteristics and qualification pathways?

Afterwards, in the more fine-grained Study 2, we focussed on six large qualifications with the OU, where we analysed the first two modules undertaken by students. We wished to explore how the paths that "new" students (in terms of studying for the first time at the OU) were taking through their qualification affected their achievement in terms of final marks of those first two modules. As highlighted by a wealth of research in higher education, and first-year experience in particular (Hillstock and Havice 2014; Rytkönen et al. 2012; Yorke and Longden 2008), the transition in the first two modules is of essential importance for successful progression and continuation of study. Furthermore, recent large-scale research on students' experiences found substantial differences in how 16,670 "new" students experienced studying at the OU in comparison to 99,976 students who already completed several modules at the OU (Li et al. 2017). If qualifications and introductory modules are well structured and assessment well-aligned, we would expect new students who are high achievers on their first module to tend to be high achievers on their second module, and low achievers on that module to be low achievers on the next (Conijn et al. 2017; Koester et al. 2016; Popov and Bernhardt 2013). Therefore, our second research question is: To what extent do the paths students take through the first two modules of their qualification impact their achievement in terms of marks?

11.2 Methods

11.2.1 Setting

Beyond our open-entry policy, another particular feature of the OU is that students may follow specific pathways within a qualification, or select modules and sequences of modules based upon their preferences. By mixing and matching modules over

time, within certain qualifications students have substantial freedom to follow their interests, and select modules that fit with their interests, while other qualifications follow more structured, fixed pathways. As previous research has found substantial differences between postgraduate and undergraduate learning designs (Li et al. 2017; Rienties and Toetenel 2016), this study included only undergraduate modules that have run from 2013 to 2017.[4]

11.2.2 Participants

For Study 1 a total of 13,966 students were included in a multi-level longitudinal analysis, whereby we selected several large qualifications for each of the four Faculties.[5] Students in this sample all have achieved minimum grade of "pass" on all modules they were enrolled in. As such, this sample represents students who were continuously "successful". For example, for students who were enrolled from October 2013 onwards, this would in practice mean that they would have needed to pass 4–6 modules consecutively to be included, while students who were enrolled from October 2015 onwards would only need to have passed at least two modules. This is a very important caveat, as OU research and practice has consistently shown that many students are not always successful in terms of completing consecutive modules (Calvert 2014; Li et al. 2016, 2017). One would assume that the selected student cohort (who passed all modules they were enrolled in) would continuously do well and perform with similar grades over time.

In Study 2, we selected a sample of students who passed at least two modules in the period from 2013 to the end of the 2016 calendar year. For each of the four Faculties, the top two qualifications in terms of student numbers were selected, apart from Faculty B, where only the top qualification was chosen, because the second most-popular qualification was fairly similar to the first. Within each qualification, the most popular paths taken by students were selected. As a result, 6794 students across these six qualifications were included in the analyses.

[4] OU processes changed significantly at 2013, meaning comparison with studies undertaken prior to this time are difficult to make and of less interest to academics seeking analysis of current offerings.

[5] In terms of demographics, commonly more female students (61%) than male students (39%) study at the OU. Most students are from the UK (96%) from a white background. Students vary considerably in age, with 24% under 25 years old, 37% aged 26–35, 22% aged 36–45, 13% aged 45–55, and 5% aged 56 and over. More than half of students work full-time (53%), while 21% work part-time, 7% are looking after the home/family, and 5% are unemployed and looking for a job. Regarding learners' qualifications, there are no formal academic entry requirements at undergraduate level at the OU. Around 40% of the students have A levels or equivalent (suggesting they had two or more years of post-compulsory schooling), 28% have less than A levels (suggesting they had not progressed beyond compulsory schooling), and around a quarter have a higher education or post-graduate degree. On average, 10% of the students report one or multiple disabilities. Participants in Study 1 and Study 2 were fairly similar in terms of demographics.

11.2.3 Data Analysis

For Study 1, a 3-level growth curve model (Rasbash et al. 2009; Rogaten et al. 2017) was fitted on student overall module grades taken each year starting from October 2013/2014 onwards[6] using MLwiN. In Study 2 the 6974 students were split into three distinct achievement groups – high, medium, or low – based on their marks in their first module.[7]

11.3 Results

11.3.1 Study 1

Figure 11.1 displays the qualification progression trajectories for some of the most popular qualifications at the OU. Each sequential module achievement is marked by the 'dot' on the line. Note that the straight lines between the assessment points are a result of regression modelling, whereby MLwiN predicts the best lines between assessment points. Of the 12 qualifications included in this analysis, all except one qualification (QUALC) had a negative trend over time, which indicates that students' performance in terms of grades dropped over time while going through a qualification. For example, the red line in Fig. 11.1 represents QUALF, a science-related qualification, and there were six assessment points associated with the current progression of students in this qualification. On average, students in QUALF obtained 79.2 (SD = 11.3) for this first module, which is an above average grade. However, as indicated in Fig. 11.1, the assessment scores over time deteriorated, with a final module score after 6 modules of 70.6 (SD = 13.2) for those students who passed all modules. In other words, "successful" students who completed all 6 modules in QUALF had a mean difference of −10.9 grade points.

[6] Multilevel growth-curve modelling allows for estimating individual students' learning trajectories by fitting an overall average module curve and allowing each individual students' curve to depart from the average module curve. Using multilevel modelling it is possible to estimate what is the variance in students' initial achievements and their subsequent grade trajectories depending on what module they are enrolled in and whether students' initial achievements and grade trajectories depend on their individual differences and socio-demographic characteristics. As students at the OU can choose different pathways and elective modules (Edwards 2017), not only can we compare how students progress within a qualification (e.g., Student 1 and Student 2) and which order of modules is most beneficial in terms of obtained grades, but we can also compare how students complete modules from other qualifications (e.g., Student 2 following Module 1 in Qualification 2; Student 3 following Module 2 in Qualification 3).

[7] I.e., Low = 40–59, Mid = 60–69, High = 70 +. For each qualification, a mixed ANOVA was carried out with time as a within-subject factor (first module to second module), and path (the top study paths, all others grouped) and achievement group (low, mid, high) as between-subject factors, and marks on the second module as the dependent variable.

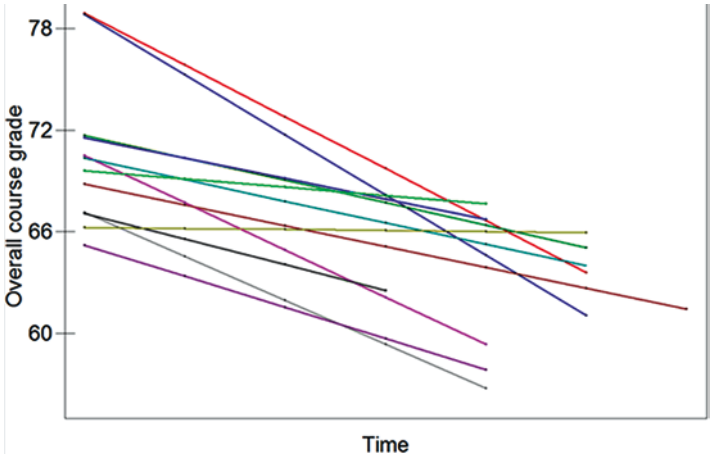


Fig. 11.1 Qualification progression over time (estimated regression lines)

After students completed their first module, our modelling indicated that students were predicted to have high grades for the next module, with the average module grade of around 70. However, in contrast to our predictions the module grades dropped as students progressed from their first module to the next module (B = − 1.746), and the range of the drop between modules progression ranged between −4.566 and 1.074 grade points. For a detailed breakdown of the statistical analyses, see (Rienties et al. 2017b). As indicated in Table 11.1, the results showed that on average OU students performed relatively well in their first module (M = 70.8, SD = 4.7), but their grades dropped as students continued their studies towards qualification to an average of 68.8 (SD = 6.0).

11.3.1.1 Level 3 Role of Qualifications on Grade Trajectories

Based upon the multi-level modelling the variance partition indicated that there was 12% of variance between the different qualifications (i.e., Level 3). Attainments achieved in any two qualifications correlated very weakly, and there was no strong alignment between selected undergraduate qualifications. In plain English, each qualification trajectory was quite independent from one another despite some groups of students taking the same modules for different qualifications. Of particular interest in the OU context, many students select and mix modules from different qualifications, so our findings might indicate that this might not be as beneficial (result in success) as perhaps hoped for. Importantly, some qualifications had high initial students' achievements on the first module, while others had relatively low initial achievements. In total, the qualification route accounted for 20% of variance in students' first module achievements. In other words, substantial differences are present when students start with a particular qualification, indicating a potential need to align introductory modules across the OU.

Table 11.1 Mean progression of students per qualification

	Qualification	Mean first module of all students who passed their first module (SD)	Mean last module of those who completed 5–6 modules (SD)
1	QUALF (n = 1736)	79.2 (11.3)	70.6 (13.2)
2	QUALE (n = 851)	80.3 (11.2)	73.9 (10.3)
3	QUALC (n = 2629)	71.8 (10.1)	80.6 (10.8)
4	QUALC (n = 405)	71.5 (11.4)	57.9 (9)
5	QUALD (n = 968)	70.9 (9.4)	68.5 (11.8)
6	Open degree (n = 3252)	70.4 (12.3)	69.9 (8.8)
7	QUALG (n = 616)	69.9 (9.5)	68.5 (9.8)
8	QUALG (n = 1079)	68.3 (10)	67.6 (9.4)
9	QUALA (n = 599)	66.7 (10)	65.6 (9.0)
10	QUALC (n = 423)	68.3 (10.1)	62.5 (2.1)
11	QUALG (n = 344)	66.2 (9.1)	64.6 (12.9)
12	QUALG (n = 980)	65.5 (10.5)	75 (10.0)
	Average across qualifications	**70.8 (4.7)**	**68.8 (6.0)**

Note: Qualifications were anonymised in six broad categories in line with OU Ethics policy. Multiple qualifications may be provided within each of these six broad categories

Furthermore, the particular qualification students were enrolled in determined their progress trajectories, and in total 30% of subsequent grade trajectories students made were due to being enrolled into a particular qualification degree. Qualifications that had first modules with relatively high average achievements tended to have a more rapid decline in their following semesters' average achievements, whereas qualifications with lower initial module grades had a lower decline. Note that given that nearly all qualifications had a negative grading trajectory, a lower decline in a way is an above average performance.

11.3.1.2 Level 2 Role of Student Characteristics in Grade Trajectories

The largest portion of variance in this model is at the student level (45%). The effort and time that students are able to put into studying will influence their performance within a module and across a qualification. Given the specific nature of OU students and the large impact of life-events on students' progression (Calvert 2014; Li et al. 2017; van Ameijde et al. 2016), it seems plausible that a large part of variation is explained by individual circumstances.

Given the widening access agenda of the OU (Li et al. 2017; Richardson 2013; Richardson et al. 2015; van Ameijde et al. 2016) one would hope that students from a widening access background, who might initially struggle on the first module, will become more successful over time. However, our multi-level analyses indicated that students with below average achievements on their first module tended to have a

steeper drop in their consequent module attainments. In contrast, students who obtained above average grades in their initial module had a lower drop in their subsequent module attainments.

11.3.1.3 Level 1 Impact of Modules on Grade Trajectories

Lastly, 43% of variance lay on a module level, or "within-students", which indicated that there was a large proportion of variation (inconsistencies) between modules that form a particular qualification route. In plain English, if a student scored 70 on the first module and 70 on the second module, one would expect that this student would also score around 70 on the third module, fourth module, etc. However, substantial variation in module scores were present in all qualifications. As illustrated in Fig. 11.2 of the QUALF, the actual scores of individual students across the six modules varied substantially.

As illustrated by Fig. 11.2 (1), the average score across the six modules in QUALF over time declined from module to module, with a notable exception for the fifth module. In terms of students' actual scores on their respective modules, Fig. 11.2 (2) illustrates the wide variations in students' scores, whereby the lines from one module to the next show substantial variation. In other words, if modules would be "perfectly" aligned with a qualification (Boud 2017; Rienties et al. 2017b), one would expect flatter lines of individual student journeys, and variation would primarily be explained by individual ability, effort, and contribution.

Finally, Fig. 11.2 (3) illustrates the predicted regression lines for each participant, which in most cases were downward sloping (i.e., indicating negative grade trajectories). Students who passed the first module with a good grade of 70+ were most likely to continue on follow-up modules, as more lines are visible in Fig. 11.2 (3) for the third module onwards, relative to students with an initially low first module grade. In particular students with very high scores (80+) continued over time, and mostly had similar grades in subsequent modules. Students who score below 65

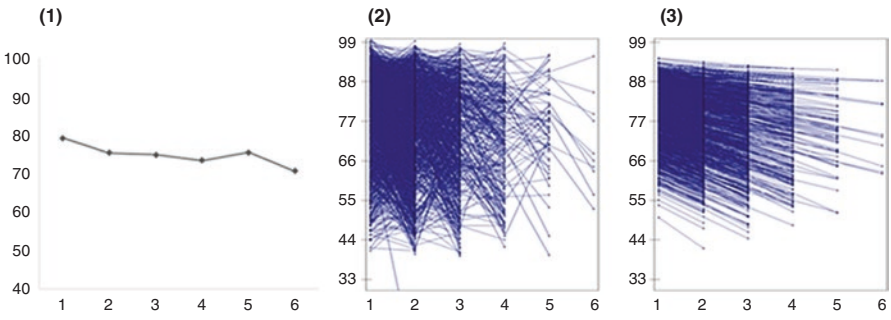


Fig. 11.2 Module scores across six modules of QUALF
Note (1) Mean across modules in the qualification, (2) Actual module scores for each participant, and (3) trellis plot predicted regression lines for each participant

mostly performed worse for the second module, and were more likely to stop after the second or third module.

11.3.2 Study 2

In Study 2 we extended our analyses by looking specifically at the first two modules that "new" students took at the OU in order to determine whether we are providing a consistent practice at the start of their journey. As expected, many new students opted for a range of pathways after following their first module (e.g., QUALA, QUALF, QUALG).[8] As highlighted previously in Table 1.2 in Rienties et al. (2017a), most students obtained a lower grade in their second module in comparison to their first module. In all but one instance, there was a significant time-path interaction ($p < .05$), and this was highly significant in five qualifications ($p < .001$). That is to say, with the exception of QUALE, students' grades changed over time depending on which study path they chose: some paths led to grades going up, and some to grades going down. Perhaps the most striking effect in this analysis is that there was a highly significant time-achievement interaction in every single case ($p < .001$). That means that students in different achievement groups (high, mid, low) had different changes to their grades over time. If assessments were well aligned, we would expect achievement groups to be on average stable over time.

For example, Fig. 11.3 shows an example of divergent changes in grades over time depending on path and achievement group for QUALA. Low achieving students on QUALAM1 (blue lines) tended to get markedly higher results on QUALAM2, but those who studied QUALAM4 got even higher grades – higher, in fact, than mid-achievers on QUALAM1 (green lines), who tended to decline in grades when they came to QUALAM4. High achieving students on QUALAM1 (red lines) achieved consistently high results when they got on to QUALAM2 (left-hand chart), but their grades declined if they instead choose QUALAM4 (right-hand chart). Again for a detailed statistical analyses of this and other qualifications, see Rienties et al. (2017a).

11.4 Implications for Assessment Practice

These studies indicate the potential value of big data to inform institutional assessment practices, which allowed us to track how students progressed over time in terms of grade trajectories, and how these were related to grading practices within

[8] For example, in the two QUALA qualifications, all students started with module QUALAM1, but afterwards some students in the first QUALA selected QUALAM2, while others selected QUALAM3 or QUALAM4. Only for QUALD did all students in the sample follow the same second module.

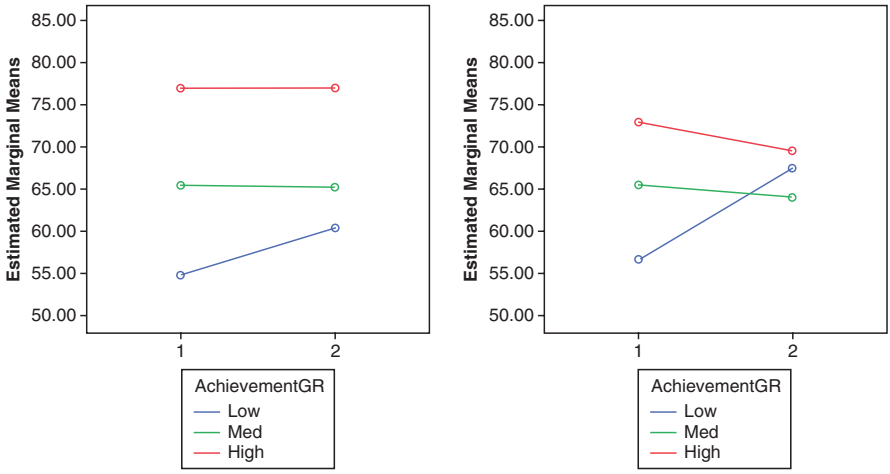


Fig. 11.3 Mean grades for students on QAULA by achievement group, for those studying QUALAM1 then QUALAM2 (left-hand chart) and those studying QUALAM1 then QUALAM4 (right-hand chart)

and across modules within one distance learning institution. After students completed their first module, our modelling indicated that students were predicted to have high grades for the next module, with an average module grade of around 70. However, in contrast to our predictions the module grades dropped as students progressed from their first module to the next module. As visually illustrated in Fig. 11.1, this negative trend continued for most qualifications. A large part of this trend seems to result from institutional (module and qualification) factors.

Our multi-level analysis indicated that 12% of variance was explained on a qualification level, whereby each qualification trajectory was quite independent from one another despite some groups of students taking the same modules for different qualifications. Depending on the selected qualification, students' progression and grade trajectories in particular were significantly impacted. This is a relatively surprising finding, as in many qualifications students have substantial freedom to mix and match modules across the OU. Some qualifications seem to help students to obtain similar/comparable learning experiences and assessment outcomes, while more variation seems to be present in other qualifications. Our big data explorations highlight a potential need to better align expectations and modules within a qualification across the OU, as students get substantially different experiences depending on the respective qualification they are enrolled into.

The largest portion of variance in this model was explained by individual student characteristics (e.g., effort, ability, socio-demographics). Given the widening access agenda of the OU (Richardson 2013; Richardson et al. 2015) one would hope that students from a widening access background, who might initially struggle on the first module, will become more successful over time. However, our multi-level analyses indicated that students with below average achievements on their first module tended to have a steeper drop in their consequent module attainments. In contrast,

students who obtained above average grades in their initial module had a lower drop in their subsequent module attainments.

Lastly, another relatively surprising finding from Study 1 was that the students' journey from one module to another caused substantial transitional problems and imbalances in students' progression (43% of variance). Substantial variation in module scores were present while students were working through modules in all qualifications, which could be explained by inconsistent alignment of grade descriptors across a qualification and variations in marking within a module. This was further strengthened and confirmed in Study 2, where we saw a highly significant effect of the study path chosen on grades in the subsequent module. This was recently highlighted by Boud (2018), who noted that researchers need to tread carefully when comparing grades across time and discipline when the underlying frameworks of assessments and grading practices are not well aligned on an institutional level.

As discussed, there are many potential explanations for some of the particular instances observed: for instance, we would hope for some small improvement in grades for low-achieving students through our efforts to support them; alternatively, different modules teach and assess different skills and knowledge, for which aptitude may not be so strongly correlated. However, the size and consistency of the findings here suggest strongly that there are some serious discrepancies in assessment between modules on the same qualification. There are substantial challenges in aligning modules which have roles in multiple qualifications.[9] This adds extra weight to the recommendation to developing university-wide, cross-faculty processes for better aligning assessment and grading (Bearman et al. 2016; Boud 2018; Dawson 2017; Rienties et al. 2017a).

11.5 Considerations for Practice

Based upon the findings from both studies, we identified three broad issues from our data: (a) substantial freedom for students to select pathways; (b) alignment of modules within a qualification; and finally (c) alignment of marking across a qualification.

Substantial freedom for students to select unique pathways: as highlighted by the detailed pathways that students can choose to complete a qualification, some programmes and qualifications have relatively fixed and structured pathways, whereby the options to choose different electives are limited. In contrast, other programmes and qualifications offer OU students wide and far reaching freedom to choose. However, like most other institutions the OU provides limited to no structural

[9] For example, QUALAM3 appears in this analysis both as a second module for QAULA and as a first module for QUALB, which are located in different faculties (FacultyC and FacultyA).

support about which pathways would fit students' needs and abilities, in contrast to other universities.

For example, in a large-scale adoption of Degree Compass, a course recommendation system, across two universities and two colleges in the US involving 40,000 students, Denley (2014) reported that the recommender analytics system steered students towards modules in which they were more likely to succeed. Similarly, in a large scale-adoption of E-advisor at Arizona State University, freshmen to sophomore retention rates increased from 76 to 84% (Phillips 2013). Likewise, Denley (2014) found that a "six-year graduation rate… increased from 33 to 37.4%" (p. 65) when introducing course recommendations to students.

Consideration 1: Institutions are encouraged to reflect on how to improve their communication to their students regarding which modules fit with their needs and abilities, and be more explicit about successful pathways for students to obtain a qualification.

Furthermore, as highlighted by the recent Innovating Pedagogy Report (Ferguson et al. 2017), it is important that institutions start to think about providing smart learning analytics to students and staff to help them to make the best decision of which qualification pathways might be the best way forward.

Consideration 2: Institutions may want to explore whether (or not) to invest in smart learning analytics recommender systems that can help staff and students to support which paths within qualifications lead to highest success.

Alignment of modules within a qualification: As also highlighted by recent research at the OU (Nguyen et al. 2017a, b; Rienties and Toetenel 2016; Toetenel and Rienties 2016) and these two studies in particular, students experience substantially different assessment practices and learning designs in general when transitioning from one module to another. Providing a consistent learning experience for students within and across a qualification will help students to adjust quickly and focus on their learning objectives, rather than spending a lot of time and effort trying to understand what is expected when a new module has a different design. Recent research in other institutions found similar inconsistencies in learning design and assessment practices (Bakharia et al. 2016; Mittelmeier et al. 2018). Therefore,

Consideration 3: Institutions should consider how they communicate and manage students' expectations of the learning designs and assessment practices from one module to another.

Consideration 4: In the longer term, it would be beneficial to align module designs across a qualification based upon evidence-based practice and what works, thereby allowing smooth transitions from one module to another in a qualification.

Alignment of marking within and across modules within and across qualifications: One potential explanation for large variations in grade trajectories in both studies is the effect of embedded expectations, norms and practice in relation to marking (Boud 2017, 2018; Boud and Falchikov 2006; Dawson 2017). Across some qualifications there appeared to be a widespread deliberate approach of making

early assessment relatively easy (Rienties et al. 2017a, 2017b), both within modules (particularly the first assessment) and within qualifications (particularly the first module). This approach is intended to reduce drop-out, but may have unintended consequences. Furthermore, given that in most OU modules associate lecturers (external teachers hired to teach OU modules) are marking relatively small numbers of 10–20 assignments, potential misalignments might be present which may not be immediately apparent when just looking at average grades and the normal distribution curves currently used in quality assurance processes. Another potential explanation is that the increasing difficulty of the material being assessed may not be completely accounted for in the marks awarded. Final-year-equivalent modules rightly contain much more difficult material than entry modules. Ensuring that this is properly accounted for in marking expectations and practice is challenging, even if there was consensus that it should be.

Consideration 5: It is good-practice that grades are aligned both within a module as well as across a qualification. For exam boards we recommend the inclusion of cross-checks of previous performance of students (e.g., correlation analyses) and longitudinal analyses of historical data to determine whether previously successful students were again successful, and whether they maintained a successful learning journey after a particular module.

Consideration 6: We recommend that clearer guidelines and grade descriptors across a qualification are developed, which are clearly communicated to staff and students to encourage effective uptake in the long-run.

Consideration 7: Given that many students follow modules from different qualifications, it is important to develop coherent university-wide grade descriptors and align marking across qualifications.

The use of big data reveals ways that assessment practices can be made more explicit that are otherwise "invisible" to educators and institutions. From these studies we found that even though substantial quality assurance and enhancement practices are in place in this institutional context, without a big data perspective the potential longitudinal misalignments of the complex students' journeys and assessment practices across several qualifications would not have been identified. By making these journeys and practices visible to staff, a start of a conversation can be made about how to potentially improve the alignment of assessment practices over time.

Acknowledgements This work is in part funded by the OfS grant "A longitudinal mixed-method study of learning gain: applying Affective-Behaviour-Cognition framework at three institutions". Study 1 was first published under a Creative Commons license as Chapter 1 in Rienties, B., Rogaten, J., Nguyen, Q., Edwards, C., Gaved, M., Holt, D., … Ullmann, T. (2017). *Scholarly insight Spring 2017: a data wrangler perspective*. Milton Keynes: Open University UK. Study 2 was first published using Creative Commons as Chapter 1 in Rienties, B., Clow, D., Coughlan, T., Cross, S., Edwards, C., Gaved, M., … Ullmann, T. (2017). *Scholarly insight Autumn 2017: a data wrangler perspective*. Milton Keynes: Open University.

References

Anderson, A. K. (2006). An assessment of the perception of learning gains of freshmen students in an introductory course in nutrition and food science. *Journal of Food Science Education, 5*(2), 25–30. https://doi.org/10.1111/j.1541-4329.2006.tb00078.x.

Ashwin, P. W. H. (2017). Making sense of the Teaching Excellence Framework (TEF) results. In *Centre for global higher education policy briefings* (Vol. 1). Lancaster.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., … Lockyer, L. (2016). *A conceptual framework linking learning design with learning analytics.* Paper presented at the proceedings of the sixth international conference on learning analytics & knowledge, Edinburgh.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the assessment design decisions framework. *Teaching in Higher Education, 21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217.

Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education, 38*(7), 769–788. https://doi.org/10.1080/02602938.2012.714738.

Boud, D. (2017). Standards-based assessment for an era of increasing transparency. In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education. The enabling power of assessment* (Vol. 5). Singapore: Springer.

Boud, D. (2018). Assessment could demonstrate learning gains, but what is required for it to do so? *Higher Education Pedagogies, 3*(1), 1–3. https://doi.org/10.1080/23752696.2017.1413671.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413. https://doi.org/10.1080/02602930600679050.

Bowman, N. A. (2010). Can 1st-year college students accurately report their learning and development? *American Educational Research Journal, 47*(2), 466–496. https://doi.org/10.3102/0002831209353595.

Calvert, C. (2014). Developing a model and applications for probabilities of student success: A case study of predictive analytics. *Open Learning: The Journal of Open, Distance and e-Learning, 29*(2), 160–173. https://doi.org/10.1080/02680513.2014.931805.

Carless, D. (2007). Conceptualizing pre-emptive formative assessment. *Assessment in Education: Principles, Policy & Practice, 14*(2), 171–184. https://doi.org/10.1080/09695940701478412.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407. https://doi.org/10.1080/03075071003642449.

Coates, H. (2016). Assessing student learning outcomes internationally: Insights and frontiers. *Assessment & Evaluation in Higher Education, 41*(5), 662–676. https://doi.org/10.1080/02602938.2016.1160273.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312.

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347–360. https://doi.org/10.1080/02602938.2015.1111294.

Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment, 9*, 61–69.

Dimitrov, D. M., & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159–165.

Edwards, C. (2017). *Understanding student experience in the age of personalised study*. Paper presented at the ALT online winter conference 2017, University of Oxford. Retrieved from http://oro.open.ac.uk/53745/.

Ferguson, R., Brasher, A., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., et al. (2016). Research evidence of the use of learning analytics; implications for education policy. In

R. Vuorikari & J. Castano-Munoz (Eds.), *A European framework for action on learning analytics* (pp. 1–152). Luxembourg: Joint Research Centre Science for Policy Report.

Ferguson, R., Barzilai, S., Ben-Zvi, D., Chinn, C. A., Herodotou, C., Hod, Y., et al. (2017). *Innovating pedagogy 2017: Open University innovation report 6*. Milton Keynes: The Open University.

Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. *New Directions for Institutional Research, 127*, 73.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809.

Higher Education Commission. (2016). From bricks to clicks. In X. Shacklock (Ed.), (pp. 1–76). London: Higher Education Commission.

Hillstock, L. G., & Havice, P. A. (2014). Exploring characteristics of retained first-year students enrolled in non-proximal distance learning programs. *Journal of College Student Retention: Research, Theory & Practice, 15*(4), 575–603. https://doi.org/10.2190/CS.15.4.f.

Johnson, J. (2015). *Fulfilling our potential: Teaching excellence, social mobility and student choice*. (Cm9141). London: Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474266/BIS-15-623-fulfilling-our-potential-teaching-excellence-social-mobility-and-student-choice-accessible.pdf.

Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of gendered performance difference in introductory STEM courses. *arXiv preprint arXiv:1608.07565*.

Li, N., Marsh, V., & Rienties, B. (2016). Modeling and managing learner satisfaction: Use of learner feedback to enhance blended and online learning experience. *Decision Sciences Journal of Innovative Education, 14*(2), 216–242. https://doi.org/10.1111/dsji.12096.

Li, N., Marsh, V., Rienties, B., & Whitelock, D. (2017). Online learning experiences of new versus continuing learners: A large scale replication study. *Assessment & Evaluation in Higher Education, 42*(4), 657–672. https://doi.org/10.1080/02602938.2016.1176989.

McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). *Learning gain in higher education*. Cambridge: HEFCE/RAND.

Meadows, M., & Billington, L. (2005). *A review of the literature on marking reliability*. Unpublished AQA report, National Assessment Agency, London.

Mittelmeier, J., Long, D., Melis Cin, F., Reedy, K., Gunter, A., Raghuram, P., & Rienties, B. (2018). Learning design in diverse institutional and cultural contexts: Suggestions from a participatory workshop with higher education leaders in Africa. *Open Learning, 33*, 250–266. https://doi.org/10.1080/02680513.2018.1486185.

Moxley, J. M., & Eubanks, D. (2015). On keeping score: Instructors' vs. students' rubric ratings of 46,689 essays. *WPA: Writing Program Administration, 39*, 53–80.

Nguyen, Q., Rienties, B., & Toetenel, L. (2017a). *Unravelling the dynamics of instructional practice: A longitudinal study on learning design and VLE activities.* Proceedings of the seventh international learning analytics & knowledge conference, Vancouver, Canada.

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, F., & Whitelock, D. (2017b). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior, 76*(November 2017), 703–714. https://doi.org/10.1016/j.chb.2017.03.028.

O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education, 9*(3), 325–335. https://doi.org/10.1080/1356251042000216642.

Phillips, E. D. (2013). Improving advising using technology and data analytics. *Change: The Magazine of Higher Learning, 45*(1), 48–55. https://doi.org/10.1080/00091383.2013.749151.

Popov, S. V., & Bernhardt, D. A. N. (2013). University competition, grading standards, and grade inflation. *Economic Inquiry, 51*(3), 1764–1778. https://doi.org/10.1111/j.1465-7295.2012.00491.x.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). *A user's guide to MLwiN*. University of Bristol: Centre for Multi-level Modelling.

Richardson, J. T. E. (2013). Approaches to studying across the adult life span: Evidence from distance education. *Learning and Individual Differences, 26*(0), 74–80. https://doi.org/10.1016/j.lindif.2013.04.012.

Richardson, J. T. E., Alden Rivers, B., & Whitelock, D. (2015). The role of feedback in the under-attainment of ethnic minority students: Evidence from distance education. *Assessment & Evaluation in Higher Education, 40*(4), 557–573. https://doi.org/10.1080/02602938.2014.938317.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333–341. https://doi.org/10.1016/j.chb.2016.02.074.

Rienties, B., Clow, D., Coughlan, T., Cross, S., Edwards, C., Gaved, M., et al. (2017a). *Scholarly insight autumn 2017: A data wrangler perspective*. Milton Keynes: Open University.

Rienties, B., Rogaten, J., Nguyen, Q., Edwards, C., Gaved, M., Holt, D., et al. (2017b). *Scholarly insight spring 2017: A data wrangler perspective*. Milton Keynes: Open University UK.

Rogaten, J., Rienties, B., & Whitelock, D. (2017). Assessing learning gains. In D. J.-t. Brinke & M. Laanpere (Eds.), *Technology enhanced assessment. TEA 2016. Communications in computer and information science* (Vol. 653, pp. 117–132). Cham: Springer.

Rytkönen, H., Parpala, A., Lindblom-Ylänne, S., Virtanen, V., & Postareff, L. (2012). Factors affecting bioscience students' academic achievement. *Instructional Science, 40*(2), 241–256. https://doi.org/10.1007/s11251-011-9176-3.

Toetenel, L., & Rienties, B. (2016). Analysing 157 learning designs using learning analytic approaches as a means to evaluate the impact of pedagogical decision-making. *British Journal of Educational Technology, 47*(5), 981–992. https://doi.org/10.1111/bjet.12423.

van Ameijde, J., Weller, M., & Cross, S. (2016). Designing for student retention: The ICEBERG model and key design tips. In D. Whitelock (Ed.), *Quality enhancement series*. Milton Keynes: The Open University.

Yorke, M., & Longden, B. (2008). *The first-year experience of higher education in the UK*. York: The Higher Education Academy.

Jekaterina Rogaten is a researcher in learning analytics at the Institute of Educational Technology at the Open University UK. Her main research interests are in learning analytics and in positive psychology with a particular focus on learning gains, academic performance, progress, assessment of educational interventions, approaches to studying, creativity, metacognition and emotion. Jekaterina worked on a number of National, European and International projects with over 50 universities looking in particular at students' progress, grade trajectories and various competences development.

Doug Clow was a senior lecturer at the Institute of Educational Technology at the Open University, UK. He currently works as an independent consultant. He is particularly interested in learning analytics: using the traces learners leave behind to improve their learning. His career at the OU over the years has enabled him to explore the affordances of new and advanced technologies, particularly in the area of learning and education, but also in healthcare and medicine, science, engineering, and a range of more social sciences.

Chris Edwards is a lecturer at the Institute of Educational Technology at the Open University, UK. He chairs the masters module, Openness and innovation in e-learning, and helps run the Online and Distance Education masters programme. His research centres around three areas: modelling student study pathways with graph databases to gain an advantage in understanding student experience; as a researcher on a large EU funded Project with 18 partner organisations, exploring the e-authentication of students in online assessment; empowering learners through enhancing their self-reflective approaches – currently within language learning.

Mark Gaved is a lecturer at the Institute of Educational Technology at the Open University, UK. He is a Data Wrangler, conducting evidence-based research to support OU faculties in better understanding their course performance and student progression. His research also includes exploration of virtual environments for distance based design education, community based networked learning, and the pedagogy of online and distance learning.

Bart Rienties is Professor of Learning Analytics at the Institute of Educational Technology at the Open University UK. He is programme director Learning Analytics and Learning Design within IET and head of Data Wranglers, whereby he leads of group of learning analytics academics who conduct evidence-based research and sense making of Big Data at the OU. As educational psychologist, he conducts multi-disciplinary research on work-based and collaborative learning environments and focuses on the role of social interaction in learning, which is published in leading academic journals and books.

Chapter 12
The Bi-directional Effect Between Data and Assessments in the Digital Age

Abelardo Pardo and Peter Reimann

Abstract Assessment and feedback, two important factors of any learning experience in higher education, are being significantly disrupted by the emergent ecosystem of increasingly technology-mediated learning environments. Digital learning experiences produce data sets with highly detailed accounts of the interactions among participants. This information brings an unprecedented potential to move the focus of assessment and feedback away from the result to the process by which such result is attained. This new focus may have a profound effect in the process to influence how students engage with their work, its comparison with an appropriate standard, and to increase their self-evaluative capacity. But at the same time these data sets pose substantial challenges on how to integrate their presence in the design, deployment and refinement of learning experiences. In this chapter, we describe the main elements that need to be considered to translate these rich data sets into actions and design aspects that achieve a positive effect in the student experience.

12.1 Introduction

The increasing mediation of technology in learning experiences has translated into widespread availability of detailed data traces. Almost every computer program supporting student learning has the potential of recording detailed logs of user interactions. These traces can be used to offer a more detailed understanding of the learning process and identify potential improvements (Macfadyen and Dawson 2010).

Despite the increasingly pervasive presence of data in learning, the connection between data and assessment remains largely under/unexplored. Intelligent tutoring systems provide fully automated support for assessment and teaching, but they

A. Pardo (✉)
UniSA STEM, University of South Australia, Adelaide, Australia
e-mail: Abelardo.Pardo@unisa.edu.au

P. Reimann
The University of Sydney, Centre for Resarch on Learning and Innovation, Sydney, Australia
e-mail: peter.reimann@sydney.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_12

typically require sophisticated models of the student, domain knowledge, and reasoning patterns (Reimann et al. 2013). However, as pointed out by Baker (2016), substantial research is needed to explore how data can enhance teacher intelligence, instead of trying to replace it.

Mediation through technology now produces data that can be used to observe how students collaboratively write a document, evolve and improve a design concept, or capture the result of a brainstorming session. Assessment should include both the collection of information about the learner, but also acting on this information in relation to the educational goals (Shute and Rahimi 2017). The increasing presence of technology in learning environments opens new avenues to provide measurements of the overall student experience (Shute et al. 2016). Thus, the design and deployment of learning environments in general and assessment in particular need to be revisited to explore how to better exploit the use of data to enhance the overall intelligence of the process.

This chapter explores two dimensions to revisit assessment in the presence of comprehensive data capturing. The first one assumes that technology may widen both the repertoire and scaling possibilities for assessment. The second dimension is based on the possibility of observing the process through which the evidence for assessment is created.

12.2 Background

Assessment is generally accepted as a ubiquitous part of the student experience, one with a significant influence on how students approach their learning experience. This importance has led to a significant body of knowledge about various aspects of assessment such as their theoretical underpinning, the production and collection of evidence, the methods to for interpreting performance, or their effect in student learning (Pellegrino 2018). Additionally, assessment poses various challenges related with aspects such as robustness, scalability, alignment with learning objectives, or lack of student understanding. Rust et al. (2005) proposed an assessment model based on a social constructivist approach with special emphasis on the alignment between assessment and the other components of a learning design—the clear definition of criteria and the creation of an effective feedback process. We posit that each of these elements can be revisited in the context of a technology-mediated learning environment and the availability of detailed data about the interactions occurring in such context.

The increasing presence of technology mediating learning experiences has transformed the context in which assessment takes place. Shute et al. (2016) point to the advancements in technology and learning sciences as two conditions that are prompting the community of experts to reconceptualize assessment as a whole. Advances in learning sciences now require considering the affective and emotional context in which assessment and feedback processes occur. Analogously, advances in technology have redefined how communication takes place among those

participating in a learning experience. Technological platforms create detail traces of the events occurring in these contexts that can be processed by software programs to detect frequently occurring patterns. These patterns can be assimilated to specific student behaviour and automatically adapt the learning environment or provide suggestions to learners. Far from reducing the already existing tension between efficiency and accuracy of assessment, these changes have instead increased the need to explore the trade-off between these two aspects. In short, there is mounting pressure to increase the accuracy and efficiency of how student success is measured; the use of data collected using technology mediation is one possible avenue to address it.

In a review of computer-based assessment in the context of elementary and secondary education Shute and Rahimi (2017) categorized the presence of technology into those used as a supplement in the classroom, those that are web-based, and those that are data-driven and continuous. This last category emerges as a consequence of the use of data-driven approaches to assessment. There have been comprehensive studies attesting the need to explore how assessments are used in so-called Technology-Rich Environments (TRE; Bennett et al. 2007) to measure aspects such as problem-solving skills. Technology is also blurring the differences between *assessment of learning* and *assessment for learning* (Bennett 2011). In the context of this chapter we will assume that a TRE is a learning environment that is highly mediated by technology **and** such mediation provides comprehensive records or data traces. These data traces can be used for attaining a measure of a learning facet and to act on that information, for example by providing students with additional information for attaining the learning goals.

In parallel with the appearance of TREs, the areas of Educational Data Mining and Learning Analytics have emerged, providing techniques to increase the understanding of learning processes and providing the knowledge to improve them (Berland et al. 2014). Educational Data Mining (EDM) proposes the use of algorithms to process the vast amount of data that is captured when learning occurs in TREs and use that information to guide quantitative research and practice. The distinctive aspect for assessment is the capturing of detailed data traces to gain further understanding of artefacts created by learners, emotional states during learning experiences, etc. For example, Heffernan and Heffernan (2014) proposed to design and deploy questions, answers, web-based videos and hints, to support students while learning mathematics. The data captured during the interactions allows for the study of how students react in the presence of difficulty, how often they ask for help, and their trajectory while attaining the learning goals. Another example of the possibilities of data capturing in TREs looks at the detection of affective states (boredom, confusion, frustration, etc.) using Natural Language processing tools while students interact with a math tutor (Slater et al. 2017).

Complementary to EDM, the area of Learning Analytics proposes the use of data about learning experience to increase the understanding of how learning occurs and to support their improvement. The focus is not only on the algorithmic part of the problem, but on how the use of data can be properly integrated in the ecosystem of design, deployment, student support, and ethical and privacy issues (see Lang et al. 2017 for a comprehensive description of the area). The use of data in this context

has proven to be useful to understand emerging social structures (e.g., Bakharia and Dawson 2011; Ferguson and Buckingham Shum 2012), detect and support students at risk (e.g., Krumm et al. 2014; Macfadyen and Dawson 2010; Waddington et al. 2016), or ethical and privacy issues (Drachsler and Greller 2016; Kitto et al. 2018; Prinsloo and Slade 2015). For the sake of argument in this document, the areas of Educational Data Mining and Learning Analytics will be referred as the single *educational analytics*.

The overlap between assessment and educational analytics is obvious: on the one hand, there is the need for assessments to be reconceptualized in learning contexts with substantial technology mediation. On the other hand, there is also a need to widen assessment beyond the attainment of learning goals (summative assessment, or assessment of learning) towards assessing the processes that lead to learning. For example, taxonomies proposed for the design of assessment items in e-learning environments (Scalise and Gifford 2006) still resort to the description of conventional items in the new context without accounting for the new knowledge that can be derived from captured data. More refined proposals such as the four levels of integration between technology and assessments (DiCerbo and Behrens 2012) consider the use of technology to accumulate information from a variety of digital activities and acknowledges dashboards as an example of that type of assessment. But the design of assessments that go beyond the accumulation of information and combine aspects such as usage patterns, engagement patterns, predictive models, etc. are not yet a reality.

Similarly, educational analytics methods are mainly focussed on how to capitalize on the vast amount of information collected, how to distil knowledge from this data, but the connection with robust assessment paradigms is still weak. Collecting detailed information about students allows for the personalisation of processes such as feedback (Pardo et al. 2018), but this area needs to deepen the study of how knowledge derived from data traces can be properly integrated in a learning design and inform assessment tasks.

We envision two areas through which this connection between assessment and educational analytics can be articulated. The first one is around the notion of new assessment techniques deployed at scale. Comprehensive data collection may offer instructors and learners the possibility of exploring assessment aspects that go beyond conventional boundaries such as courses, topics, or even degrees. Learning is becoming a life-long endeavour and comprehensive data collection may offer insights never observed before. Analogously, some assessment typically restricted to contexts with a small number of participants can now be scaled through technological support. Algorithms may support large communities of learners to assess their contributions, suggest adequate next steps, or recommend peers for interaction, typically restricted to small student cohorts.

The second area to connect assessment and educational analytics is to widen the focus of assessment from final artefacts to processes. Assessments typically focus on specific artefacts that are the result of learners engaging with a task. Technology and comprehensive data capture may pave the way to increase our understanding of these processes and perhaps evolve towards a *correct by construction* type of

assessment where learners are guided throughout the artefact-creation process to accomplish the learning goals.

12.3 Widening Assessment Repertoire and Scaling

ICT and educational analytics can extend the assessment repertoire in a number of ways: by (i) offering established assessments more efficiently and more frequently to large numbers of students; by (ii) extending the kinds of assessments; and by (iii) extending what gets assessed. The use of learning management systems (LMSs) has contributed greatly to extending the reach and frequency of assessments in Higher Education, thus making it easier to employ assessment as a tool for learning, not only of learning (Pellegrino 2018). Multiple-choice tests and many forms of assignments—for individual and group work—can be deployed with ease, across hundreds and thousands of students. Relevant technologies and processes, including MOOCs, are by now in place in higher education around the world. A main benefit of educational analytics lies in the potential to analyse relations between assessments across courses and over time (semesters, years), thus exploiting the potential of big data on learning. For example, Poquet et al. (2018) used data about how users participate in a discussion forum in a MOOC over ten iterations of the course. The results identified three patterns of engagement that appear consistently throughout the editions, and they are maintained even when the number of participants decreases significantly. The type of interactions in the forum also evolved towards a higher focus on tasks related to the course and patterns similar to conventional Q&A spaces. A further example of an analysis of relations across assessments is found in Chap. 11 (Rogaten et al. this volume).

Of particular relevance for the second type of widening is the use of ICT for integrating assessment with learning activities. The guiding vision is the TREs in which all relevant learning (inter-)actions are captured and can be interpreted as such because the environment is designed accordingly (Shute et al. 2016). A typical example is the patient simulator for medical education, described in (Blanchard et al. 2012). TREs are much less prevalent in Higher Education than learning management systems. Because of the time and effort required to develop such environments, which are always domain-specific, they get developed for learning domains that are fairly stable and highly relevant. The main role of educational analytics is similar to the first case: integration of data from different environments and courses and comparative analysis across specific applications. The latter task is made easier by storing data in repositories such as the CMU data shop (pslcdatashop.web.cmu.edu). This repository provides access to data sets stemming from the use of intelligent tutoring systems and facilitates the analysis of these data from a number of perspectives (Koedinger et al. 2010).

The third form of widening pertains to assessing properties of learning and the learners that have not been the subject of assessment in higher education, from self-regulation to graduate qualities and "21st century skills". The two main strategies to

get to measurements are the use of psychometric tools such as scales for self-regulation and the use of analytics to identify and track indicators in students' activities. For instance, Fincham et al. (2018) proposed the use of data capture about how students engage in a learning experience to first obtain indicators of *study sessions* and then use them to identify different study tactics and how they evolve over the duration of a course adopting a flipped classroom design strategy. Although there is not a clear notion of a *correct* or *incorrect* study tactic, the information extracted from the traces provides a nuanced account of how students approach a learning experience and the potential to support them more effectively throughout the process. The key elements to use these traces effectively are (1) the connection with relevant elements of the learning design, and (2) the derivation of actions that promote aspects already present in learning designs such as self-reflection, regulation, goal setting, etc. The first element, the connection with elements of the learning design requires an explicit relation between data traces and the type of interactions that are desired within a learning experience. For example, if learners are given a set of exemplars to analyse and then discuss, the data capturing process should be aligned with these two steps and identify which exemplars are being accessed and which events occur in the discussion space. The wide variety of possible scenarios makes this relation highly sensitive to the context, but when achieved it guarantees that the indicators identified by software programs can be interpreted within the context of a task in the design. The second element, the derivation of actions to support learners requires a tight integration between the conclusions derived from interpreting the traces and the type of actions. For example, elements within the learning context can be labelled as to support certain aspects and therefore be part of a set of *recommendations* automatically provided to the students when certain patterns are identified. A hybrid approach in terms of automation could require the intervention of the instructor (expert) to provide coaching advice on how to adopt a more appropriate learning strategy. The key element in this context is to align the comprehensive collection of data with existing models of what constitutes good learning and assessment practices.

In these scenarios educational analytics plays a role not only in integrating data sets into big data and performing comparative analyses but promoting new approaches for assessment design and engineering. This is because the student activities relevant for assessment take multiple forms and are distributed over multiple contexts. For instance, when assessing the development of collaboration skills, evidence for skill proficiency can consist of self-reports, tests, psychometric scales, and a range of observations captured in log files and in portfolios in a variety of databases. This task is made the more challenging as university teachers (other than those teaching in a liberal arts college perhaps) are by and large not experts in the development and assessment of student qualities that are not subject-matter specific. This and the fact that educational analytics practitioners are also not experts in—for a lack of better word—general pedagogy has contributed to an approach to assessment design that does not build on knowledge about the nature and the development of such capacities.

In this knowledge-lean approach, the intrinsically complex measurements of the quality of collaboration can be approximated through indicators that are mainly identified by way of data mining. For instance, assume we know students' scores on a psychometric scale for self-regulation and have all their data from the learning management system they use. In such circumstances, new indicators for self-regulation can be identified by correlating the scores from the scale with variables derived from the LMS data—such as regularity of contributions, adherence to deadlines, etc. An example of this combination of results was described by Ellis et al. (2017). The conceptual framework of student approaches to learning (Pintrich 2004) was combined with the Revised Study Process Questionnaire (R-SPQ) to collect data self-reported by the students (Biggs et al. 2001). The results were processed using Exploratory Factor Analysis and combined with data derived from interactions with resources in a blended learning environment. The authors show the increase in variance explained when both data sources are combined.

The knowledge-lean approach, while being practical, has a number of disadvantages, owing to its inductive-correlational nature. When statistical significance is used as the criterion, then the selection of indicators depends on the number of students; the size of correlations depends on variability (hence, processes that are necessary but show less variability will be missed); and including less or more variables might change the correlations significantly. Another disadvantage is that knowing that two variables correlate does not necessarily lead to actionable knowledge. For instance, raising student's adherence to timelines may not lead to higher self-regulation; it may as likely have no effect or a negative effect on self-regulation. Knowledge-rich approaches to assessment engineering, such as provided with the Cognitive Design Systems framework (Embretson 1998) are better suited to develop valid indicators. However, they require more effort and the combination of domain expertise with pedagogical and assessment expertise. A major advantage we see in their deployment is the closure of the gap between assessment and learning processes (Shute et al. 2016, p. 52).

12.4 Observing the Process

The second assessment aspect that will be significantly impacted by the ubiquity of data sets is the capacity to observe the process of creating artefacts for assessment. In typical assessment scenarios, learners are given the description of a task, a set of assessment criteria and they have to produce an artefact that is going to be assessed with respect to those criteria. A more holistic view of this process includes the deployment of feedback and the interplay of existing knowledge and beliefs, goal setting, strategies to achieve these goals, etc. Existing models such as, for example, self-regulated learning (Winne 1997, 2014) consider all elements in a common context that include explicitly the notion of task and performance (see Winne 1997, p. 399). But the existence of comprehensive data collection and analysis methods requires a reconceptualization of the relationship between learners, assessment and

feedback (Pardo 2018). A new landscape is emerging in which assessment tasks can be conceived as a learning trajectory that may include frequent interactions with multiple agents (human and non-human) informed by data collection. Various elements in this trajectory such as task descriptions, clarifications, relevant resources, or even strategic steps, may be adjusted or personalised depending on the captured contextual data. The increased relevance of this trajectory may be because data provides unprecedented levels of detail to be deployed at scale (large number of learners) and consequently may lead to more comprehensive learner support.

A representative example of the combination of technology, human intervention, and a tight integration with existing learning design is provided by reflective writing assignments (Gibson et al. 2017). Natural language processing technology was used to analyse the rhetorical moves of 120 submissions from 30 students. The study highlighted the need to interpret textual data in the context of the assignment and a theoretical framework to capture the structure of such texts. But once these indicators are obtained, it was equally important to frame the feedback in adequate terms to effectively support learners towards improvement of their submissions.

In these contexts, assessment designers not only need to contemplate the production of criteria to evaluate the final artefact, but also need to develop rules to analyse the indicators derived from the creation process, relate these observations to actions and deploy them in the assessment scenario. This new way of assessment design is complex. The process of connecting data with actionable items is influenced by the learning design and heavily mediated by the learner's affective reactions, motivation, self-regulation, etc.

Additionally, the use of indicators derived from data traces requires us to reconsider the notion of validity of assessment instruments (Hickey et al. 2000; Maier et al. 2016). An assessment instrument is considered valid when it provides a sound distinction between different levels of attainment of the learning goals. Indicators derived from captured data are highly unlikely to provide such level of validity. Typically, the indicators will *correlate* with other aspects of the learning experience, thus providing only a partial or approximate measure of the process. In other words, indicators derived from observing how students engage with an assessment task will contain an inherent level of uncertainty (Macfadyen and Dawson 2012).

Assessment criteria need also to be revisited when considering the shift of focus towards the process to create the artefact. Under this new lens, scoring the final outcome is just one out of multiple elements of the assessment process. Criteria now need to morph into more sophisticated statements describing the aspects to observe during the process of creating an artefact, the way observations are combined to derive insights, the actions that are derived from these insights, and how such actions are deployed in the learning experience. Technology mediation opens the possibility of reconsidering the overall process as a continuously iterative reflective loop involving learners and instructors that simultaneously promotes the capacity of learners to assess their work, but also to achieve the desired outcomes. We envision the presence of technology as the catalyzer of assessments that blur the distinction between formative and summative.

This shift in focus is even more relevant when trying to assess the so-called *higher order skills*. Although arguably these skills have always been desirable and promoted in traditional learning experiences, the existence of data extracted from technology mediation widens the range of indicators to assess them. For example, concept or mind mapping techniques are typically used to promote critical and analytical skills (Davies 2011). There is a wide variety of software tools that help students create visual representation of concepts and relations. Observing students throughout the process of constructing these diagrams offers the opportunity to support them towards the creation of useful maps.

12.5 The Effect on Strategies for Assessment Design

The new conceptualisation of assessments in settings with comprehensive data capturing influences design and deployment. There are various aspects when designing an assessment that can be reconsidered in the presence of data. We envision a bi-directional influence between data and how assessments are designed. One of these design aspects is a potentially wider consideration of learning goals. Rather than stating the learning goals, an assessment may also identify the skills, strategies or attitudes that would facilitate the attainment of such goals. The data captured in relation to these skills may then be available to instructors and learners to increase their awareness of the whole process. This relation is an example of how assessment definition may influence how data is captured.

Analogously, if learning experiences are deployed in contexts that provide detailed observations of elements relevant to the attainment of the learning goals, the assessment may be reframed to take these observations into account. For example, if a platform allows student to specify their goals and strategic steps to tackle a problem, the assessment may be redesigned to include the provision of suggestions or support actions to take full advantage of this feature. Researchers are already obtaining increasingly reliable measures of learner affect (Bosch et al. 2015); assessments can now be designed to take these observations into account and personalise their form depending on these values. For example, a student known to have *test anxiety* may be identified in advance and given a combination of alternative tasks to obtain the required valid measure of attainment. Identifying this type of aspects undoubtedly has ethical ramifications that need to be taken into consideration when assessing the type of indicators as well as the conclusions derived from these observations. But the relationship between how assessments are conceived and analysed is bi-directional. Obtaining detailed accounts of how assessments are delivered may prompt reflection about features of the learning design. For example, identifying a large percentage of students with test anxiety may prompt a reconsideration of the assessment items in a design.

An additional aspect to be considered is how assessment results are reported. As discussed in the previous section, the indicators derived from the data captured in a learning environment may offer a partial view of how learners engage with a task,

but even in the presence of uncertainty, and being fully aware of their limitations, they could also be part of the assessment result that are distributed among stakeholders. This modified report containing indicators would also contribute to increase the transparency of the overall procedure as a mechanism to increase its reuse and adaptation to other contexts.

12.6 Conclusions

The availability of data captured by technology when mediating a learning experience is an element that is transforming numerous aspects of education. But as is already happening in other areas, the presence of data is not enough to produce tangible improvements. Data needs to be analysed, situated in the proper context, and connected with existing elements in a learning design. Assessment is no exception to this observation, quite the contrary. This chapter has explored two aspects of assessment that are heavily influenced by this change. The first one is that data availability is clearly widening the repertoire of assessment tasks that are feasible, but also the scale at which these can be deployed. Massive Open Online Courses are examples of how assessments can be shaped and the trade-offs that emerge under these demands. Systematic data collection allows for the creation and refinement of models to characterise how students participate in learning experiences. These models can be used to predict goal attainment, detect strategies, and ultimately provide more effective student support.

The second aspect that will be significantly influenced by the presence of data is the capacity to observe the process leading to the production of the artefact subject to assessment. This shift in focus away from the finished product and to the process leading to it opens the possibility to consider a holistic view including student prior knowledge, beliefs, goals, strategies, affect, etc. Although at the risk of significantly increasing their complexity, new assessment tasks may include criteria and instructions to extract insights from existing data and deploy actions to support students towards goal attainment.

Although these two aspects paint a promising horizon of improvements, there are numerous caveats that are intrinsic to these approaches need to be carefully considered. Although technology may provide the aforementioned detailed data traces, they may still not be enough to describe what exactly is happening in a learning experience. The best approach to account for this limitation is to frame this as an *approximation problem*. Technology is now offering us the possibility to have a more approximate idea of how learning experiences unfold, increase our level of understanding and therefore our possibility to improve them. This increase in perception cannot be derived from the mere presence of detailed data. Context is essential, thus the need for a solid connection between data, analytics methods and the learning design. Such connection is not trivial to establish and may significantly raise the overall complexity of the problem.

Aside from these caveats, the deployment of any paradigm that involves comprehensive collection of personal data, needs to be implemented under the strictest rules to observe privacy, ethics and transparency. These aspects not only apply to how data is captured and managed, but they need to be considered in the early stages of the design. The potential improvements in learning experiences cannot be an excuse to ignore these aspects.

In addition to these aspects, as with any context in which the use of data is introduced as a key element, it is desirable for all stakeholders to have solid data literacy skills. If learners receive detailed accounts of their progress, they need to be able to interpret the data and extract useful knowledge. These skills are even more important in the case of instructors as they are typically required to interpret data in the context of the overall application and assess both individual and population measures. The most powerful educational analytics solution would have a significantly reduced impact in the hands of stakeholders with poor data literacy skills.

As all these elements point out, the irruption of data in the assessment space has the potential of improving learning experiences, but at the same time requires a reconceptualization of the design procedures in which conventional assessment design techniques and the presence of data are continuously influencing each other.

References

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education, 26*(2), 600–614. https://doi.org/10.1007/s40593-016-0105-0.

Bakharia, A., & Dawson, S. (2011). *SNAPP: A bird's-eye view of temporal participant interaction.* Paper presented at the 1st international conference on learning analytics and knowledge, Banff, Canada.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594x.2010.513678.

Bennett, R. E., Persky, H., Weiss, A. R., & Jenkins, F. (2007). *Problem solving in technology-rich environments. A report from the NAEP technology-based assessment project, research and development series (NCES 2007-466)* (p. 196). Washington, DC: National Center for Education Statistics.

Berland, M., Baker, R., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1), 205–220. https://doi.org/10.1007/s10758-014-9223-7.

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133–149. https://doi.org/10.1348/000709901158433.

Blanchard, E. G., Wiseman, J., Naismith, L., & Lajoie, S. P. (2012). *A realistic digital deteriorating patient to foster emergency decision-making skills in medical students*. Paper presented at the IEEE international conference on advanced learning technologies, Rome, Italy.

Bosch, N., D'Mello, S., Baker, R. S., Ocumpaugh, J., Shute, V. J., Ventura, M., … Zhao, W. (2015). *Automatic detection of learning centered affective states in the wild*. Paper presented at the 20th international conference on intelligent user interfaces, Atlanta, Georgia.

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: What are the differences and do they matter? *Higher Education, 62*(3), 279–301.

DiCerbo, K. E., & Behrens, J. T. (2012). *From technology-enhanced assessment to assessment-enhanced technology*. Paper presented at the National Council on Measurement in Education 2012 Annual Meeting, Vancouver, Canada.

Drachsler, H., & Greller, W. (2016). *Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics*. Paper presented at the 6th international conference on learning analytics & knowledge, Edinburgh, UK.

Ellis, R. A., Han, F., & Pardo, A. (2017). Improving learning analytics – Combining observational and self-report data on student learning. *Educational Technology & Society, 20*(3), 158–169.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3*(3), 380–396.

Ferguson, R., & Buckingham Shum, S. (2012). *Social learning analytics: Five approaches*. Paper presented at the 2nd international conference on learning analytics and knowledge, Vancouver, Canada.

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2018). From study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies, 12*, 59–72. https://doi.org/10.1109/TLT.2018.2823317.

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). *Reflective writing analytics for actionable feedback*. Paper presented at the 2nd international conference on learning analytics and knowledge, Vancouver, Canada.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientist and teachers together for minimal invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education, 24*(4), 470–497.

Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. H. (2000). Assessing learning in a technology-supported genetics environment: Evidential and systemic validity issues. *Educational Assessment, 6*(3), 155–196. https://doi.org/10.1207/s15326977ea0603_1.

Kitto, K., Buckingham Shum, S., & Gibson, A. (2018). *Embracing imperfection in learning analytics*. Paper presented at the 8th international conference on learning analytics and knowledge, Sydney, Australia.

Koedinger, K. R., Baker, R. S. J. D., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. In C. Robero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of educational data mining* (pp. 43–56). Boca Raton: Chapman & Hall/CRC.

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A learning management system-based early warning system for academic advising in undergraduate engineering. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 103–119). New York: Springer.

Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds.). (2017). *Handbook of learning analytics*. Vancouver: Society for Learning Analytics Research.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008.

Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society, 15*(3), 149–163.

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education, 95*, 85–98. https://doi.org/10.1016/j.compedu.2015.12.002.

Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education, 43*(3), 428–438. https://doi.org/10.1080/02602938.2017.1356905.

Pardo, A., Jovanović, J., Dawson, S., Gašević, D., & Mirriahi, N. (2018). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*, 128–138. https://doi.org/10.1111/bjet.12592.

Pellegrino, J. W. (2018). Assessment of and for learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 410–421). New York: Routledge.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x.

Poquet, O., Lim, L., Mirriahi, N., & Dawson, S. (2018). *Are MOOC forums changing?* Paper presented at the 8th international conference on learning analytics and knowledge, Sydney, Australia.

Prinsloo, P., & Slade, S. (2015). *Student privacy self-management: Implications for learning analytics*. Paper presented at the 5th international conference on learning analytics and knowledge, Poughkeepsie, USA.

Reimann, P., Kickmeier-Rust, M., & Albert, D. (2013). Problem solving learning environments and assessment: A knowledge space theory approach. *Computers & Education, 64*, 183–193. https://doi.org/10.1016/j.compedu.2012.11.024.

Rogaten, J., Clow, D., Edwards, C., Gaved, M., & Rienties, B. (This volume). Are assessment practices well aligned over time? A big data exploration. In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world* (Chapter 11). Dordrecht: Springer.

Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education, 30*(3), 231–240. https://doi.org/10.1080/02602930500063819.

Scalise, K., & Gifford, B. (2006). Computer-based assessment in E-learning: A framework for construction "intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning, and Assessment, 4*(6).

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*(1), 1–19. https://doi.org/10.1111/jcal.12172.

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. (2016). Advances in the science of assessment. *Educational Assessment, 21*(1), 34–59. https://doi.org/10.1080/10627197.2015.1127752.

Slater, S., Ocumpaugh, J., Baker, R., Almeda, M. V., Allen, L., & Heffernan, N. (2017). *Using natural language processing tools to develop complex models of student engagement*. Paper presented at the 7th international conference on affective computing and intelligent interaction, San Antonio, USA. https://doi.org/10.1109/ACII.2017.8273652.

Waddington, R. J., Nam, S., Lonn, S., & Teasley, S. D. (2016). Improving early warning systems with categorized course resource usage. *Journal of Learning Analytics, 3*(3), 263–290. https://doi.org/10.18608/jla.2016.33.13.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89*(3), 397–410.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning, 9*(2), 229–237. https://doi.org/10.1007/s11409-014-9113-3.

Abelardo Pardo is Professor and Dean of Programs at UniSA STEM at the University of South Australia. His research interests include the design and deployment of technology to increase the understanding and improve digital learning experiences. More specifically, his work examines the areas of learning analytics, personalized active learning, and technology for student support.

Peter Reimann is Professor in the Sydney School of Education and Social Work at the University of Sydney where he co-directs the Centre for Research on Learning and Innovation (CRLI). His primary research area is the study of learning with emphases on educational computing, multimedia-based and knowledge-based learning environments, e-learning, and the development of evaluation and assessment methods for the effectiveness of computer-based technologies. Peter is

well-known for his work on learning by self-explanation, on augmentation approaches to support computer-mediated group learning, and for his contributions to computer-based analysis of temporal data. His recent book *Measuring and visualising learning in the information-rich classroom* has been published by Routledge.

Chapter 13
Standards for Developing Assessments of Learning Using Process Data

Sandra Milligan

Abstract Digital technology is changing assessment of learning. Digitised assessment can be more administratively efficient, more easily scaled, more effectively targeted to individual levels of performance, more integrated into the learning environment, more interactive, and it can support more imaginative, colourful, interactive and timely feedback. However, in this chapter, it is argued that 'more, faster and prettier' is only part of the assessment story of the first quarter of the twenty-first century. Education institutions are also being pressed to make distinctive shifts in what is learned and thus in what is assessed. Students now need to establish mastery of complex learning outcomes that extend beyond the cognitive domain, and beyond mastery of content knowledge, to mastery of competence and skill, including soft skills, or general capabilities. This chapter explores this assessment frontier, examining whether and how large quantities of digital, process-oriented data generated from learning management systems and other digital learning tools can be used to make reliable and valid judgments about the degree to which student have mastered complex general capabilities. It is argued that 'metrolytic' standards for development of assessment tools can be applied to ensure the requisite validity and reliability.

13.1 Introduction

Shifts in thinking about assessment are not new. Assessment practice has always been responsive to the educational and social concerns of the day (Pellegrino 1999). For instance, the contemporary methods of psychometrics can be traced back to the interest in individual differences of anthropologists and eugenicists in the nineteenth century. The statisticians of the day, including Fisher, Spearman, and Pearson developed methods to manage evidence of individual attributes, many of which are still routinely applied. The multiple choice question was introduced in the early

S. Milligan (✉)
Assessment Research Centre, University of Melbourne, Parkville, VIC, Australia
e-mail: s.milligan@unimelb.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_13

twentieth century, in the attempt to improve objectivity and fairness into wide-scale assessments of applicants, to rank them reliably for positions in the US military. During the 1980s and 1990s, community demand for fairer selection methods, and greater accountability for student learning resulted in refinement of standardized testing methods. Statistically sophisticated methodologies were used with automated administration in support of large-scale ranking and monitoring, usually focused broadly on assessing 'scholastic aptitude', or mastery of basic learning like literacy, or numeracy.

Today, digital tools used in university teaching are visibly changing assessment practice, and the character of the relationship between assessor and assessed. Teachers have embraced digitally mediated tools to set assessments, scaffold student response to them, monitor cheating, harvest responses, mark responses, provide feedback, grade, compile and report. Digitisation of assessment procedures is administratively efficient, allows for scaling to accommodate large classes, allows more effective targeting to individual levels of student performance, is more interactive, and it can support more imaginative, colorful, interactive and timely feedback and faster more direct reporting. Digital scaffolding of assessment makes feasible assessment methods that require complex administration, including for example, peer and self-assessment (see Tai and Adachi, Chap. 15, in this volume). Embedding assessment into the learning environment of digital learning management systems allows better integration of teaching and learning and supports formative assessment practices. Digital mediation of teaching and assessment via ubiquitous learning platforms is now common, even in small on-campus classes. These forms of technological enhancement have arguably enabled assessment processes to become more efficient, faster and prettier, more responsive, more formative, more timely.

However, a further frontier for assessment is emerging that goes beyond technological enhancement of teacher assessment practice, to exploring the use of powerful digital technologies and digital data, particularly process data, to better assess and report on the growth in learning, particularly learning of complex competencies and general capabilities. This chapter focuses on this frontier, describing the difficulties associated with generating valid and reliable assessments. It argues that new standards for development of assessment tools—metrolytic standards—are required. These standards draw on methods used in the learning analytics community, combined with those commonly used in educational measurement, to provide a framework designed to ensure that any assessment can be trusted to have the required level of reliability and validity to warrant use for educational purposes.

13.2 Contemporary Pressures on Assessment Practice

The frontier arises directly from contemporary pressures on universities to change what students learn and how they learn it, which consequently changes what teachers assess and report (Griffin and Care 2015). The core idea is that the '4th industrial revolution' is in train, requiring educators to produce learners with different skills

and capabilities than required by generations past (Tremblay et al. 2012; OECD 2018; Milligan et al. 2018). No single, simple driver can be identified. A range of factors are involved, including the ubiquity of digital communications and computing technology, the rapid expansion of knowledge, the impact of globalisation, and the increasing commitment to ensuring sustainable, equitable development for human wellbeing. All of this is redefining how we live and work in the twenty-first century. The net effect on education institutions is that they are being pressed to redefine what learners need to know, and to produce graduates who can demonstrate attributes other than just mastery of knowledge in a content domain. Students are now required to demonstrate *knowhow* in a domain of study, as well as content mastery. Curricula are being altered to supplement the cognitive outcomes of traditional disciplines with requirements that learners develop the constellations of knowledge, values, attitudes, skills and beliefs required for competent performance in any field (Dreyfus and Dreyfus 1980). This is sometimes characterised as a shift from *content* to *competence* in curricula (Griffin 2007). In this context, curriculum outcomes are extended beyond mastery in discipline or professional domains, to encompass attributes variously referred to as 'soft' or '21st Century skills' or 'transversal skills', or (as referred to in this chapter) 'general capabilities', that might be applied in any field or domain (Asia-Pacific Education Research Institutes Network 2015; Griffin and Care 2015). For instance, the World Economic Forum (2015) described a range of required general learning capabilities, including critical thinking skills, communication, creativity, collaboration, scientific and ICT literacy, persistence and curiosity, amongst others.

A related shift requires that students develop skills as lifelong learners (Bransford et al. 2003). The argument is that it is not enough that students can learn when directed by teachers, in formal educational settings. They must also be able to learn by their own initiative. In modern times, this ability to learn is not considered to be a matter of IQ or any innate ability – it is more to do with the mastery of a set of knowledge, skills, understandings and beliefs about learning that equip individuals to a greater or lesser degree with the capability they need to learn (Milligan and Griffin 2016).

Formal inclusion of these general capabilities in curricula represent a professional challenge for teachers, especially in courses which use traditional approaches to higher education assessment, such as summative assessments made at the end of course, based on the evidence presented to a teacher via student essays. Now, the challenge is to assess the degree to which a learner has also mastered general capabilities.

Assessment of this sort is several degrees more difficult than assessment of mastery of content knowledge in the cognitive domain. It is a new field for teachers, and is especially challenging in large classes where teachers do not necessarily know their students. Assessment tasks that enable individuals to demonstrate complex capabilities are usually themselves complex, and often need to be conducted in non-standardized environments, involving performances or building of artifacts, or working with peers, sometimes in teams. Mastery of these capabilities usually takes time and practice, often in 'authentic' learning environments unlike traditional

lecture halls. Feedback on performance is required at various stages of development, enabling learners and other stakeholders to chart the learner's gradual increase in mastery. Thus, in parallel with changes in curriculum, there are requisite changes in the purposes and methods of assessment. Assessments need to assist learners and teachers to determine the degree to which a learner has mastered complex general capabilities in a domain of learning. Important methodological frontiers for assessment are opening in this area.

13.3 The Promise of Big Data to Assess General Capabilities

Scardamalia et al. (2013) reported on the findings of a large, international research project established to explore how best to assess these general capabilities. They concluded that the best assessments are "embedded in the technologies used in the learning environment, concurrent and transformative" (p. 34). They pointed out that embedded technologies can generate automatic feedback, provide on-demand assessments, and prevent or reduce the separation of assessment from the learning experience. Response rates are unproblematic, as participant activities are always reflected in the log stream, and data collection requires no additional effort by participants. Use of digital traces of learner activity for assessment purposes has the potential to provide real-time calculation of scores, and could provide greater timeliness in calculation and feedback to participants during a course.

On the face of it, this approach seems feasible. There are copious quantities of learning-related digital information now available, including click stream data that captures every mouse click, swipe, or keyboard action of every learner as they use digital learning applications. Other information can be obtained from sophisticated digital data sensors in classrooms that capture anything from eye-gaze direction to heart-rate, from speech to physical movement. The capacity to 'see' what students say, do, make, or write in learning environments is vastly enhanced. Available data goes well beyond the inputs to traditional university assessments, such as observations by teachers in classrooms, or responses to assessment tasks, or results from standardized tests. Data now systematically captures information on the *process* of learning not just the outputs from it.

Such data can also be interrogated for meaning by a plethora of modern analytical methods, including, for example social network analysis, text analysis and various forms of data mining. These enable the construction of statistics that can, in theory, be used as indicators of performance, for assessment purposes. Teachers can 'see' the degree of connectedness to others in the class (from network analysis), the focus of a student's interests (from text analysis), the systematicity of study habits (from time-series analyses) and so on. Techniques of data mining and artificial intelligence can be applied to these indicators (He et al. 2016) potentially adding value. Data or analyses of this sort are now routinely presented in digital dashboards, or otherwise provided as feedback to teachers and learners (Corrin and DeBarba 2014).

Researchers early into the learning analytics field (Carmean and Mizzi 2010; Gasevic et al. 2015; Siemens and Long 2011; Greller and Draschler 2012) expected that such data would have many benefits for learners and teachers, making visible the process of teaching and learning, supporting reflection on practice by learners (on their learning) and by teachers (on their teaching); predicting and modeling learning, leading to better intervention; and making possible personalization of learning through real-time tracking and analysis of each individual. If backed with artificial intelligence tools, it is argued, digital responders might become more capable than their human predecessors in assessing learning.

This optimism was supported by belief that use of big data was not only possible, but it is the *preferred* option when it comes to assessing individuals' performances in the kind of complex general capabilities which are now the focus of curriculum (Scardamalia et al. 2013). The traditional techniques for assessing individuals' attributes or capabilities include use of self-report scales, direct observation by experts, employing think-aloud protocols, analysing artifacts like respondent diaries, and using micro-analytic approaches involving codification of behaviours such as eye-gaze or facial expression to infer an individual's level of an attribute (Cleary et al. 2012). But, such techniques are impractical for use in real learning environments. They are too costly and labour-intensive, leading teachers and assessors to look for better, more practical approaches. Naturally, they look to the possibility of using big, digital data, derived from sensors embedded in the learning environment, that provide systematic evidence about the process of learning used by learners.

13.4 The Need to Ensure Validity and Reliability

While the optimism is high, there are acknowledged difficulties. Scholars of learning analytics have always been quick to point out that that bigger digital data generated as a by-product of learning is not always better data (Greller and Draschler 2012; Siemens and Long 2011). A key question not yet convincingly answered is whether or not these digital traces can be used to construct indicators of learning, or whether they merely record processes that *may or may not* reflect the degree of learning attained. In addition, it is not clear whether the process data embodies sufficient information: perhaps missing elements may be exactly those required to explain learning. Platforms or digital sensors cannot capture all 'off-line' activity such as reflection or note taking, or what student are thinking, but these missing elements might be vital (Gunnarsson and Alterman 2013). Big natural databases tempt the search for interesting relationships based on correlations, or factor and cluster analysis. When interesting patterns are found, as they are bound to be, and if they attain statistical significance, the temptation is to impute explanatory value, and to infer meaning about learning. But it is not at all clear that these interesting, statistically significant patterns are a suitable basis on which to judge an individual's learning. The patterns may in reality be products of chance, or be inconsequential for learning, and may not evidence an underlying phenomenon with explanatory

value. Statistical relationships demonstrate only that discovered relationships are unlikely to be random, which is an insufficient basis for interpreting any numbers as measures of learning for an individual.

The most robust of the contemporary work using process data to assess and report on the growth in complex competencies and general capabilities uses a combination of analytics methods and the methods of educational measurement (Buckingham Shum and Deakin Crick 2016; Griffin and Care 2015; He et al. 2016; Milligan and Griffin 2016; Shute and Ventura 2013; Wilson et al. 2016; Polyak et al. 2017). Measurement principles and techniques such as those that underpin Wilson's (2005) Constructing Measures approach, or the Evidence Centered Design approach (Mislevy and Haertel 2006) are designed to engender trust in assessments by ensuring that scores measure something of value, to a requisite standard. A careful, methodical, iterative, curriculum-focused process is used to develop scores for individuals, involving the development of constructs and evidence maps, sampling of evidence using rules and procedures, and so on. The measurement sciences establish standards that evidence claims that measures are appropriately used to assess learning – that is, they are valid and reliable, and can safely be used to judge an individual's learning progress.

It is notable that researchers who have adapted traditional measurement techniques to analytics-based process data tend to be cautious, reflecting a growing awareness in the assessment and analytics communities that such work is still in its infancy. The key to understanding the difficulties associated with this frontier is that, when constructing measures of complex constructs from digital data, assumptions crucial to quality of assessment need to be made explicit, and tested (Wright and Masters 1982). For instance, assessment of learning is based on an assumption that individuals can possess different amounts of an attribute being assessed, and that a description of that attribute should provide the basis of assessment. The underlying attribute must make sense and be plausible, and there should be practical benefit in assessing it. The attribute must have dimensionality and it must be possible to understand how it is that people can have more or less of it. 'More' or 'less' must be capable of consistent representation as a progression for all individuals, using units of equal value, and the units should be additive and repetitive. Although attributes cannot be directly observed, it must be possible to understand how different levels of performance is explained by differences in observable behaviour of individuals: what individuals do, or say, or make, or write (Glaser 1994a, 1994b). The behavioural differences must have explanatory value, and it should be possible to infer from these observed differences in behaviours the amount of the attribute that a person has. Such assumptions need to be tested and evidence and argument presented to satisfy stakeholders that the use of assessments for educational purposes are warranted. An important source of thinking about the standards that should apply can be found in discussion of validity in measurement science (Cronbach and Meehl 1955; Kane 2013; Messick 1995; Wolfe and Smith 2007) and in discussion of quality in learning analytics (Dringus 2012; Greller and Draschler 2012).

Table 13.1 outlines a set of indicative 'standards' that might be used to interrogate the quality of any score, before it is used to make decisions about an individual. These standards are derived from both the practice of measurement science and learning analytics. They are best described as metrolytics standards, derived from the Greek words *metron*, which is the root of the word measurement and means limited proportion, and *analutikós*, meaning to analyse. In an ideal world, the standards would provide the basis for presentation by an assessment designer of evidence and argument to support the validity and reliability of any assessment constructed from process data, in exactly the same way that high-stakes test developers are required to present arguments as to the validity and reliability of their tests. A set of indicative standards is presented in Table 13.1.

Recently, the learning analytics community expressed concern about the burgeoning array of analytics apps, and whether or not there is a sufficient evidence base about the use of analytics to warrant trust by stakeholders (Bergner et al. 2017; Ferguson and Clow 2017). The concerns intensify when the results are used to control or shape the treatment a person receives, as is often the case in assessment of learning. It would not be unreasonable for stakeholders (learners, teachers, employers, professional associations), to mistrust that an instrument can assess complex attributes, especially if it combines a variety of data types and utilizes complex data transformation or algorithms.

Adoption of metrolytic standards provides one way to address this issue, requiring methods that ensure that an assessment has utility for the intended purpose. The standards provide the framework for presenting the evidence and argument to engender trust in the assessments. In this, it is important that evidence presented should not merely defend the interpretation of an assessment. Rather, it should be capable of convincing a reader that the assessment design or method tested the assumptions on which it was based, and that no other plausible alternative interpretations of what is being assessing are supported. This involves imaginatively identifying and examining potentially disconfirmatory evidence as well as confirmatory evidence, to address what are the risks that could apply to a particular assessment.

13.5 Methodological Frontiers

Consideration of the metrolytics standards outlined in Table 13.1 highlights some of the practical difficulties facing analysts attempting to build reliable and valid assessments of complex general capabilities using process data. For instance, the standard requires as a prerequisite clarity about what is to be assessed. In traditional classrooms, what is to be assessed is usually operationally defined as 'what has been taught', often content-based. Assessments of the newer general capabilities require a similarly clear view of what is being assessed. In practice, this demands specification of the

Table 13.1 Indicative metrolytics standards for assessments derived from digital process data

Utility There is a clear purpose for making assessment(s) that are of value to stakeholders.
There is clarity about the nature of the attribute There exists a clear definition of the attribute being assessed, expressed as the constellation of knowledge, understandings, skills, beliefs, attitudes, and values required for mastery, differentiated at each level of mastery. This definition should be understandable and acceptable to stakeholders, including teachers and learners.
The attribute has dimensionality It is plausible to assume that individuals have more or less of the attribute of interest, and can be arrayed along an underlying continuous scale from more to less. Ideally, the scale reflects typical learning trajectories of learners' progress from level to level as their mastery grows. These trajectories, typically called 'progressions' or 'learning continua' must make sense to learners and teachers alike.
Data relates to learning behavior Selected data comprises representations of what learners say, do, make or write when learning. Data does not include characteristics or status of individuals that might be related to learning, (such as dispositions or social or demographic features) but which do not reflect learning gains in a particular environment.
Process data is 'clean' and understood Data husbandry methods are used, including, for instance: checking plausibility of the range and distribution of values; transforming raw log stream data into variables, counts or categories; identifying and mitigating corrupt, incomplete, misleading or miscoded data; ensuring that data definitions do not change over time; ensuring data definitions are stable and uniform; ensuring granularity matches the purposes to which data were being put (e.g., in time series studies, should analyses be precise to the second, to the hour, to the week or to the year); adopting analytical and sampling techniques to manage high volume; and so on.
Data is mapped to the levels of progression using statistical indicators The data elements selected to construct the assessments should be capable of generating stable behavioural indicators for each learner. For instance, when using network analysis, interactions between students might generate statistical indicators of connectedness. Differences in behavior of individuals on each behavioural indicator should plausibly explain different levels of attribute of an individual.
Interpretation of indicators is stable Direct comparability of scores over time is maintained. This is especially important when machine learning is used to construct scores or indictors. If the algorithms change indicators, the construct being measured might change. Any teaching policy changes are identified which may change inferences that can be made from data: for instance, voluntary participation in forums might normally reflect student engagement, but, if it is compulsory to post, it might reflect compliance.
Attribute is fully represented The behavioural indicators comprising the score together provide a balanced and full representation of the attribute along the full range of levels; indicators are not skewed by missing or irrelevant data; there are no important features missing. For instance if off-line work is vital to a process, it is difficult to see how an online automated assessment can be other than skewed.
Scoring and data transformation is transparent There is a transparent audit trail describing each stage of the transformations of data to indicators to scores. Metrics and algorithms at all stages of development are transparent.
Technical quality is adequate The psychometric qualities of accuracy, discrimination and reliability of indicators and scores are demonstrated: the indicators cohere with a simple developmental integrity, generating a scale with even intervals, and the criterion of conjoint measurement applies, there are no biases evident for sub groups. These characteristics can be examined by testing fit to a measurement model.

(continued)

Table 13.1 (continued)

Scores are interpretable There is no other plausible interpretation of the assessments other than that they reflect differences between individuals in the level of competence or capability they possess.
Alternative methods are canvassed There are no simpler, alternative methods for arriving at assessments.
Unintended consequences are identified There are likely to be no unintended negative consequences arising from use of the assessment that arise from shortcomings of the assessment.
Appeals are possible There is a process of appeals for review, important if assessments are the product of complex algorithms that are difficult for stakeholders to understand.

constellation of knowledge**,** understandings, skills, beliefs, attitudes and values that define different levels of mastery, based on understanding of typical trajectories or progressions described by learners as they learn. One difficulty for assessment designers seeking to use process-based data is that there are as yet not many examples of such progressions that describe the development trajectories for general capabilities. The first task of teachers or analysts seeking to design assessments is, therefore, to define a progression that plausibly describes the likely patterns of behaviours of individuals who have more or less of the attribute, so that they be arrayed along an underlying continuous scale from more to less. Defining a progression is not a simple task, and is often neglected by analysts or teachers keen to skip straight to the data. Unfortunately, without a theoretical progression on which to base empirical work, there is no means of determining the validity, utility or interpretability of assessment scores.

There are also difficulties in assessing individual performance when the individual is performing with others in a group. This is especially difficult for assessment of general capabilities like teamwork and collaboration, since such attributes are exhibited only in social environments. Teachers understand this, and work through it every time they assess group work. The relationship between performance of the group and the performance of the group member is complex. This complexity is endemic in data derived from digital forum participation, or collaborative work, or multi-user interactive activities. Current measurement and assessment methodologies struggle to untangle the complexities sufficiently to support individual assessments from confounded data of this sort. It is notable that, after reviewing the last 10 years of years of solid, large-scale psychometrically-based work on the topic of assessment of collaborative problem solving capacity, a recent report for the prestigious National Assessment of Educational Progress (NAEP) in the US acknowledged that the world is not yet in a position to a confidently make claims to be able to measure the collaborative problem solving ability of students (Fiore et al. 2017). Solutions to this problem are still developing (Wilson and Scalise 2016). A non-technical approach is to recognise that when the performance of a group in a domain is at issue, it is appropriate to focus assessment on the performance of the group rather that the performance of the individual. This approach would involve assessors taking seriously a socio-centric view of knowledge generation rather than a psycho-centric view of it. A more technically oriented team (von Davier and Halpin 2013) noted that the confounding effects of group membership

are traditionally treated as ‘random error’ in assessments. They have proposed turning this around, psychometrically, so that indicators of the degree of error in individual measurement might be used to indicate presence of group capacity independent of individuals’ skill.

Further work is also required to determine whether or not an individual’s capacity to perform at particular levels on measures of general capabilities such as problem solving, or communication, or persistence, are generalisable, and thus worth assessing in one context or domain, because they are transferable to another. For example, will a learner who demonstrates good problem solving skills in a chemistry classroom demonstrate good problem solving skills in workplace, or in a physics lab? Will a student who collaborates well in online games be a good collaborator in face-to-face groups? Early research suggests that in general, transfer of these general capabilities, with some notable exceptions, is low (Perkins and Salomon 1992). The interpretations of scores relating to complex general skills in any given context or domain might need to be treated circumspectly, regarded as being context dependent, and used in a way that recognises their limits.

Teachers or assessment designers needing to collect evidence about complex competencies may decide to combine diverse evidence sources (such as peer, self and teacher assessments of portfolios, or combining patterns of participation in forums and in lectures). Unfortunately, when the relationship between diverse indicators is unexamined, it may be that the combined results are of poor quality. Unless each indictor taps into the same underlying attribute, the resultant assessment is likely to be characterized by poor validity, low reliability, inaccuracy and poor utility.

A legitimate response to technical difficulties is to question whether some general capabilities might best be left unmeasured, creating no-go zones for individual assessments. For instance, Masters (2018) gives the example of ‘creativity’, notoriously difficult to assess in any environment. He suggests that it may be appropriate to question whether there exists a generalisable attribute called ‘creativity’ that can be assessed, even in a specific domain. Unless it is possible for the community of stakeholders to reach agreement on a clear definition and progression, it may not be possible to assess.

In addition, the construction of measures that meets metrolytic standards is time consuming, expensive and demands technical skill. It becomes economically feasible only when used at scale. This in itself is likely to limit use of these methodologies in the short term.

This discussion highlights some of the practical difficulties associated with using process data to generate high quality assessments of general capabilities. Methodological challenges include that there is a paucity of clear, empirically derived progressions for these capabilities which need to underpin assessment, the scores generated may lack generalisability, there are questions about feasibility, and technical methodologies are currently limited when it comes to disentangling the effects of group membership on achievement. Further there are dangers in combining scores. Even the robust methodological tools of measurement science may fall short, for a range of technical and transparency reasons. Skepticism is warranted

that any particular assessment tool will have basic qualities such as reliability, validity, accuracy, utility or interpretability.

13.6 Conclusion

The contemporary frontier for assessment focuses on whether and how to use powerful digital technologies to make use of digital data, particularly process data, to better assess and report on the attainment in learning, particularly of general capabilities. The underlying danger is always that 'data' will be mistaken for 'assessments'. This area of assessment is challenging, and it is still uncertain to what degree the new, big data sets, especially process data, will support valid and reliable assessment of learning. It is also uncertain whether current analytical methodologies are up to the task, or whether stakeholders (including learners, teachers and employers) will trust the results. Because of the technical complexity often associated with use of process data, including complex algorithms and data transformation, there are inevitable questions about whether these data have utility for assessment. The impact on learners of inaccurate, unreliable and invalid assessments of general capabilities should not be underestimated in the digital, automated, self-regulated world. Feedback and reporting generate powerful, real-life consequences for learners, which can be positive, but which can also be negative or destructive (Hattie and Timperley 2007). This concern is likely to be acute when assessments are automated to shape the treatment a person receives, and especially if there is no obvious appeal to an independent person who is familiar with the work of the assessed.

Metrolytic standards provide a framework for building confidence in assessments that use process data. It might be that the expected level of rigour inherent in the standards is too great to be feasible. The standards embody quality requirements for reliability or validity, accuracy, or interpretability of the sort normally only applied to high stakes, scaled assessments such as PISA, or SAT or GMAT. However, the application of metrolytic standards is one option available to those working at this assessment frontier, combining the methodological strengths of learning analytics and measurement science, and underpinned by a solid understanding of assessment and its role in learning.

References

Asia-Pacific Education Research Institutes Network. (2015). *Regional study on transversal competencies in education policy and practice*. Bangkok/Paris: UNESCO. Retrieved from: http://www.unesdoc.unesco.org/images/0023/002319/231907E.pdf.

Bergner, Y., Lang, C., & Gray, G. (2017). Measurement and its uses in learning analytics. In C. Lang, G. Siemens, A. Wise & D. Gasevic (Eds.), *Handbook of learning analytics*

(pp. 35–48). New York: SOLAR. https://doi.org/10.18608/hla17. Retrieved from: https://www.solaresearch.org/hla-17/.

Bransford, J.D., Brown, J.D., & Cocking, R.R. (2003). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington DC: National Academy Press.. Retrieved from http://www.nap.edu/read/9853/chapter/1

Buckingham Shum, S., & Deakin Crick, R. (2016). Learning analytics for 21st century competencies. *Journal of Learning Analytics, 3*(2), 6–21.

Carmean, C., & Mizzi, P. (2010). The case for nudge analytics. *Educause Quarterly, 33*(4), 4.

Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing self-regulation as a cyclical, context-specific phenomenon: Overview and analysis of SLR microanalytic protocols. *Educational Research International., 2012*, 1–19. https://doi.org/10.1155/2012/428639.

Corrin, L., & de Barba, P. (2014). Exploring students' interpretation of feedback delivered through learning analytics dashboards. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and reality: Critical perspectives on educational technology. Proceedings ASCILITE Dunedin 2014* (pp. 629–633). Dunedin: ASCILITE.

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Retrieved from http://www.dtic.mil/get-tr-doc/pdf?AD=ADA084551

Dringus, L. P. (2012). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks, 16*(3), 87–100.

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Proceedings of the 7th International learning analytics & knowledge conference*. Vancouver**,** Canada.

Fiore, S. M., Graesser, A. G., Greiff, S., Griffin, P. G., Gong, B., Kyllonen, P., et al. (2017). *Collaborative problem solving: Considerations for the national assessment of educational progress*. Alexandria: National Center for Education Statistics. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/researchcenter/collaborative_problem_solving.pdf.

Gasevic, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71.

Glaser, R. (1994a). Criterion-referenced tests: Part I: Origins. *Educational Measurement: Issues and Practice, 13*(4), 9–11.

Glaser, R. (1994b). Criterion-referenced tests: Part II: Unfinished business. *Educational Measurement: Issues and Practice, 13*(4), 27–30.

Greller, W., & Draschler, H. (2012). Translating learning into numbers: A framework for learning analytics. *Educational Technology and Society, 15*(3), 42–47.

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation, 33*(1), 87–99.

Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach* (Vol. 2). Dordrecht: Springer.

Gunnarsson, B. L., & Alterman, R. (2013). *Understanding promotions in a case study of student blogging.* Paper presented at the 3rd international conference on learning analytics and knowledge, Leuven, Belgium.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487.

He, J., Rubinstein, B. I. P., Bailey, J., Zhang, R., Milligan, S., & Chan, J. (2016). *MOOCs meet measurement theory: A topic-modelling approach*. Paper presented at the 30th AAAI conference on artificial intelligence (AAAI-16), Phoenix, Arizona.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Masters, G. (2018, July 31). But can we measure it. *Teacher Magazine*. Retrieved from https://www.teachermagazine.com.au/columnists/geoff-masters/but-can-we-measure-it

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5–8.

Milligan, S. K., & Griffin, P. (2016). Understanding learning and learning design in MOOCs: A measurement-based interpretation. *Journal of Learning Analytics, 3*(2), 88–115. https://doi.org/10.18608/jla.2016.32.5.

Milligan, S. K., Kennedy, G., & Israel, D. (2018). *Assessment, credentialling and recognition in the digital era: Recent developments in a fertile field* (Seminar series 272). Melbourne: Centre for Strategic Education.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues & Practice, 25*(4), 6–20. https://doi.org/10.1111/j.1745-3992.2006.00075.x.

OECD. (2018). *The future of education and skills: Education 2030*. Retrieved from http://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf

Pellegrino, J.W. (1999). The evolution of educational assessment: Considering the past and imagining the future. William H. Angoff Memorial Lecture. Retrieved from https://www.ets.org/Media/Research/pdf/PICANG6.pdf

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. Husen & T. Neville Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.). Oxford: Pergamon Press.

Polyak, S. T., von Davier, A., & Peterschmidt, K. (2017). Analyzing game-based collaborative problem solving with computational psychometrics. In *Proceedings of 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Halifax, ACTNext.

Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2013). New assessments and environments for knowledge building. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (Vol. 1, pp. 231–300). New York: Springer.

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: MIT Press.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.

Tai, J., & Adachi, C. (this volume). The future of self and peer assessment: Are technology or people the key? In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world (chapter 15)*. Dordrecht: Springer.

Tremblay, K., Lalancette, D., & Roseveare, D. (2012). Assessment of higher education learning outcomes. *Feasibility study report, volume 1 – Design and implementation*. OECD. Retrieved from http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf

von Davier, A. A., & Halpin, P. F. (2013). *Collaborative problem-solving and the assessment of cognitive skills: Psychometric considerations*. ETS: Princeton.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York: Taylor & Francis Group.

Wilson, M., & Scalise, K. (2016). Learning analytics: Negotiating the intersection of measurement technology and information technology. In J. M. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, design, and technology*. New York: Springer.

Wilson, M., Scalise, K., & Gochyyev, P. (2016). Assessment of learning in digital interactive social networks: A learning analytics approach. *Online Learning, 20*(2), 97–119.

Wolfe, E. W., & Smith, E. V., Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part 1 & part 2. In E. V. Smith Jr. & R. M. Smith (Eds.), *Rasch measurement: Advanced and specialized applications* (pp. 202–290). Minnesota: JAM Press.

World Economic Forum. (2015). *New vision for education: Unlocking the potential of technology*. Geneva: WEF & The Boston Consulting Group.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Sandra Milligan is Director of the Assessment Research Centre at the University of Melbourne. Sandra's current research focuses on assessment, micro-credentialing and warranting of hard-to-assess curriculum or competency areas, including use of 'big data', learning analytics and developmental assessment. Sandra has an unusually wide engagement with the education industry, and in educational research. Originally a teacher of science and mathematics, she is also a former Director of Curriculum in an Australian state education department, and has held senior research, management and governance positions in a range of educational organisations, including government agencies, not-for-profits, start-up businesses and large, listed, international corporations.

Part IV
Practical Exemplars

Joanna Tai

In this section of the book, we turn now towards the practical. While the previous sections dealt with the theoretical, conceptual, and imagined future world in which assessment might take place, this section contains a selection of chapters which detail ways of doing assessment differently. Significantly, the chapters contained in this section discuss tools, designs and implementations which have *already occurred*: the future is now, the future is here. While the adage 'those who do not study the past are doomed to repeat it' applied in this context might remind us of assessment dilemmas which occurred prior to the advent of the digital, we suggest readers consider how they might indeed repeat, replicate, or indeed adapt these already existing future-oriented ideas into their contexts.

To satisfy the potential diversity of readers, there are multiple perspectives on and positioning of assessment within this section. As editors, we did not feel it was our place to constrain thinking about assessment to a particular focus. Diversity is a strength that should be capitalised on where possible. The resultant chapters in this section, we hope, can inform a broad range of assessment-related practice and satisfy the multiple purposes of assessment: of learning, for learning, and as learning. Furthermore, the chapters deal with assessment from the microscopic task level (e.g., Apps et al. (Chap. 17), and Cain et al. (Chap. 16)), to the macroscopic policy level (Tai and Adachi (Chap. 15) and Jorre de St Jorre (Chap. 19)), while also considering the meso-level curriculum design (Apps et al. (Chap. 17), Cain et al. (Chap. 16), Kim and Rosenheck (Chap. 18), Corrin and Bakharia (Chap. 14), and Tai and Adachi (Chap. 15)). Themes of student agency, the role of peers in assessment, and moving beyond traditional assessment configurations and structures are also conceptual threads which link the chapters in this section, and merit further discussion here.

Student Agency

The concept of reimagining the role of the student in assessment is a key part of reworking what assessment looks like within a digital world, whether explicitly articulated, as in Kim and Rosenheck's chapter on a playful way to involve learners in the development of rubrics (Chap. 18), or as an integral part of the assessment design, as in Cain et al.'s chapter on a standards-based assessment framework and accompanying digital curriculum and feedback platform (Chap. 16), or Jorre de St Jorre's digital credentialing (Chap. 19), where the graduate-institution-employer relationship is altered. The digital affords different learner interactions, with peers (e.g., Corrin and Bakharia, Chap. 14)) and educators (e.g., Cain et al., Chap. 16), which can drastically change how assessment is done. The recognition of learners as active constructors of knowledge and experience rather than passive vessels may be one of the underlying requirements to fully exploit the affordances of a digital world, and indeed these chapters demonstrate the types of reimaginations necessary to do so.

The Role of Peers in Assessments

A parallel concept to student agency is the increasing role of peers in assessment. Corrin and Bakharia (Chap. 14) share two very different examples of online peer assessment, while Tai and Adachi (Chap. 15) provide a broad overview of technological options and also address the question of necessary knowledge, skills and attitudes for those implementing and undertaking peer assessment. In these examples, peers form an important source of assessment information on a range of tasks. However, Apps et al. (Chap. 17) also identified that peers might contribute to additional unintentional confusion and misunderstanding within online assessment tasks, especially where the task design is novel or unfamiliar to students. Overall, the potential for peers to contribute either formally or informally to assessments being undertaken in digital settings should be considered: it may provide new opportunities for learning, and new opportunities for information gathering, however unintentional effects may also transpire and need to be monitored – a necessary consideration when educators relinquish some control over assessment.

Moving Beyond Traditional Assessment Configurations and Structures

The final key concept encountered in this section, which might also be one of the key tenets of the book overall, is the need to go well beyond traditional assessment configurations and structures. The innovative work here does not just rehash what

has previously been done in face-to-face or even distance settings. It drastically reimagines the way in which educators, learners and others are configured with regards to time, space, relationships and materials to achieve new forms of assessment. This is evident in Cain et al.'s chapter (Chap. 16), where formative interactions occur in relation to learning objectives through a digital platform, and Jorre de St Jorre's chapter on using digital microcredentials to illustrate learners' certifications in a way that is not possible via a traditional academic transcript (Chap. 19). Kim and Rosenheck's chapter (Chap. 18) asks students to participate in the design of rubrics, an activity usually accomplished by educators, and also asks students to move beyond considering assessment as a place for knowledge reproduction.

While one cannot push the boundaries in all directions at once, we hope this section provides a smorgasbord of ways to move forward when reimagining assessment within a digital world, and that readers can select a way forward that meets the requirements and appetites of their local context.

Chapter 14
Re-imagining Peer Assessment in Self-Paced Online Learning Environments

Linda Corrin and Aneesha Bakharia

Abstract In higher education the ways courses are delivered online continues to change and evolve, prompting a need to rethink how assessment can be designed to suit these learning environments. This chapter explores how technology can be used to help deliver assessment designs in self-paced online learning environments that provide flexible, customisable, and scalable peer assessment and feedback. Two innovative approaches are profiled and used to illustrate the affordances of such digital environments in providing ways to support peer assessment activities and the provision of feedback for student learning.

14.1 Introduction

In higher education the ways courses are delivered online continues to change and evolve, prompting a need to rethink how assessment can be designed to suit these learning environments. As universities seek to attract a broader range of students from a diverse range of backgrounds, increasingly flexible online learning formats are emerging which allow students to move through a course at their own pace, without a specific start/end date. This format of delivery enables students to better align their learning with life and career needs and commitments. They can choose to work through the whole course content in a short period of time, or alternatively spread the load over many months. The ability to set the pace of learning is particularly important where students may be trying to fit their study around existing work commitments.

L. Corrin (✉)
Learning Transformation Unit, Swinburne University of Technology, Hawthorn, VIC, Australia
e-mail: lcorrin@swin.edu.au

A. Bakharia
UQx, Institute for Teaching and Learning Innovation, University of Queensland, Lucia, QLD, Australia
e-mail: a.bakharia1@uq.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_14

While self-paced online learning environments can offer many benefits to students, they also present challenges from a learning design perspective, especially in relation to the design of assessment. In an online course involving a single cohort who work through the course at the same pace it is possible to have shared assessment deadlines and communication between students as part of learning activities. Peer assessment is regularly used in these courses to enable students to receive personalised feedback on their work without the need for a teacher to mark each individuals' work. A student's work may be reviewed by multiple peers, which potentially gives them a wider range of comments than had their assessment been marked by a single teacher. The use of peer assessment also allows for more open-response forms of assessment to be delivered at scale (Kulkarni et al. 2013).

In self-paced courses the flow of students through the learning environment varies, making collaborative activities and the use of peer assessment more difficult to implement. The ability to get students to review the work of their peers is complicated by the fact that they are not all submitting their work at the same time. To address these challenges, new and innovative approaches to assessment are needed. This is especially important for large-scale courses where the number of students makes teacher-marked assessment impractical. In order to respond to this need, educators have to reimagine how assessments can be designed and facilitated in self-paced online learning environments.

In this chapter we will profile two innovative approaches, 360-degree feedback and PerspectivesX, and explore the potential benefits and challenges they provide to the delivery of peer assessment and feedback in self-paced online environments. The 360-degree feedback activity enables students to gain the benefit of rich feedback from workplace peers on their assessment tasks without the need for common deadlines or heavy marking loads for teachers. PerspectivesX is a collaborative multi-perspective elaboration tool designed to both encourage student idea submission and the development of curation skills, a twenty-first century digital literacy that fosters critical inquiry and exploration (O'Connell 2012). Both options represent different ways of incorporating people and technology into assessment approaches, while still taking advantage of the pedagogical benefits of peer learning.

14.2 Peer Assessment in Large-Scale Online Learning Environments

There is a long history of the use of peer-assisted learning in higher education. Peer-assisted learning involves "people from similar social groupings, who are not professional teachers, helping each other to learn and by so doing, learning themselves" (Topping and Ehly 1998, p.1). A common form of peer-assisted learning is peer assessment where students are responsible for evaluating the work of their peers and providing feedback (Liu and Carless 2006). Peer assessment offers a number of benefits to student learning including cognitive gains (e.g., error and misconception

identification, identification of knowledge gaps, etc.), development of students' evaluative judgment skills, improvement in student writing, greater provision of feedback, and improvements in preparation for future assessments (See also Tai and Adachi, Chap. 15 this volume; Cho and Cho 2011; Russell et al. 2017; Sadler and Good 2006; Tai et al. 2018; Topping 2009). It is used widely in higher education in both face-to-face and online learning environments.

As the use of large-scale online learning environments has increased in higher education, interest in how peer assessment can help facilitate assessment for large numbers of students has grown. Delivering effective assessment tasks to students in such environments presents many challenges to teachers and learning designers. Initially such courses tended to rely on multiple choice tests that could be automatically marked to cater for the large number of students enrolled (del Mar Sánchez-Vera and Prendes-Espinosa 2015). However, as the topics and outcomes of these courses expanded there was a need for the incorporation of a wider variety of assessments to evaluate student learning, including more written work. The large enrolments in these courses made the marking of these types of assessment difficult, even with a team of teachers and teaching assistants. It was largely in response to this issue that peer assessment began to be incorporated into these large-scale online learning environments.

While peer assessment offers many benefits for supporting student learning, there are also concerns about the ability of students to effectively evaluate their peers' work and to provide helpful feedback (Piech et al. 2013). With large numbers of students these concerns can be greater due to teachers having less opportunity to mediate and supervise the process (Suen 2014; Adachi et al. 2018). Previous studies and meta-analyses of peer assessment research have reported mixed findings as to how well the assessment and feedback of peers correlates with what a teacher would have provided (Russell et al. 2017). Recommendations that have been made on the basis of this research for improving the reliability and validity of peer assessment include: making sure that the criteria are clear and well understood (Falchikov and Goldfinch 2000); ensuring that multiple peers assess the work (at least four was suggested by Cho et al. 2006); and consider giving a mark for the peer assessment process to encourage engagement and commitment to the task (Ballantyne et al. 2002). In online learning environments a variety of technical approaches have been employed to improve the accuracy of peer assessment outcomes. For example, the use of an automated guidance tool that applies a scaffold to the written file to be assessed drawing the markers' attention to important elements that relate to predefined rubrics for the assessment (Vista et al. 2015).

The potential benefits of peer assessment for student learning and the ongoing efforts to improve its reliability and validity make it an appealing option for assessment design in large-scale online learning environments. However, most current models of peer assessment are founded on the requirement that students all complete the assessment and marking at the same time. In self-paced learning environments, this core requirement is not possible. This challenges educators to design ways that harness the pedagogical benefits of peer-assisted learning and assessment whilst still allowing flexibility for students to complete the learning and assessment

activities at their own pace. This chapter profiles two innovative approaches to solving this problem while still allowing students to learn from their peers.

14.3 360-Degree Feedback

360-degree feedback models have traditionally been used in business environments for employee performance evaluation. The model involves the gathering of information from self, peers, supervised staff, and managers about the employee in order to make an overall judgement of performance (London and Beatty 1993). In an educational context the model involves the completion of a work-related project. Feedback on this project is sourced from the student's colleagues within the workplace, rather than (or in addition to) other students within the same course. A major advantage of this approach is that the student's work colleagues are in a good position to be able to provide feedback that draws on knowledge of the organisation, domain and context of the industry to which the project relates. Additionally, the 360-degree feedback model can provide important links between the learning outcomes of the course and the work-related goals of students (Egloffstein and Ifenthaler 2017).

The application of this method is not new in the higher education environment (Tee and Ahmed 2014), but is now commonly seen as part of the assessment design in large-scale online learning environments. The 360-degree feedback approach is useful in online courses where students can complete authentic workplace-related assessment tasks such as courses that focus on change management, project management or leadership. In the self-paced online learning environment this form of assessment can still make use of feedback provided by peers, however because the peers are colleagues drawn from the student's workplace, it is not tied to any cohort of students in a course. This allows the students to undertake the assessment at any time that fits with their enrolment and study schedule.

In addition to the timing flexibility for completion of the 360-degree feedback assessment task, there are other benefits offered by this form of assessment. By receiving feedback from colleagues from within the same industry domain the feedback is more likely to relate to the expectations and culture of the industry. In contrast to course peers who may give simplistic feedback due to a lack of knowledge of the domain, colleagues have a better understanding of the terminology and concepts that are key to the specific domain, enabling them to give more focused feedback (Morris et al. 2017). This can result in more contextualised advice that can be directly integrated with the student's work practices. By involving colleagues, it also has the potential to raise awareness in the workplace of knowledge and skills that the employees are developing through continuing education. For the student, the ability to see a direct link from the assessment task to their professional context can increase their motivation by increasing their value beliefs (Wigfield and Eccles 2000).

14.3.1 Challenges Related to 360-Degree Feedback

There are however challenges in this form of assessment. While the use of colleagues can enable the provision of greater industry-specific feedback, these colleagues may not be as capable of giving feedback with respect to course subject matter. Providing workplace colleagues with adequate support materials to make them aware of the kind of material covered in the course could help to reduce this problem, but this shouldn't be too onerous else it may discourage colleagues from taking part in the process. For this form of assessment to be viable it is necessary that there is support from the employer for the time of colleagues to contribute to the process. Without this support the assessment cannot be completed or if insufficient time is allowed then the efficacy of receiving the feedback could be reduced. There is also a concern about the impact that the receipt of negative feedback from colleagues could have on the student and their motivation. When 360-degree feedback is used as a strategy for performance evaluation in the workplace there are, ideally, consultants and/or counsellors who can help to mediate such negative occurrences. However, the same safeguards may not be available in the workplace where this task takes place and how such situations will be given attention needs to be considered when the negotiation of the task takes place with the workplace employer. There is also the possibility that workplace colleagues are not available (e.g., in the case of a sole trader), and, in these situations, colleagues from a similar organisation or external stakeholders could be approached to take part in the feedback process.

While the core design of a 360-degree feedback assessment task in a large-scale online environment may not be substantially different from that given to a small cohort studying face-to-face, the larger number of students and flexible mode of delivery means that systems need to be built to support this assessment design in this environment. There are some similarities between the 360-degree feedback assessment and other peer assessment activities embedded within a Learning Management System (LMS) platform. The key differences include allowing the learner to specify the contact details of workplace peers and providing workplace peers not enrolled in the course access to the individual assessment item submitted by the learner. Rather than building a new tool, we envision the tool being an extension to existing peer assessment tools provided in popular LMS platforms. The following sections detail the design specifications of a tool built to support a 360-degree feedback assessment design (Morris et al. 2017).

14.3.2 360-Degree Feedback Tool Design

The first design specification relates to the ability of the teacher to create a 360-degree feedback task that is suited to the course learning outcomes. Flexibility is required in order that the task can not only be tailored to the course content, the discipline context and the learning outcomes, but also allow it to accommodate different work

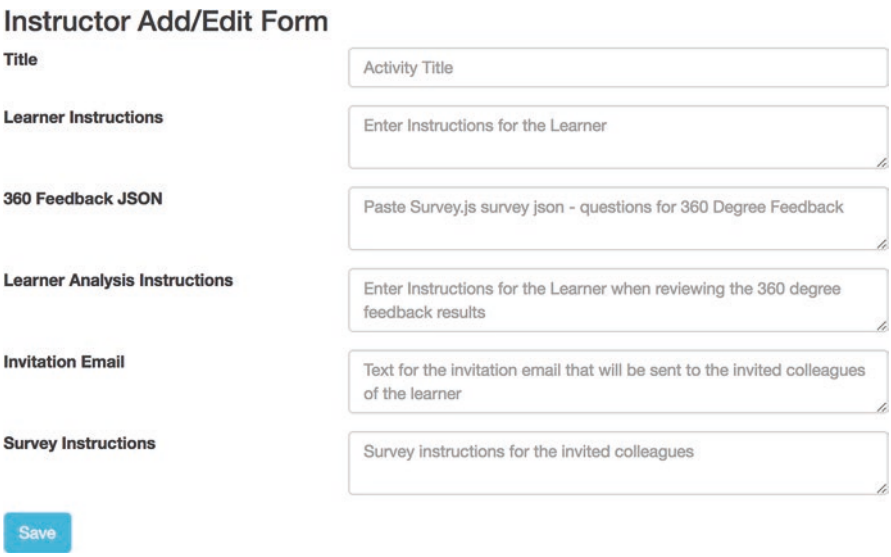


Fig. 14.1 Design and communication elements of the 360-feedback assessment activity

conditions and local adjustments. The form below (Fig. 14.1) prompts the teacher to outline the important design and communication elements of the activity. The activity is given a title and instructions can be specified for the learner about the design and assessment criteria for the activity.

The system includes the use of an online rubric to enable a student's work colleagues to evaluate the outcomes of a workplace assessment item and submit feedback for the student. The rubric is built using the Survey.js library (2018) and includes a full features survey builder user interface. Instructors can use a visual tool to create the rubric and then paste the resulting JSON (JavaScript Object Notation) format code within the form shown in Fig. 14.1. The rubric can include different items (tailored to the activity requirements) and users can give a response in free text, Likert scale, multiple choice, multiple selection, and slider formats.

The next three fields on the form (see Fig. 14.1), allow the teacher to enter text which forms part of email communications with the student about their feedback and the student's colleagues, both to invite them to participate and to provide them with instructions about the rubric if they agree to take part. It should also be noted that the system can also allow for the students' peers within the course to provide feedback if appropriate and if their study schedules allow this to take place in a timely manner.

Students participating in the activity can nominate work colleagues to provide feedback information on their assessment task (see Fig. 14.2). The student is required to enter the name and email address of a colleague and can be given the option to specify the relationship they have to that colleague (e.g., manager, external stakeholder, etc.). Students are also able to preview the rubric. Emails are automatically generated using the text specified by the teacher in the invitation email field

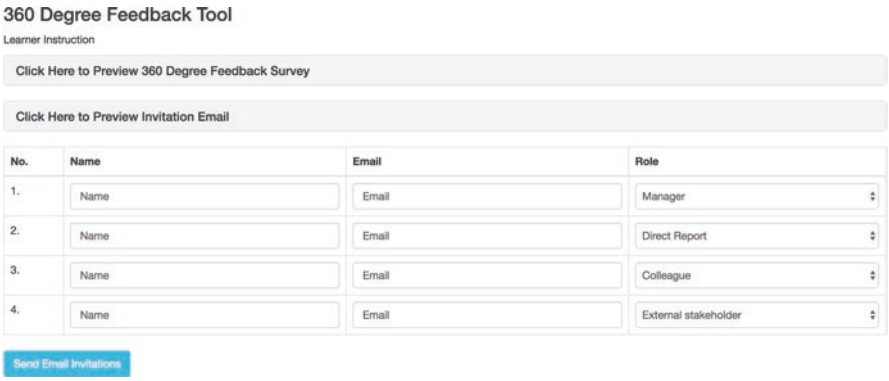


Fig. 14.2 Form for nominating colleagues to take part in the peer review

(see Fig. 14.1), including a description of the assessment task, the assessment criteria, and the date on which their review is due. The colleague can accept the invitation by clicking on a link in the email they receive. To avoid reviewers needing to establish accounts in the system, they receive a unique, encrypted link which provides them with direct access only to the student's work and the assessment rubric for completion. However, they are required to accept a terms of service agreement before they can access the student's work for review. These terms of service relate to issues such as appropriate use of the system, confidentiality, and fairness in giving feedback. Teachers may also consider providing guidance materials for colleagues on the assessment criteria and how they can give constructive feedback for the student.

14.3.3 Learning Analytics

Another important design feature of the 360-degree feedback assessment system is the analytics it can provide to teachers and students. Some of these analytics relate to the process, allowing teachers and students to see, for example, whether reviews have been completed by workplace reviewers. The system can also send out automated reminders to reviewers to remind them of approaching deadlines for reviews. Other analytics on the performance of the students are also available. Summary statistics on each rubric question is provided and can be analysed at the individual level by students and teachers or at a cohort level by teachers and course coordinators (for example, see Fig. 14.3). The collection of these analytics assists in evaluating whether this model of assessment is working effectively within the context of the course.

The 360-degree feedback assessment design offers the flexibility required in self-paced learning environments while also allowing students to benefit from peer-assisted learning. As students in a self-paced course could be completing this

Rubric Criterion	Your Self Evaluation	Average Cohort Peer Score	Average 360 Degree Peer Score
Key Components of Leadership Development	80%	85%	75%
Personal Goals	70%	60%	85%
Organisation and Structure	100%	70%	85%
Grammar, Punctuation and Spelling	100%	95%	65%

Fig. 14.3 Analytics on a student's 360-degree assessment activity performance

assessment at different times, a system such as the one outlined above can be useful to facilitate the different stages of this activity and ensure that the review process runs smoothly. Analytics collected by the system can be used to monitor the quality of the reviews, and to this end useful types of analysis include:

- Comparing rubric scores obtained from enrolled cohort peers with those received from workplace colleagues;
- Comparing written free text feedback information provided by enrolled cohort peers with those received by workplace peers;
- Analysing differences in feedback provided by the types of peers that are nominated (e.g., direct supervisor, indirect supervisor, direct report, other internal stakeholder, other external stakeholder); and
- Analysing whether workplace peers provide more contextualised and detailed information.

These analytics can help to evaluate the effectiveness of this form of assessment for students studying in self-paced learning environments. Analytics on students' access to this information also helps give some insight into whether they are engaging with the feedback provided by their workplace peers. It is recommended that teachers also build in guidance for students in the interpretation of this feedback information to support their learning.

14.4 Collaborative Multi-perspective Knowledge Construction and Curation

The second peer-based assessment approach profiled here is a knowledge construction and curation assessment design, facilitated through a tool called PerspectivesX (Bakharia 2017). PerspectivesX is designed to support the scaffolding of learning activities that require students to elaborate on a problem, scenario or topic from different assigned perspectives. It can support a range of custom and predefined idea

generation and multi-perspective activities such Strengths, Weakness, Opportunities and Threats (SWOT) analysis, SCAMPER (Ederle 1996), Fishbowl (Miller and Benz 2008), and Six Thinking Hats (de Bono and de Bono 1999). PerspectivesX can be used to encourage active participation in online discussions by allowing students to explore, review and curate information from their peers' submissions. It provides a scaffold to improve student engagement in an online collaborative learning environment, even when students are not all present in the learning environment at the same time.

The need for a tool such as PerspectivesX came about due to a lack of scaffolded collaborative tools, particularly tools that could integrate with online learning platforms. Most online learning environments have rich support for content authoring, but the discussion forum is usually one of the few tools available for collaboration. Strategies to encourage student engagement with online discussions have been the focus of much research over time (e.g., Hewitt 2005; Salter and Conneely 2015). Research examining Massive Open Online Courses (MOOCs) found that it is typical for only 5% to 10% of learners to participate within a discussion forum (Manning 2013). It has also been shown that learners who actively participate in discussion forums achieve better grades (Corrin et al. 2017). Better discussion activity design and support tools that encourage learners to share their own ideas and review the ideas of their peers are therefore required. PerspectivesX has been designed to provide a scaffolded collaborative experience for students.

14.4.1 Learning and Assessment Design of the PerspectivesX

Central to this tool is a searchable knowledge base that contains a record of all students' contributions. The knowledge base can persist over re-runs of self-paced courses allowing students access to this rich resource collection regardless of when they are completing a course. This approach is inspired by the Knowledge Community and Inquiry (KCI) model (Slotta and Najafi 2013) which is based on three main principles: (1) a knowledge base is created as a result of the collaborative work of students; (2) the knowledge base is used as a resource base for collaborative inquiry activities; and (3) the learning goals of the course are evidenced through the outcomes of the inquiry activities (Slotta and Najafi 2013).

In an assessment activity facilitated through PerspectivesX, students are required to access the contributions of other students, to critically review these, and to curate relevant material to prepare their own contribution. Curation is an important digital literacy that involves the development of student search and evaluation strategies as well as promoting critical thinking, problem solving, and participation in networked conversations (O'Connell 2012). The learning design objectives for a PerspectivesX activity are to:

1. Encourage students to submit and curate ideas across all perspectives;

2. Encourage students to share ideas (even if they initially submit ideas as anonymous or not shared);
3. Encourage students to curate a list of diverse ideas within and across perspectives; and
4. Trigger discussion among students in a post activity forum.

The design guidelines that underpinned the development of the PerspectivesX tool aim to support:

- the design of structured knowledge construction, critical thinking and multi-perspective elaboration activities;
- the ability for students to receive a participation score;
- opt-in and anonymous learner knowledge sharing;
- teacher moderation and intervention;
- student curation;
- temporal independence (i.e., asynchronous student activity);
- knowledge base growth and collective knowledge sharing across course re-runs and multiple cohorts; and
- the facilitation of the delivery of scalable, customised and personalised feedback.

In Fig. 14.4, the form used by instructors to add a PerspectivesX activity to a LMS is shown. The teacher selects an existing grid template or creates a new custom grid. PerspectivesX is configurable, allowing instructors to define whether students choose their own perspective or are assigned a random perspective on which to contribute. A perspective refers to the point of view from which the student is asked to address a question or problem as part of the learning activity. Configurable options are available for curation and for enabling the knowledge base settings.

The student contribution user interface is displayed in Fig. 14.5. Each of the perspectives are displayed in a grid format. Each grid has its own submission input element and the student can specify whether the item is shared, anonymous or not shared by selecting one of these options from a drop-down box. The default option is set to be shared, but students either not comfortable or confident with sharing are easily able to opt out whilst still receiving a contribution grade. A visually clear distinction is also made between the items a student has submitted versus the items they have curated. At any point in time the student can export the grid they have assembled to a Microsoft Word document to be saved offline and/or printed.

14.4.2 Learning Analytics

The provision of analytics is important for teachers and moderators to facilitate the activity and provide feedback information as the activity is progressing. A prototype of the teacher dashboard is provided in Fig. 14.6. The metrics and visualisations were chosen to provide evidence to support and evaluate the learning design objectives in real time. Table 14.1, below, maps the tool learning design objectives against

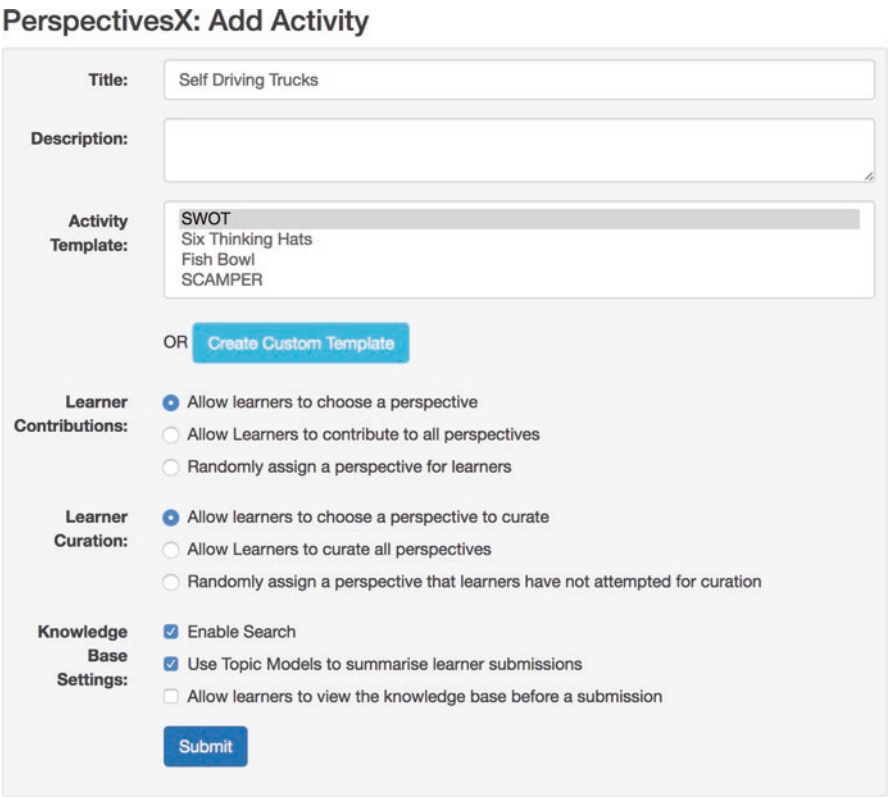


Fig. 14.4 The PerspectivesX user interface for instructors

the selected analytics. Currently these analytics are generic to all activity types, but there is potential for customised reports to be incorporated later.

The instructor dashboard currently includes a timeline chart with a series for each sharing option (i.e., shared with other students, shared anonymously or not shared). The timeline gives the instructor an indication of student sharing behaviour over time. It is important for the teachers and moderators to have an indication of whether student sharing behaviour is changing over the duration of the activity. Submission count distributions are included for each perspective with the aim of providing a summary of the number of student submissions.

Natural Language Processing techniques can be used to provide analytics to address the learning design objective "Encourage students to curate a list of diverse ideas within and across perspectives". To evaluate this objective, teachers need a high-level overview of student submissions to see how students, as a group, are contributing and the topics being covered. Topic modelling algorithms (Lee and Seung 1999) will be incorporated within the instructor dashboard to provide a breakdown of the submissions for each perspective by topic. The teacher will also be able to analyse the student submissions that map to a topic for misconceptions to provide students with appropriate feedback.

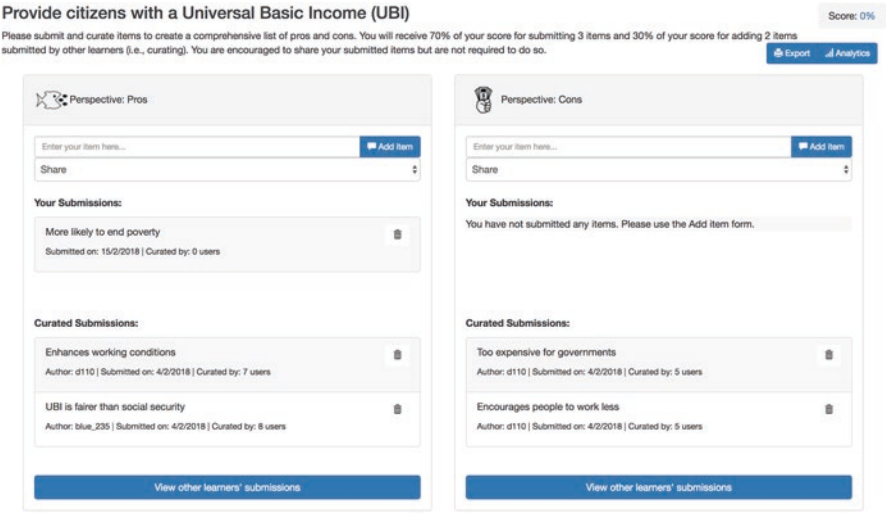


Fig. 14.5 The PerspectivesX learner contribution and curation user interface using a 'pros and cons' custom design

Peer assessment activities facilitated through PerspectivesX have the potential to help students learn from and contribute to discussions with peers about course content while also engaging in content curation, an important digital literacy skill. The tool provides flexibility for the teacher to implement a range of activity designs and customise these to the course learning outcomes and content. Most importantly, it provides an opportunity for students to participate in collaborative assessment asynchronously, which makes it a suitable option for self-paced online learning environments.

14.5 Continuing to Re-imagine Peer Assessment for Self-Paced Online Learning Environments

In this chapter we profiled two options for designing and facilitating peer assessment in self-paced online learning environments. Both these designs involve key elements that ensure students gain the benefits of peer-based learning and assessment. These include connecting students with the work and/or feedback of their peers, who could either be fellow students or workplace colleagues. The designs allow students to complete assessment activities at a time that fits with their own schedule, so they are not constrained by shared assessment deadlines with a course cohort. Both designs enable teachers to facilitate the process of the assessment activities. The learning analytics available through the tools are closely related to the learning design of these activities so that they can provide meaningful information to both teachers and students (Bakharia et al. 2016) and be used to evaluate the effectiveness of the assessment activities.

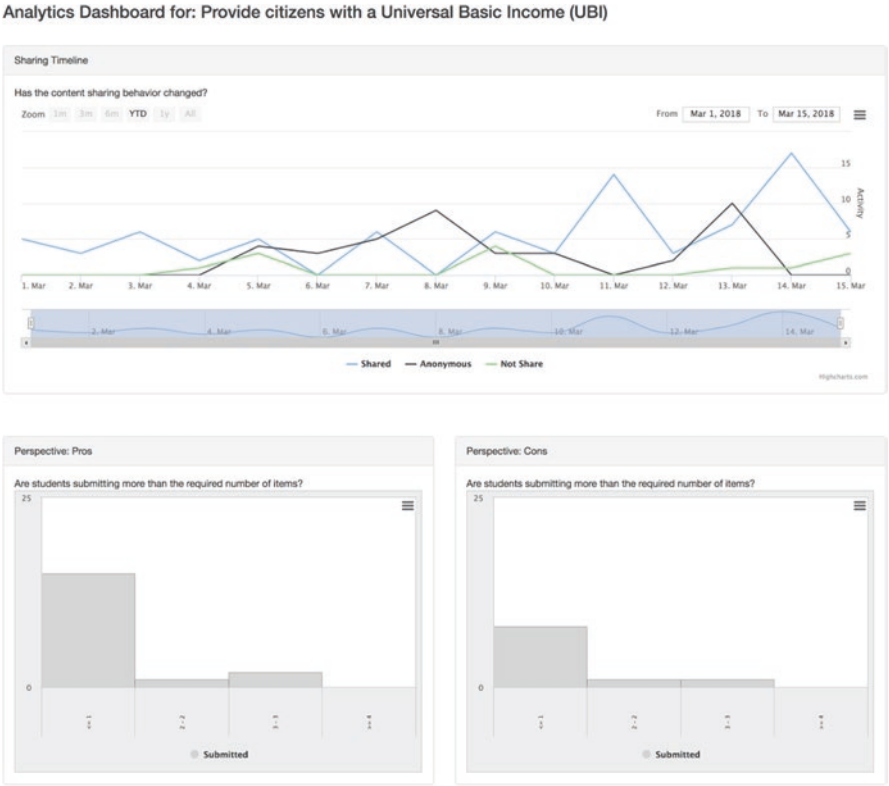


Fig. 14.6 A prototype instructor dashboard for PerspectivesX

Table 14.1 Mapping learning design objectives to analytics

Learning design objectives	Metrics and visualisations to support objectives
Encourage students to submit and curate ideas across all perspectives	Histogram of number of perspectives submitted to by students
Encourage student to share ideas (even if they initially submit ideas as anonymous or not shared)	Timeline of submitted ideas by sharing options (i.e., see how sharing patterns change during tool use)
Encourage students to curate a list of diverse ideas within and across perspectives	Show topics covered by items submitted to a perspective Show list of top curated items
Trigger discussion among learners in post activity forum	Sharing and curation networks that lead to discussions

With the increasing use of self-paced online learning environments in higher education it is important that educators consider new and innovative ways of designing assessments that are relevant, pedagogically-sound and flexible whilst providing students with useful feedback on their learning. Conversations about the future directions of peer assessment in online environments often focus on harnessing

technology in conjunction with the information we know about students to improve the reliability of marking and feedback. These approaches can include the dynamic assignment of peers to give feedback on work based on their previous peer assessment feedback experience and ratings (del Mar Sánchez-Vera and Prendes-Espinosa 2015). There have also been advances in the automated marking of assessments using technology to provide feedback. However, while this provides flexibility and feedback, it doesn't enable peers to learn from each other.

A great opportunity for the higher education community is to think of new ways that the power of students learning from each other can be built into assessment designs and the tools used in self-paced online learning environments. The two suggested approaches in this chapter have focused on providing flexible, customisable, and scalable peer assessment activities that enable the provision of feedback, the moderation of the assessment process, and the evaluation of student engagement and outcomes using learning analytics. These approaches utilise technology to enable the delivery of peer assessment tasks to students regardless of when they start their studies and the pace at which they proceed though their course. Other designs that build on these important elements can add to the possibilities in this area. Already suggestions have been made in the literature to expand the involvement of peers through ideas such as the crowdsourcing of expertise and the alignment of student work to industry relevant projects (Kulkarni et al. 2013).

The digital environment affords many opportunities to support more sophisticated assessment designs for students that respond to changing models of course delivery in higher education. The two approaches profiled here, 360-degree feedback and PerspectivesX, demonstrate the importance of aligning assessment and tool design, while still providing flexibility to customise elements of the assessment to the content and cohort of the course. New technologies and techniques such as machine learning and learning analytics offer even greater possibilities to improve, personalise and support such assessment designs in online learning environments. Incorporating artificial intelligence to not only monitor the process of these types of assessment, but also the quality of work and feedback received could help to enable the scalability of such assessment tasks. The challenge now is to develop the designs and tools to support the facilitation of such activities and to ensure that they are appropriately evaluated to verify that there are benefits to student learning.

References

Adachi, C., Tai, J. H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*(2), 294–306. https://doi.org/10.1080/02602938.2017.1339775.

Bakharia, A. (2017). PerspectivesX: A proposed tool to scaffold collaborative learning activities within MOOCs. In C. D. Kloos, M. P. Sanagustin, P. Jermann, D. Seaton, S. Whity, D. Jansen, & M. Calise (Eds.), *Proceedings of work in progress papers of the experience and research tracks and position papers of the policy track at EMOOCs 2017* (pp. 59–64). Aachen: CEUR-WS.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., et al. (2016). A conceptual framework linking learning design with learning analytics. In T. Reiners, B. R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), *Proceedings of the 6th international conference on learning analytics and knowledge* (pp. 409–413). New York: ACM.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education, 27*(5), 427–441.

Cho, Y., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629–643.

Cho, K., Schunn, C., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98*(4), 891–901.

Corrin, L., de Barba, P., & Bakharia, A. (2017). Using learning analytics to explore help-seeking learner profiles in MOOCs. In X. Ochoa, I. Molenaar, & S. Dawson (Eds.), *Proceedings of the 7th international conference on learning analytics and knowledge* (pp. 424–428). New York: ACM.

De Bono, E., & De Bono, E. (1999). *Six thinking hats* (Vol. 192). New York: Back Bay Books.

del Mar Sánchez-Vera, M., & Prendes-Espinosa, M. P. (2015). Beyond objective testing and peer assessment: Alternative ways of assessment in MOOCs. *International Journal of Educational Technology in Higher Education, 12*(1), 119–130.

Ederle, B. (1996). *Scamper: Games for imagination development*. Waco: Prufrock Press.

Egloffstein, M., & Ifenthaler, D. (2017). Employee perspectives on MOOCs for workplace learning. *TechTrends, 61*(1), 65–70.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287–322.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *The Journal of the Learning Sciences, 14*(4), 567–589.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., et al. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI), 20*(6), 1–31.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*(6755), 788.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management, 32*(2–3), 353–372.

Manning, J. (2013, July 18). How widely used are MOOC forums? *A first look*. [Blog post]. Retrieved from https://teachingcommons.stanford.edu/teaching-talk/how-widely-used-are-mooc-forums-first-look

Miller, R. L., & Benz, J. J. (2008). Techniques for encouraging peer collaboration: Online threaded discussion or fishbowl interaction. *Journal of Instructional Psychology, 35*(1), 87–93.

Morris, A., Bakharia, A., Corrin, L., & de Barba, P. (2017). 360-degree feedback in professional development MOOCs. In L. Corrin, S. Milligan, N. Law & U. Ringtvad (Eds.), *Proceedings of the 2nd design LAK workshop*, Retrieved from https://sites.google.com/site/designlak17/workshop-schedule.

O'Connell, J. (2012). Content curation in libraries: Is it the new black? *Collected Magazine, 6*, 4–6.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. In *proceedings of the 6th international conference on Educational Data Mining (EDM 2013)*. Retrieved from https://arxiv.org/pdf/1307.2579.pdf

Russell, J., Van Horne, S., Ward, A. S., Bettis, E. A., & Gikonyo, J. (2017). Variability in students' evaluating processes in peer assessment with calibrated peer review. *Journal of Computer Assisted Learning, 33*(2), 178–190.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.
Salter, N. P., & Conneely, M. R. (2015). Structured and unstructured discussion forums as tools for student engagement. *Computers in Human Behavior, 46*, 18–25.
Slotta, J. D., & Najafi, H. (2013). Supporting collaborative knowledge construction with Web 2.0 technologies. In C. Mouza & N. Lavigne (Eds.), *Emerging technologies for the classroom. Explorations in the learning sciences, instructional systems and performance technologies* (pp. 93–112). New York: Springer.
Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distributed Learning, 15*(3), 312–327.
Survey, J. S. (2018). *Tallinn: Devsoft Baltic OÜ* [Computer software]. Retrieved from https://surveyjs.io
Tai, J., & Adachi, C. (This volume). The future of self and peer assessment: Are technology or people the key? In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world (Chapter 15)*. Dordrecht: Springer.
Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*, 467–481.
Tee, D. D., & Ahmed, P. K. (2014). 360 degree feedback: An integrative framework for learning and assessment. *Teaching in Higher Education, 19*(6), 579–591.
Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20–27.
Topping, K., & Ehly, S. (1998). Introduction to peer-assisted learning. In K. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 1–24). London: Routledge.
Vista, A., Care, E., & Griffin, P. (2015). A new approach towards marking large-scale complex assessments: Developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. *Assessing Writing, 24*, 1–15.
Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.

Linda Corrin is Associate Professor, Learning Analytics at Swinburne University of Technology in Melbourne. Linda's research interests include students' engagement with technology, learning analytics, feedback, and learning design. Currently, she is working on several research projects exploring how learning analytics can be used to provide meaningful and timely feedback to academics and students. Linda is co-founder of the Victorian and Tasmanian Learning Analytics Network.

Aneesha Bakharia is currently the development team lead at UQx, a department at The University of Queensland focused on developing MOOCs for edX platform and blended on-campus courses. Dr Bakharia's PhD research focused on text analytics with case studies in collaborative learning environments. Dr Bakharia has also published broadly within the field of Learning Analytics and has written 10 books on programming and web development.

Chapter 15
The Future of Self and Peer Assessment: Are Technology or People the Key?

Joanna Tai and Chie Adachi

Abstract Self and peer assessment are frequently facilitated through educational technologies. In this chapter, we outline theoretical frameworks of self and peer assessment, and design principles, and compare these with the affordances of current technologies. We then suggest how the digital could be harnessed more fruitfully to achieve pedagogical goals and best practice in self and peer assessment. Finally, we argue that non-technological factors (i.e., educators and learners' knowledge, attitudes and skills) are also critical to the success of self and peer assessment.

Self and peer assessment are pedagogical strategies which have been used for many years. While they can be considered separate concepts and practices, with different theoretical bases, they have a natural symbiosis and are frequently used together. Within the context of higher education, a key goal is to prepare learners for future professional and personal life in an increasingly digital and complex world. In their future roles, graduates are unlikely to have expert assessors monitoring all of their outputs. Self and peer assessment therefore can be used as an authentic strategy within higher education to provide students with relevant learning experiences for future work purposes, and to provide additional feedback from different perspectives. It is also an opportunity for students to develop their own self and peer assessment skills.

Increasingly, just as with other learning activities within the higher education sector, self and peer assessment can be facilitated through digital technologies. Software packages can augment the ability to implement self and peer assessment processes within higher education, overcoming distance, scale and a/synchronicity.

J. Tai (✉)
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: joanna.tai@deakin.edu.au

C. Adachi
Deakin Learning Futures, Deakin University, Melbourne, VIC, Australia
e-mail: chie.adachi@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_15

While many technologies exist, their alignment with theories of self-assessment and peer assessment vary. The use of technology may also present challenges in implementation which go beyond those inherent in the pedagogical strategies alone. These are likely to be linked to the digital literacy of both educators and learners. Selwyn (2008) points out that there can be a gap between the 'state of the art' and the 'state of the actual' in terms of what it is envisaged might occur with the use of technology, compared to what *actually* plays out when contextual and practical factors are included. In this chapter, we take this lens to the topic of self and peer assessment and the educational technology used to facilitate it.

In this chapter, we take a broad view of self and peer assessment even though the existing literature outlines various types (Adachi et al. 2018a). Self-assessment involves students judging their own work against appropriate standards, whereas peer assessment involves students evaluating their peers' work similarly. The assessment may take several forms: written or verbal feedback, a decision about performance in relation to a rubric or other form of criteria, or a grade.

First, we outline design principles of self and peer assessment and feedback and compare these with the affordances of currently available technologies, examining technologies' fitness for pedagogical purposes. We then contend that non-technological factors (e.g., educators and learners' knowledge, attitudes and skills) are likely to be critical to the success of self and peer assessment. Finally, we propose elements of digital technologies which may enhance self and peer assessment practices.

15.1 Principles Informing Self and Peer Assessment

Just as there are many types of self and peer assessment, there are a range of reasons why it might be employed within higher education. We take the view that self and peer assessment are learning processes, from which learners might develop their ability to perform independently. This aligns with Boud and Soler's (2016) argument that sustainable assessment must be a key goal of education, in preparing students for a future outside of higher education, where decisions on quality must be made by the individual often in conjunction with their peers. The use of self and peer assessment also ties in with the concept of developing students' evaluative judgement. i.e., understanding notions of quality, and the ability to apply such notions in a judgement about their work (Tai et al. 2018). Students can exercise their evaluative judgement through self and peer assessment activities. With these goals in mind, a formative approach is usually suitable for self and peer assessment, where the focus is on a developmental trajectory. However, in some cases it is also appropriate for peer assessment to form a component of summative assessment, if the aspect of the assessment can be appropriately informed by peer perspectives, such as in the case of teamwork, or professional behaviours (Tai and Adachi 2017), though we need to be cautious about such use for summative ends (Sridharan et al. 2018). Self-assessment for summative purposes is a fraught area, as discussed below.

15.1.1 Principles and Evidence Supporting the Use of Self-Assessment

Self-assessment is crucial for effective feedback dialogue and more broadly in self-regulated learning, where the learner must have some idea or notion of their actual performance in relation to goals or standards, as a starting point for action to close the gap between them (Molloy and Boud 2013; Sadler 2010). Panadero and Alonso-Tapia (2013) consider self-assessment important both in formative stages, and in the process of working on the final product.

Self-assessment is acknowledged to be fallible, where the self-assessor is relatively inexperienced in the domain where they are self-assessing (Eva and Regehr 2008; Falchikov and Boud 1989; Kruger and Dunning 1999). There may be many factors influencing the accuracy of self-assessment, including the criteria used, the comparability of student and educator assessments (i.e., they may not be valuing the same aspects of work, or even using the same scale), and if the scores were to be used for formal assessment purposes (Boud and Falchikov 1989). The need for external evidence such as feedback from peers, tutors, or even automated systems to inform self-assessment is key (Boud and Falchikov 1989; Eva and Regehr 2008). Recently, it has been demonstrated that providing information to students on the gap between their self-assessment and the educator's assessment, has improved students' ability to self-assess on subsequent related tasks (Boud et al. 2015).

Though opportunities to practise the skills associated with self-assessment are important, increasing domain knowledge is also a critical factor in improving the accuracy of self-assessment (Falchikov and Boud 1989; Kruger and Dunning 1999). Several conditions have been proposed as necessary for self-assessment: being aware of the value of self-assessment; having access to relevant criteria; and having a well-defined task or product (Panadero and Alonso-Tapia 2013).

Therefore, several key features or design elements should be considered when implementing self-assessment:

- Self-assessments should pertain to areas/domains where students already have some knowledge/skill
- Self-assessments should be employed primarily for formative purposes; where it is used in conjunction with final work, marks could be awarded for accuracy only
- Students should have multiple opportunities to self-assess and refine this skill over time
- Students should have some external information/comparison to inform their self-assessment (e.g., an exemplar and criteria)
- Students should receive feedback on their self-assessment which highlights discrepancies in judgement.

15.1.2 Principles and Evidence for Peer Assessment

Peer assessment may serve several practical aims, including providing additional feedback on learners' work, developing learners' evaluative judgement capability, and evidencing teamwork skills (Gielen et al. 2011; Tai et al. 2016). Many theories come into play when peer assessment is implemented. Social dynamics are likely to play a role between peers who are assessing each other, and this may positively or negatively motivate students to apply themselves to a task, depending on their orientation to it (Raat et al. 2013). This is likely to feed more broadly into students' self-regulation of learning, contributing to self-judgement, self-reaction, and self-motivation beliefs (Panadero 2016). Depending on how the peer assessment is configured, the activity may also introduce or induct students to communities of practice they are likely to join, where team members may critique each other on ideas or work in a similar fashion (Lave and Wenger 1991).

Generally, peer assessments should adhere to educational principles, such as those regarding feedback (Boud and Molloy 2013), and constructive alignment of the curriculum (Biggs and Tang 2007). For instance, if peer feedback is a component of the peer assessment, then students will need to know how to participate in a dialogic feedback interaction prior to undertaking the task. In terms of alignment, students should be assessing each other on appropriate elements of the work which relate to learning outcomes and activities. Peer assessments that are irrelevant and not explicitly made meaningful for students may be regarded as tokenistic or inauthentic (Boud et al. 1999).

A framework for designing peer assessment, synthesised by Adachi, Tai and Dawson (2018a) describes nineteen key design elements to be considered when developing a peer assessment. They are arranged in six clusters (see Table 15.1).

Table 15.1 Nineteen key design elements to be considered when developing a peer assessment

Cluster	Constituent components
Basic decisions	The subject area; intended learning outcomes for students; objectives & benefits for staff; temporal elements, type of work assessed; formality/weighting.
Relationship with rest of unit/course	Links to other assessments and learning activities; links to self-assessment.
Interaction configuration	Type of feedback provided; anonymity of peers in the feedback process; how do students then utilise the feedback? (i.e., opportunity to revise work).
Composition of peer groups	How many students per group; are students matched for ability (differing or similar).
Management of assessment procedure	Standards of work/performance; opportunity to calibrate judgements (i.e., scaffolding); moderation of feedback.
Contextual elements	Technology used or required; other resources required; policy environment regarding use of peer assessments.

Adapted from Adachi, Tai, Dawson (2018a)

Technology is highlighted as one of the nineteen elements, indicating its importance and centrality to how peer assessments are executed at present. Use of technology can greatly facilitate many of the other elements, such as timing, links to self-assessment, anonymity, peer group configuration (including matching according to previous ability), provision of standards (e.g., access to exemplars or a rubric), and moderation of feedback.

15.2 An Overview of Current Technologies

There are many digital technologies available which facilitate self and peer assessment. Reasons for their development include logistical considerations for efficiency and scalability that digital technologies afford – e.g., facilitating distribution of comments/marks, calculation of grades, pairing of students in digital learning environments, but also because of a desire to foster pedagogically sound models of self and peer assessment. Some of the well-known technologies include software such as SPARK^{PLUS} (sparkplus.com.au), PeerMark™ (turnitin.com), ReView (http://academ.com.au/review/) and CATME (info.catme.org), which all serve different purposes and address particular issues in peer assessment.

In response to a university's needs to address innovative approaches to self and peer assessment, we were involved in evaluating and trialling digital technologies for self and peer assessment that could be adopted across the institution. The project had two aims:

(i) to investigate the current pedagogical approaches and practices around self and peer assessment across the university; and
(ii) to evaluate and make recommendations for technological solutions to cater for and improve our practices.

Led by the second author, more than 30 academics and academic developers were consulted to provide input across all faculties. The reasons for and practices in which self and peer assessment were employed varied. Some educators only focused on self-assessment as a formative opportunity for students to more critically self-evaluate their own work (i.e., drafts) before final submission, while others implemented quite complex, summative use of peer assessment that incorporated the evaluation of teamwork processes and final products among team members.

Table 15.2 outlines the analysis of our subsequent market research conducted in 2016 on self and peer assessment tools. Essential requirements and functions were identified, and 11 tools were evaluated against these criteria.

It quickly became evident that each tool provided specific capabilities that offered solutions to particular problems. As seen in Table 15.2, these tools were created by tertiary institutions (aside from Turnitin), to address specific needs in their self and peer assessment practices. This diversity of tools highlights the varying institutional focus and the immaturity of the market around self and peer assessment as digitally enabled learning opportunities.

Table 15.2 Market research on self and peer assessment tools – basic functionalities and requirements

Digital tool, URL, Developer	Group formation and evaluation (process of teamwork)	One-to-one peer review (product of individual or group work)	Open source	Further development and cost required	Integration with learning management system	Interface
CATME http://info.catme.org/ North Carolina State University, US	Yes	No	No	Yes	No	OK
WebPA http://webpaproject.lboro.ac.uk/ Loughborough University, UK	Yes	No	Yes	Yes	No	OK
TEAMMATES http://teammatesv4.appspot.com/ National University of Singapore	Yes	No	Yes	Yes	No	Basic
QuackBack https://goldapplesoftware.ca/portfolio/quackback-peer-review University of British Columbia Okanagan	Yes	No	Yes	Yes	No	Basic
SPARK^{PLUS} http://sparkplus.com.au/ University of Technology, Sydney	Yes	No	No	Yes	No	Basic
Calibrated Peer Review (CPR) http://cpr.molsci.ucla.edu/Home.aspx University of California, US	No	Yes	No	Yes	No	OK

iPeer https://lthub.ubc.ca/guides/ipeer/ University of British Columbia, Canada	No	Yes	Yes	Yes	No	Basic
Peerceptiv https://www.peerceptiv.com/ University of Pittsburgh, US	No	Yes	No	Yes	Yes	OK
PeerMark™ https://guides.turnitin.com/01_Manuals_and_Guides/Instructor_Guides/Turnitin_Classic_(Deprecated)/23_PeerMark Turnitin	No	Yes	No	No	Yes	OK
PRAZE http://peerreview.cis.unimelb.edu.au/tools/about-praze/ University of Melbourne	Yes	Yes	No	Yes	No	Good

Broadly, we identified two major functions which these technologies serve: Firstly, they provide a means to assess the product/outcome of students' work (e.g., an essay, poster, group report, presentation, performance), and secondly, they offer a means to assess the process of a team activity (i.e., how team members contributed to a group task). These two primary use cases also offer future (both pedagogical and technological) processes that could be negotiated and managed by teaching teams. Within these categories, there are many variations on the peer assessments' design elements (Adachi et al. 2018a).

With regards to technical requirements, a key issue was the integration of the tool with the current online environment, especially the Learning Management System (LMS), and whether further (potentially costly) development was required to ensure seamless learning experiences via 'hook-ups' and to improve the interface and software capabilities. From the user experience perspective, it was considered superior to have these self and peer assessment tools available and integrated with the main LMS as deeply as possible (e.g., without having to create additional user accounts) and allow students and staff to focus on learning opportunities rather than technical issues.

Based on these key requirements, two tools were selected for the pilot implementation: SPARK^{PLUS}, which allows for evaluation of teamwork, i.e., process of peer work, and PeerMark™ (as part of Turnitin) which allowed evaluation on products of peers and group work. Subsequently, a set of resources (see http://teachassist.deakin.edu.au/self-and-peer-assessment/) and professional development workshops were designed and run to roll out these tools into faculties. These workshops were targeted at both teaching academics as well as learning and teaching support team members (e.g., academic developers and learning technologists) across all faculties. They covered firstly the theoretical purposes and learning opportunities around self and peer assessment, then the technological how-to of usage and assessment workflow with these tools in relation to the already available assessment engine within the LMS. During and after the pilot phase was completed faculties took up the tools in different ways. We present two short cases that illustrate how these tools were employed.

15.2.1 Case 1

Despite the less user-friendly interface and complex nature of the tool, SPARK^{PLUS} was largely taken up by one of the faculties in which teamwork for project-based and problem-based curricula was commonly found (both undergraduate and postgraduate). The faculty continued to increase the number of academics and students using this tool since 2016. In 2017, it was recorded that more than 20 units with approximately 3500 students used it. SPARK^{PLUS} provided a digital space within which students could share their evaluations of each other's teamwork skills and contributions. SPARK^{PLUS} contains a sophisticated algorithm which calculates factors incorporating both self and peer evaluations (i.e., the Self and Peer Assessment

factor or SAPA). Students can also clearly see qualitative feedback alongside the numeric scores, and this was deemed a strength of the tool. This adhered to the principle that students should have some familiarity with the area they are assessing, in order to make valid judgements. An additional advantage for this faculty, however, was that there was an academic developer working closely with academics, who could support the redesign of assessments and implementation of the tool.

15.2.2 Case 2

PeerMark™ was mostly introduced as a peer review tool for formative assessment and feedback opportunities. In one case, students were asked to peer-review early drafts of the two summative assessment pieces using the same rubrics subsequently used for marking. This therefore worked as a great opportunity to cultivate evaluative judgement skills applied to the peer assessment. It was an advantage that students were already exposed to using Turnitin so that they could readily use the functionalities similarly found within PeerMark™. PeerMark™ within Turnitin facilitated rich qualitative feedback among students, which allowed students to revise and improve their work for final submission. This aligned with overall feedback principles which focus on opportunities to demonstrate improvement. Even though there were some technical issues with allocation of work/peer markers and with releasing feedback information to students, these were relatively quickly resolved. The enthusiasm from the unit coordinators and the maturity of postgraduate students were likely factors that led to higher tolerance for the technical and logistical issues.

Across the two cases, and other implementations within the university, it was found that considerable technical and pedagogical support was usually required to implement new tools. Even when the pedagogical aims and learning opportunities for the given units/courses were clearly identified, to fit and work effectively with the digital tools required further careful consideration and adequate support.

15.3 Are Digital Technologies Fit for Purpose?

As we outlined above, digital technologies can bring a range of affordances to enhance learning opportunities: automation of the allocation of students (including group formation), mark calculation and distribution, anonymisation of feedback, and improved scalability of assessment with large cohorts. Many tools allow (or at least promise) integration with pre-existing assessment engines (e.g., via the LMS), and some allow for tracking of student progress over time, that gives sophisticated learning analytics.

However, no current technology tool encompasses all common principles of self and peer assessment or existing types of use in teaching. Current ‘out of the box’

functionalities do not necessarily promote good assessment design. For example, they may hamper the ability for feedback information to be acted upon, when there is no opportunity to refine work further and demonstrate improvement, and, when assessors are anonymous, it does not allow students to develop relationships with peers where they can learn to trust others' judgements. Technologies may also become obsolete rather quickly, and some may only function through specific learning management systems. Ongoing support for configuring and troubleshooting software is also a consideration. New tools for online collaboration and self and peer assessment continue to be developed. Since the time of the previous tool compilation, we have become aware of other available products such as ReView, Feedback Fruits, and Collab Hero. Any list is likely to be soon outdated.

Educators and policy decision makers therefore need to be aware of the specific affordances and limitations of the tools available to them, and to be aware of new developments. A comparison against the pedagogical aims and design should be made to identify synergies and gaps. Ideally, pedagogy should drive the use of technology, rather than technology constraining pedagogical possibilities. Appropriate support for the technology should also be arranged. Realistically, however, it is possible that the pedagogy will be modified according to the tool affordances, and so careful attention must be paid to the overall aims of the self or peer assessment exercise, to ensure they are still being achieved. Academics also may require additional training to use the tools efficiently.

15.4 Non-technological Factors which Also Need to Be Addressed

Despite the promise of technology, the use of technology alone is not enough to achieve change in peer assessment practices (Adachi et al. 2018b). While we are critical of technology (Selwyn 2007), we are equally critical of human and system factors which prevent better implementations of self and peer assessment. These factors can be broken down into those at micro, meso, and macro levels.

15.4.1 Micro: Students' and Educators' Knowledge, Skills and Attitudes

Both students and educators need to be aware of and competent at their role in any type of assessment. For educators, this includes understanding what outcomes it can be used for, and how to set up an activity. Educators firstly need to be amenable to using self and peer assessment in the first place, and then motivated to learn and change their practice. Attitudes can be difficult to shift, as resistance to change is frequently encountered (Tai et al. 2017). Students also need to have some level of

buy-in to the assessment: they must perceive that undertaking the task is valuable. More frequently than not, marks are assigned for completion of such a task. While marks and grades may incentivise students' participation, their participation may still be tokenistic if they don't fully understand the learning opportunities associated with the task. Explicitly aligning the act of the assessment with learning outcomes can help students to recognise how this activity contributes to their learning. Students also require appropriate skills to undertake the self or peer assessment: this may include knowing how to participate in effective feedback processes, and understanding the standards or quality indicators for the task or aspect of performance they are assessing, as part of their development of evaluative judgement (Tai et al. 2018). Upskilling students and educators is likely to be important: both in their assessment roles, and in their use of technology.

15.4.2 Meso: Availability of Support; Course Design

Structures surrounding educators and students may also help to facilitate self and peer assessments. In a technology-enabled environment, providing support for the use of the technology is important. Most current software tools require some integration with the LMS, customisation and set-up for individual use, and are likely to also require troubleshooting during use. Not all educators are technologically savvy, and even those who are, may not have the correct level of access to systems, to install software and fix fundamental issues. Therefore, ensuring support accompanies technology implementation is important.

In addition to this, the design of units or series of units from a course perspective may better support the use of self and peer assessment. Having similar tasks (at different levels of complexity/difficulty) in subsequent units may allow students to develop better self and peer assessment skills, including refining feedback dialogues, and improved understanding of quality as they gain additional expertise (Boud et al. 2015). This also ties into ensuring various aspects of self and peer assessment skills are embedded across different units and year levels over the duration of a course. Course coherence therefore may be a deciding factor for educators in designing tasks and implementing activities.

15.4.3 Macro: The Policy Environment; Institutional Culture

Institutional policy may also dramatically help or hinder effective implementation of self and peer assessment. If assessment policy has no explicit guidance on how self or peer assessment can be used, educators are likely to feel less supported and be hesitant in implementing it (Adachi et al. 2018a). If policy states that self or peer assessment cannot be used for summative purposes, then this sends a strong message about the situations in which it is appropriate. Conversely, making self and

peer assessment compulsory might result in tokenistic implementations. Institutional culture is also likely to influence how policy is implemented: while individuals might have personal views about self and peer assessment, institutional values will also influence behaviours and practices.

Overall, effective and successful implementation of self and peer assessment will stem from addressing all three levels (micro, meso and macro). Making technology available with no support or education on how to utilise it most efficiently is likely to result in poor outcomes. Both educators and students are crucial players, with surrounding supportive structures, such as support in both pedagogy and technology and institutional support around assessment policy. These factors can either facilitate or hinder the use of self and peer assessment.

15.5 Looking to the Future: Research and Practice

Several aspects could be investigated to improve self and peer assessment in a digitally-enabled higher education context. We divide these into practical aspects that could be implemented already from current evidence, and those that require more research to understand their value.

Digital technologies can enable the carrying out of self and peer assessment in the most authentic ways possible as part of a trajectory of life-long learning within a digital world– e.g., collaborating with peers on teamwork, provide feedback on peer assessments, and options for identifying or anonymising peers. Several technical aspects will also benefit from digital technologies. Rubrics are already integrated in some systems. The next step here is to develop the ability to view multiple overlays of marks, where self, peer, and tutor rubrics use can be viewed simultaneously to be able to interrogate ideas of standards and quality. Sophisticated algorithms and programs can allow more fine-grained considerations in the matching and allocation of peer groups, drawing upon information sourced not only from previous performance, but also learner preferences and individual goals for grade attainment. Furthermore, the provision of learning environments or ecosystems in which students interact is critical. This includes the creation of secure drop box facilities for assignments to reduce concerns about plagiarism, and safe spaces for online feedback dialogue, to facilitate peer interactions. With regards to self-assessment, initial compulsory formative exercises which require little technological use (i.e., submission of an additional file/rubric) may normalise the process of self-assessment, with grading for this task based on alignment to an external assessment. Deeper integration of self and peer assessment tools within already-existing LMS structures will contribute to seamless learning experiences and help ensure data privacy for our students and their assessment work.

More complex and sophisticated strategies may also be needed, relating to overall learning and assessment design. Firstly, audio or video feedback could be required as a standard within peer assessment exercises, which allows students to attend to non-verbal cues and communicate in more responsive ways (see Pitt and

Winstone, Chap. 7 this volume). A required structure for such feedback formats may be necessary to avoid "vanishing" feedback, where nothing impactful is said. Secondly, longitudinal self and peer assessment could be implemented across a program of study, to enable formative feedback opportunities over time, and give students opportunities to practise and develop their skills.

Beyond these suggestions, which are enactments of the principles outlined in the first section of this chapter, there are also other less understood aspects of self and peer assessment which could be investigated through the implementation of digital technologies. Exemplars are known to improve students' understandings of quality (Carless and Chan 2016; To and Carless 2016). The combination of peer assessments through an adaptive exemplar database with opportunities for virtual a/synchronous peer discussion might help us understand how many and what types of peer assessments are necessary for learners to understand quality and develop accuracy in assessment. Even within face-to-face settings, use of audience polling in preparatory discussions prior to and during peer assessments, where individuals' responses are longitudinally tracked may help us understand how standards are adopted by students. Monitoring and reflecting progress over time through a dashboard or portfolio is also likely to be useful for self-assessments. While this has been done in a short-term sense (Boud et al. 2015), the long-term effects beyond the educational context are unknown: are students being prepared for making judgements in their future lives?

In conclusion, our work extends the notion of 'the state of the actual vs state of the art' outlined by Selwyn (2008). In the context of higher education where the use of digital technologies is somewhat romanticised with effective and scalable (self and peer) assessment (i.e., the state of the art), we highlight the importance of considering all the other factors that are contextual and non-technological (i.e., the state of the actual). Humanistic, social and cultural factors have an important role to play for any digital assessment practices to work and we ought to make our best efforts in supporting both academics and students in using digital technologies. Furthermore, institutional facilitators and constraints should also be considered when exploring new technology for self and peer assessment. By exploring the principles in this chapter that guide self and peer assessment, we demonstrate that successful examples of self and peer assessment are a marriage of technology and humanistic elements that foster learning opportunities, rather than a simple inclusion of digital technologies alone.

References

Adachi, C., Tai, J., & Dawson, P. (2018a). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development, 37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913.

Adachi, C., Tai, J., & Dawson, P. (2018b). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*(2), 294–306. https://doi.org/10.1080/02602938.2017.1339775.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Maidenhead: Open University Press.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529–549. https://doi.org/10.1007/BF00138746.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462.

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education, 24*(4), 413–426. https://doi.org/10.1080/0260293990240405.

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: Disruptive effects of assessment patterns. *Higher Education Research & Development, 34*(1), 45–59. https://doi.org/10.1080/07294360.2014.934328.

Carless, D., & Chan, K. K. H. (2016). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education, 42*(6), 930–941. https://doi.org/10.1080/02602938.2016.1211246.

Eva, K. W., & Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions, 28*(1), 14–19. https://doi.org/10.1002/chp.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430.

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137–155. https://doi.org/10.1080/02602930903221444.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121.

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

Molloy, E., & Boud, D. (2013). Changing conceptions of feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding and doing it well* (pp. 11–33). Milton Park: Routledge.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 247–266). New York: Routledge.

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551–576. https://doi.org/10.14204/ejrep.30.12200.

Pitt, E., & Winstone, N. (this volume). Dialogic feedback in the digital world (Chapter 7). In M. Bearman, P. Dawson, J. Tai, R. Ajjawi, & D. Boud (Eds.), *Reimagining assessment in a digital world*. Dordrecht: Springer.

Raat, J., Kuks, J. B. M., van Hell, E. A., & Cohen-Schotanus, J. (2013). Peer influence on students' estimates of performance: Social comparison in clinical rotations. *Medical Education, 47*(2), 190–197. https://doi.org/10.1111/medu.12066.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015.

Selwyn, N. (2007). The use of computer technology in university teaching and learning: A critical perspective. *Journal of Computer Assisted Learning, 23*(2), 83–94. https://doi.org/10.1111/j.1365-2729.2006.00204.x.

Selwyn, N. (2008). From state-of-the-art to state-of-the-actual? Introduction to a special issue. *Technology, Pedagogy and Education, 17*(2), 83–87. https://doi.org/10.1080/14759390802098573.

Sridharan, B., Tai, J., & Boud, D. (2018). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education, 77*(5), 853–870. https://doi.org/10.1007/s10734-018-0305-7.

Tai, J., & Adachi, C. (2017). Peer assessment and professional behaviours: What should we be assessing, how, and why? *Medical Education, 51*(4), 346–347. https://doi.org/10.1111/medu.13254.

Tai, J., Molloy, E., Haines, T., & Canny, B. (2016). Same-level peer-assisted learning in medical clinical placements: A narrative systematic review. *Medical Education, 50*(4), 469–484. https://doi.org/10.1111/medu.12898.

Tai, J., Canny, B. J., Haines, T. P., & Molloy, E. K. (2017). Implementing peer learning in clinical education: A framework to address challenges In the "Real World.". *Teaching and Learning in Medicine, 29*(2), 162–172. https://doi.org/10.1080/10401334.2016.1247000.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

To, J., & Carless, D. (2016). Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education, 40*(6), 746–764. https://doi.org/10.1080/0309877X.2015.1014317.

Joanna Tai is Senior Research Fellow at the Centre for Research in Assessment and Digital Learning, at Deakin University, in Melbourne, Australia. Her research interests include peer learning, self and peer assessment, evaluative judgement, student perspectives on learning and assessment, and research synthesis.

Chie Adachi is Director, Digital Learning at Deakin University, Melbourne. She works on innovation projects around digital learning at the central learning and teaching unit, Deakin Learning Futures. Her research interests lie within the areas of digital and social learning and innovative assessment practices through educational technology.

Chapter 16
Using Technology to Enable a Shift from Marks to Outcomes-Based Assessment

Andrew Cain, Laura Tubino, and Siva Krishnan

Abstract Doubtfire is a web application that brings together an educational model based on constructive alignment and self-regulated learning to support students and educators. This model uses a task-oriented portfolio-based approach to assessment to scaffold and evidence different standards of achievement; backed by frequent formative feedback to support learning. Doubtfire's technology not only draws together educational design principles, it further supports educators by minimising overheads associated with accessing and assessing student work and provides visualisations to monitor student progress. This chapter describes Doubtfire's underpinning educational model and explores the experiences and advantages associated with teaching using this model and associated software.

16.1 Introduction

This chapter focuses on exploring a web application called Doubtfire, and how it supports evidence-based teaching and learning practices, as an example of a system of re-imagined learning and assessment facilitated by technology. Doubtfire (Cain et al. 2018), is an open source web application that provides a fast and efficient means of assessing student submissions (Renzella and Cain 2018). Doubtfire brings together an educational model based on constructive alignment and self-regulated learning enabled by technology to support both educators and students. Doubtfire was created in the context of teaching programming to undergraduate students with a view to promote authentic assessment. This original purpose was combined with the capabilities of a software system to support educators in delivering task-oriented portfolio-based assessment by minimising the overheads associated with accessing and assessing student work, and through the provision of visualisations to help staff

A. Cain (✉) · L. Tubino · S. Krishnan
Deakin University, Melbourne, VIC, Australia
e-mail: andrew.cain@deakin.edu.au; laura.t@deakin.edu.au; siva.krishnan@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_16

and students monitor progress. In this way, educational technology has been used to simultaneously support educators in their work, and improve learning and assessment design for students.

In this chapter we describe Doubtfire's underpinning educational model and its implementation. We start with a summary of the system, and then expand on the task-oriented portfolio-based assessment approach used in Doubtfire to scaffold and evidence different standards of achievement of learning outcomes. The formative feedback process incorporated into Doubtfire will also be described as will be the way in which Doubtfire promotes self-efficacy, time-management skills and evaluative judgement in students.

The experience and advantages that Doubtfire provides to educators will also be explored. The educator's tools, how their workload is impacted by Doubtfire's feedback and marking expectations, and how data can be used to make informed decisions will also be described in this chapter.

16.2 Summary of the Doubtfire Educational Model

Doubtfire was originally designed to support task-oriented portfolio assessment, which promotes the use of multiple small assessment tasks, each aligned to specific unit learning outcomes and a grade. In Doubtfire, students are encouraged to develop self-regulated learning by deciding what grade they are working towards. Students' target-grade will determine what tasks they will need to complete and can be updated at any time.

In Doubtfire, students are able to act upon the feedback information obtained, as tasks are only graded at the end of the unit. Moreover, students can go through iterative feedback cycles until they achieve the learning outcome in question. In this way, Doubtfire supports the conditions necessary to ensure that feedback does improve student learning. Figure 16.1 provides an illustration of the components involved in this web application (Cain et al. 2017).

Doubtfire allows educators to use the information obtained from students' target-grade and their submitted tasks to plan timely interventions, either personalised, addressing individual student's needs, or for the whole group, through educational resources. They can also plan their workload and collect evidence to request further resourcing.

At the conclusion of the unit, Doubtfire makes it easy for students to combine together the artefacts created through their engagement with unit tasks and the reflections in the form of a Learning Summary Report. The Learning Summary Report is a student's reflection on their learning progress and achievement of outcomes, together with their justifications for how they have demonstrated achievement of a particular grade.

Final grading is then performed by teaching staff. This involves reviewing the student self-assessment, grade justification, and reflections from the Learning Summary Report, together with the evidence submitted. Thus, the deliberate

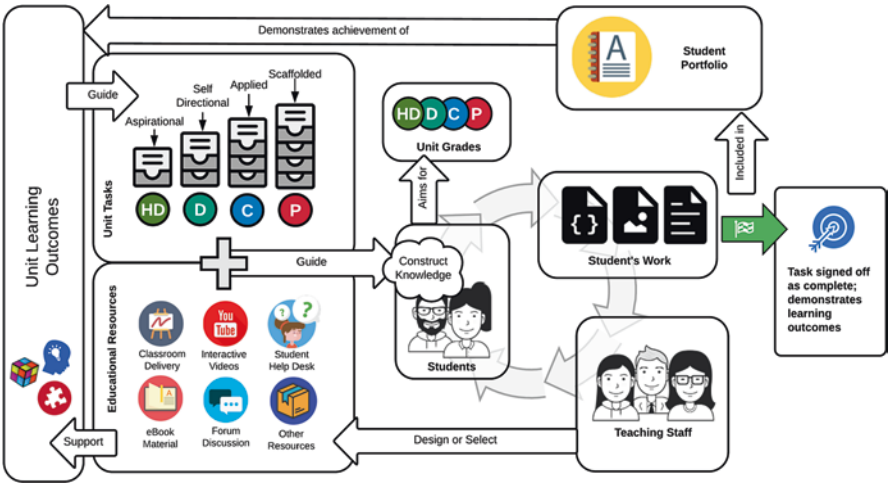


Fig. 16.1 Components of the Doubtfire approach

application of the constructive alignment framework and self-regulation in the design and delivery of task-oriented portfolio-based assessment enables students to develop the required domain knowledge along with evaluative judgement and a capacity to demonstrate achievement of learning outcomes.

16.3 Doubtfire's Assessment Design

16.3.1 Constructive Alignment Using a Portfolio Assessment Approach

Doubtfire assessment is designed to achieve constructive alignment (Cain et al. 2017). Procedurally, constructive alignment involves the definition of learning outcomes using verbs that are likely to require appropriate cognitive levels of activity to achieve (Biggs and Collis 1982). These outcomes then become the focus for assessment, as well as teaching and learning activities. Students work toward demonstrating achievement of these outcomes through engaging with unit assessment tasks (Biggs and Tang 2011). In many situations, assessment is what defines the curriculum from the students' perspective (Ramsden 1992).

Doubtfire promotes a flexible task-oriented portfolio assessment approach. This choice was based in part on Biggs (2011), describing portfolio assessment as achieving a 'web of consistency' through which educators can optimise the likelihood of students engaging appropriately with learning activities and assessment tasks. Educators can incorporate a range of tasks including group and project work, scaffolded tasks, and examinations where appropriate (Cain et al. 2017). In Doubtfire,

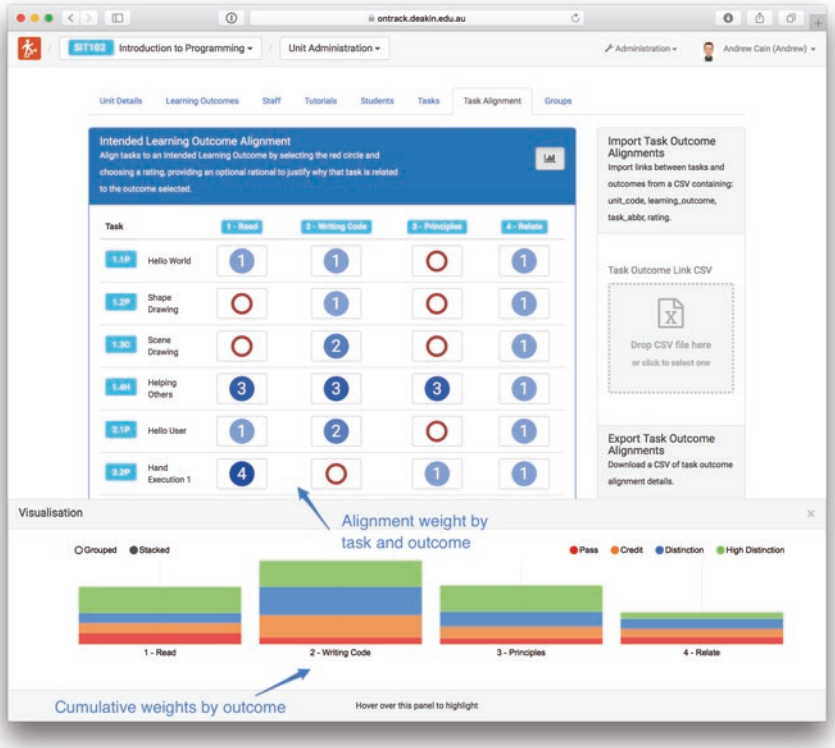


Fig. 16.2 Task alignment to unit learning outcomes illustrate focus of the unit

students prepare a collection of their work and reflections at the end of the unit, that demonstrate they have met unit learning outcomes. The overall grade is determined at the completion of the unit through the holistic assessment of the tasks and the reflections included by the student in their portfolio.

Educators must explicitly state the purpose of assessment tasks in Doubtfire by linking each assessment task to relevant unit learning outcomes. In addition, Doubtfire asks students to link each task to unit learning outcomes and justify their choices by reflecting on what they are expected to know, do and demonstrate in order to successfully complete that particular task. This helps further validate students' achievement of unit learning outcomes. However, this feature of Doubtfire is not compulsory and thus, not always used by educators or required of students.

Figure 16.2 shows an example of the alignment between unit learning outcomes and tasks. In this example the Unit Learning Outcomes relate to reading and analysing program code, writing code to implement program features, exploring the concepts associated with the principles of structured procedural programming, and developing the ability to reflect upon learning through relating tasks to learning

outcome achievement. Teaching staff and students each have their own alignment, using the same visualisation. In Fig. 16.2, we can see the focus of the unit tasks are more around the second intended learning outcome, as shown in the visualisation at the bottom of the screen. This visualisation shows a weighted sum of the task weightings for all tasks that align to this outcome, which takes into consideration both higher graded and higher weighted tasks.

16.3.2 Evidencing Different Standards of Achievement in Doubtfire

Internally, Doubtfire promotes multiple small assessment tasks, each aligned to the unit learning outcomes, with specific tasks targeted at each grade standard. Unit delivery is centred on student activity rather than discipline content. Feedback is incorporated in such a way that it supports students in achieving and demonstrating every unit learning outcome. The tasks provide a framework to guide students through a learning journey that should culminate in them having sufficient work to demonstrate achievement of each one of the unit learning outcomes. In other words, tasks will allow students to both, develop and demonstrate their knowledge, understanding, skills and attitudes.

Tasks are designed around grade levels, so educators must be explicit about the different standard of performance required at each level. The tasks and grade levels relate to different standards of achievement, and can be designed around the levels of the structure of observed learning outcomes (SOLO) taxonomy (Biggs and Collis 1982). If we consider the grade level outcomes to range from Pass to High Distinction, Pass and Credit would require at least a multi-structural level of understanding, meaning that the different concepts have been learned but the relationships between the concepts is not yet understood. Distinction and High Distinction tasks would require a relational or extended abstract level, which means that the relationship between concepts needs to be clear so these can be seen as a coherent whole and can ultimately be used as the basis to predict, generalise and create new understanding.

An incremental design allows students to evidence their chosen standard of achievement. All students undertake the Pass tasks: the scaffolding in these tasks helps support student learning to achieve the learning outcomes. The output of the Pass tasks demonstrates achievement of the learning outcomes to a minimally acceptable standard. Students aiming to achieve a higher than Pass grade need to complete the Credit tasks in addition to the Pass tasks. Credit tasks build upon the Pass tasks to provide additional challenges. Students may need to review associated concepts or build further depth in the concepts presented. Distinction tasks build upon the challenge level set in the credit tasks. Tasks at this level would usually be less structured to give students greater flexibility and responsibility in how this standard is met. One strategy for the Distinction tasks is to have students apply what they have learnt in a domain or context of their own choosing, or to have them

demonstrate the ability to synthesise aspects covered in the unit in a more complex scenario. These tasks can generally push students beyond the content covered in the unit, allowing them to demonstrate the use of the knowledge and skills learnt in a broader context. High Distinction tasks are then generally set at an aspirational level, providing students with a challenge through which they will be able to demonstrate excellent achievement of the unit learning outcomes. With students having already demonstrated achievement of the distinction level standard, high distinction tasks can be more open-ended allowing students to explore related concepts and skills.

For example, in an undergraduate programming unit the pass tasks could scaffold student development of their programming knowledge and skills. This may include guided tasks where students apply the concepts covered to build small programs, and accompany these with explanations of different aspects that are the focus of each task. Credit tasks could challenge students to re-apply what they have learnt with less guidance and in more challenging scenarios. At the Distinction level, students could take what they have learnt and demonstrate the ability to use these concepts in the creation of a program of the student's own design, allowing them to engage with the material in a way that is relevant to their own desires and aspiration. High Distinction tasks could require the custom program to meet certain quality indicators, showing a higher degree of sophistication or elegant application of the principles covered.

In this way each of the grade levels has a different focus, which should be clear to both students and teaching staff. Clearly indicating this in the task sheets, and associated instructions, can help reinforce the focus. The greater challenge is in determining the levels initially, and planning out tasks that will both develop student capability and demonstrate desired outcomes.

16.3.3 Feedback for Learning

Substantial learning gains can be made by initiatives designed to strengthen the feedback students receive (Black and Wiliam 1998). Good feedback practice can be broadly defined as anything that might strengthen the students' capacity to self-regulate their own performance (Nicol and Macfarlane-Dick 2006). Doubtfire's assessment design capitalises on the benefits of formative feedback for student learning. Each completed task can be submitted to teaching staff for formative feedback during established timeframes. Submission of work for formative feedback provides staff with a contact point they can use to direct student efforts, helps ensure students are demonstrating the expected standard and keeps them on track.

Students can also submit their tasks for feedback more than once, as giving feedback on work in progress can enhance student learning (Carless et al. 2011). Providing feedback rather than marks on drafts is also useful because giving marks changes students' views of the learning task and obscures feedback (Dahlgren et al. 2009). However, only writing feedback comments on students' work and not

providing a grade can result in more learning (Black and Wiliam 1998). A staged process that gives feedback on drafts, like the one promoted in Doubtfire, is likely effective because it provides the motive and the opportunity to engage with and use the feedback (O'Donovan et al. 2016).

Grades will only be provided at the conclusion of the unit based on the portfolio submitted by the student. If a student fails to submit tasks for feedback they can still include these completed tasks in their portfolio at the conclusion of the unit, to be considered in their summative assessment. A very important feature about this feedback process is that it doesn't require more work on the educator's part. The marking effort is only moved forward to the draft stage with a grade given at the end for the final submission (O'Donovan et al. 2016).

16.3.4 Formative Feedback Process in Doubtfire

Within Doubtfire, staff provide formative feedback through comments and through a task status. The task status helps communicate what is required by the student. When students submit work, staff provide feedback that may require the student to address issues associated with their work and resubmit it for further assessment. This usually occurs once, but there is also the possibility that resubmitted work may not have fully addressed the issues identified and the student may need to make further corrections if they have not achieved the intended learning outcomes at the required standard.

The following lists enumerates the different task statuses available in Doubtfire. Student associated statuses include:

- **Not Started** – all tasks start here.
- **Working on it** – for students to indicate their current focus.
- **Ready for Feedback** – this is set when students submit work for feedback and indicates to teaching staff that this work is ready to be checked.

Staff can then provide feedback using the following statuses:

- **Fix and Resubmit** – indicates that the task is close, but there are some issues that need to be addressed. The student needs to fix these issues and resubmit the work.
- **Redo** – similar to Fix and Resubmit but indicating that there are significant issues with the work. For example, the work was submitted to the wrong task, or completely missed the point. This would indicate a need to start again.
- **Discuss** – means that the task is sufficiently demonstrating the learning outcome, however, if this discussion identifies misunderstanding, then the student may be asked to develop their understanding further and resubmit. This mechanism can also help ensure the student is the author of the work submitted.
- **Complete** – indicates the task demonstrates the required learning and should be included in the portfolio as is or with very minor fixes.

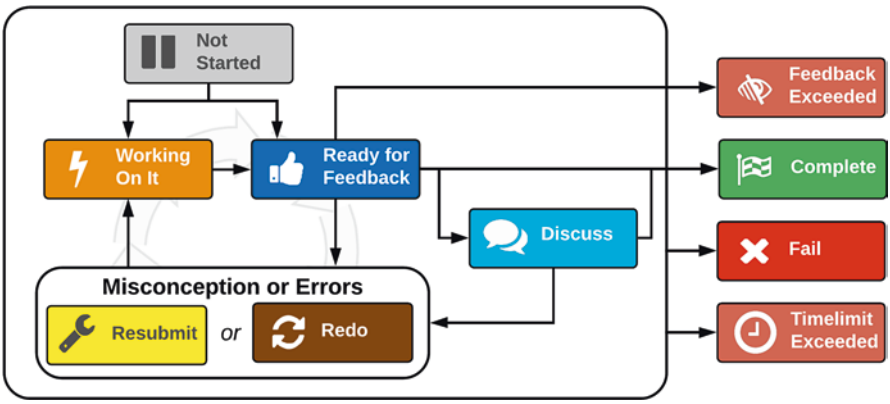


Fig. 16.3 Task statuses, and the transitions between these states

- **Feedback Exceeded** – provides staff with an option to break the feedback loop. This would be used when the student has needed more than a reasonable number of attempts at the task. This indicates to the student that they need to fix the task and ensure it up to standard before it is included in their portfolio.
- **Time Exceeded** – similar to Feedback Exceeded, except that this indicates the task was submitted after the set deadline for the task. As with Feedback Exceeded, students would need to ensure the task was up to standard before it was included in their portfolio.
- **Fail** – indicates that some aspect of the task was failed. For example, this may be due to a time limit for successful completion of a task. It indicates the task was not completed successfully. This relates to a specific task, not to the overall portfolio.

Figure 16.3 shows the different statuses available in Doubtfire and the possible transitions between them. A sequence of interactions indicates the iterative formative feedback loop embedded within Doubtfire.

16.4 Doubtfire Promotes Self-Regulated Learning

Self-regulated learning is an active process in which students set goals for their learning and monitor, regulate, and control their thinking, behaviours and motivation, while responding to external feedback and their environment (Nicol and Macfarlane-Dick 2006). Self-regulation is essential for students to be able to continue learning outside of formal education and thus promotes sustainable assessment (Boud and Soler 2016). Doubtfire provides many opportunities for students to apply and develop self-regulation, assuming control over their own learning. We will frame Doubtfire's opportunities for developing and demonstrating self-regulated learning within Zimmerman's proposed model, which organises self-regulated learning in three phases: forethought, performance and self-reflection (Zimmerman 2000).

16.4.1 Forethought Phase: Goal Setting and Self-Efficacy

Self-efficacy is the belief in one's ability to succeed in specific situations or accomplish a task (Bandura 1991). Stronger self-efficacy results in higher goals and aspirations and higher commitment to them (Bandura 1991; Tosi et al. 1991) and can be used to predict someone's desire or ability to complete a task (Gist and Mitchell 1992; Stajkovic and Luthans 1998).

Doubtfire promotes structuring assessment tasks around graded levels, where students choose a target-grade and complete all the task at that level and below but do not have to complete any tasks above that level. This provides an opportunity for students to be actively involved in setting goals associated with what they want or believe they are able to achieve and might improve their self-efficacy. Where a student is struggling with the targeted-grade they can change their target, ensuring that they are able to demonstrate all of the learning outcomes to a lower grade standard and thus will develop a better understanding of what their personal standard is. The opposite can also occur, where a student finds the targeted-grade too easy to achieve and decides, or is encouraged, to try the next level up.

16.4.2 Performance Phase: Self- and Time-Management

In terms of student self-management, as the summative grading of the students work only occurs at the end of the unit, there is a need for students to demonstrate better self- and time-management skills. Doubtfire helps students further develop these skills by tracking progress through the unit tasks (Cain et al. 2013; Woodward et al. 2013) and enabling students to visualise their progress toward achieving the unit learning outcomes (Law et al. 2016).

Doubtfire's student dashboard can be seen in Fig. 16.4, this is the main view through which students interact with it. Student attention is focused on the task list, on the left of the screen, which provides an ordered list of tasks for the student to work through to achieve their target grade. Different colours and icons are used to visually convey the status of each task in the list, which progress from the grey un-started state to a green tick when complete. Assigning small visual badges to the tasks helps gamify the experience, with students working to get all of the tasks 'green'. Students need to select tasks from the task list to view associated resources, and to trigger state changes.

This is further supported by a progress visualisation using a burn-down chart to show the volume of work remaining and their current velocity; an approach adapted from agile software engineering practices (Sutherland and Schwaber 2007; Woodward et al. 2013). The burn-down chart plots progress against time, with student's average historical progress being projected into the future to give an estimated completion time (Law et al. 2016; Woodward et al. 2013). These visualisations aim to communicate progress, encourage engagement and ensure students do not unwittingly fall behind.

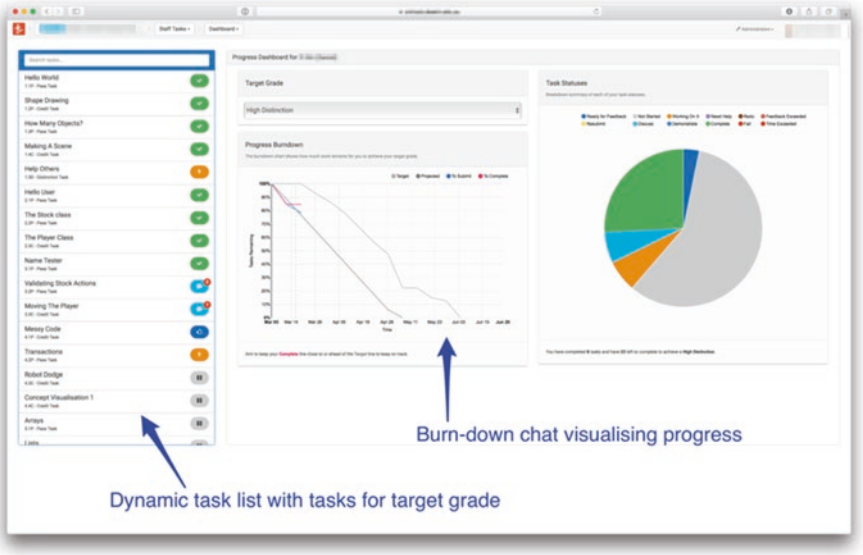


Fig. 16.4 Student dashboard showing dynamic task list and progress visualisation

16.4.3 Self-Reflection Phase: Self-Assessment and Evaluative Judgement

In Doubtfire, the goal setting and self-assessment processes have great implications for students' summative assessment. As mentioned earlier, it is well recognised that summative assessment defines the curriculum for the student (Ramsden 1992). By making self-assessment part of summative assessment, and by placing the emphasis of self-assessment on eliciting and using criteria and standards (Thompson 2016), the Doubtfire model helps students develop evaluative judgement. Evaluative judgement is the capability to make decisions about the quality of one's own and others' work, and it helps students become independent learners (Tai et al. 2017). Furthermore, according to Cowan (2010), students who have first defined their own goals are more determined to achieve them and don't doubt they are worthwhile. They also monitor their progress against their chosen goals frequently and adjust their approach accordingly. Consequently, these students may rate their learning experience as being more effective and purposeful (Boud 1997).

In Doubtfire, once all the tasks are completed, students perform a self-assessment of their performance in relation to the unit learning outcomes and justify how they have demonstrated achievement of their target-grade. Students also need to reflect on their growth throughout the unit. In their reflection, students can refer to evidence captured in the artefacts they created as they progressed through the unit, to the tasks performed and how these helped them develop skills or to the feedback obtained and how it helped them better understand standards and learning outcomes.

Additional evidence can be attached to further support the student's claims, for example, evidence of assisting their peers.

Student's self-assessment, reflections on their learning progress, and justification of their proposed grade, make up their Learning Summary Report. Once complete, the Learning Summary Report becomes the cover document for the portfolio and is submitted for summative grading. A parallel can be drawn between this format, and typical job applications where applicants need to demonstrate how their experience meets the job requirements. The Report may therefore be also perceived as a relatively authentic assessment component.

16.5 Summative Assessment

Summative assessment and grading is performed at the conclusion of a unit, by teaching staff on each student's portfolio as a whole. This involves reviewing the student self-assessment, grade justification, and reflections from the Learning Summary Report, together with the evidence submitted.

The task status for each task from the iterative feedback process can help inform the summative grading. Usually the work produced by students has already been checked and marked off as complete, which means they have demonstrated the expected learning outcomes. Where not all work is marked as complete, a judgement needs to be made to determine which grade standard has been demonstrated through their work. Typically, the grade level details are easily determined based upon the status of each of the student's tasks.

Final numeric results can then be determined through the use of a rubric, as shown in Fig. 16.5. Once the grade level is determined based on the artefacts submitted, as described above, the quality of the work presented, depth of reflections, and clarity of the justifications can be used to position the grade within the awarded grade band.

The Learning Summary Report plays a crucial role in guiding the assessment process and understanding student's standard of achievement. In cases where doubt exists, a face-to-face or video discussion is used to identify any misunderstanding and confirm the standard of achievement of learning outcomes by the student. This mechanism can also help ensure the student is the author of all work submitted.

To ensure consistency, the whole teaching team is usually involved in this process. Each portfolio can be assessed by two tutors, both of whom review the Learning Summary Report and quickly inspect key tasks. They then agree upon a grade band, and on the position within this band. Initially these decisions can be shared across the whole team to ensure consistency across marking pairs. Once agreement is reached in relation to standards, then only edge cases need to be brought back to the whole group for discussion.

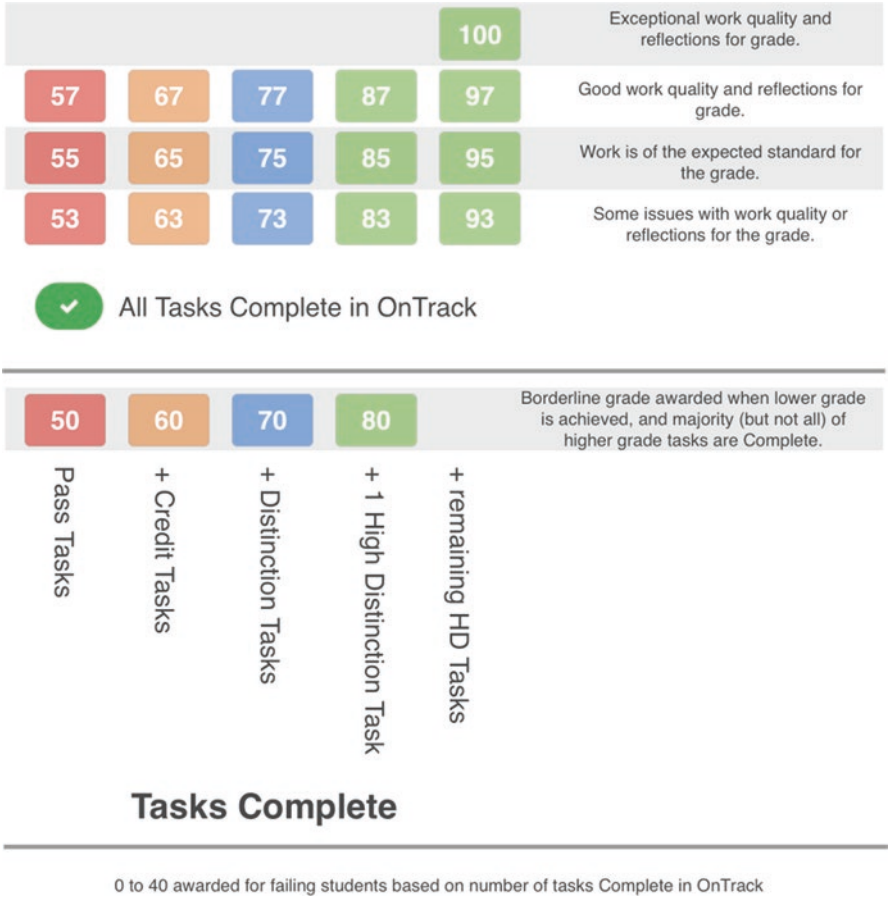


Fig. 16.5 Rubric used to determine student grades

16.6 Doubtfire Supports Educators

Doubtfire's technology not only draws together the educational design principles described above, it also supports educators in applying these by minimising the overheads associated with accessing and assessing student work and providing visualisations related to student progress.

For educators, the main advantages of using Doubtfire are:

- Scaffolds them through designing their unit in a way that incorporates some of the best practices in teaching and learning.
- Makes tracking students' goals, submissions and feedback manageable for big cohorts.
- Distributes their marking workload throughout the unit rather than leaving it all to the end.
- Gives them readily available data that they can use for personalised or group interventions.

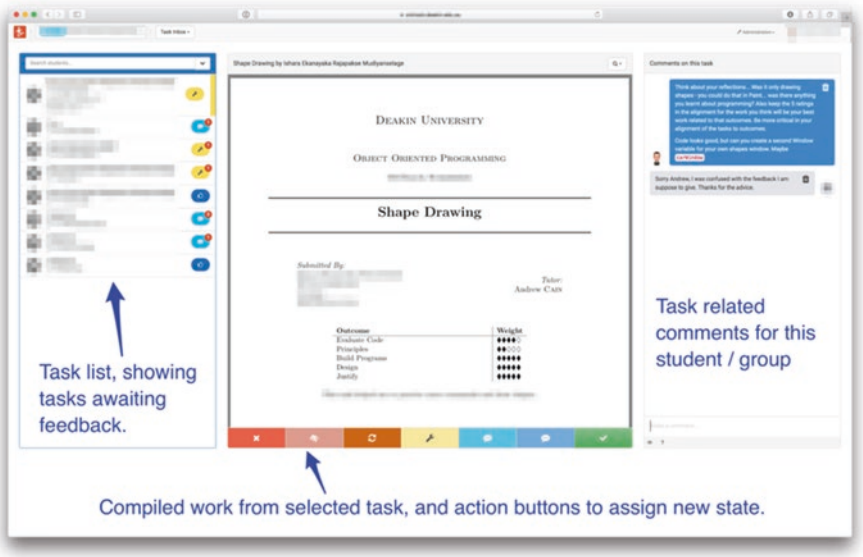


Fig. 16.6 Staff inbox with tasks awaiting feedback and fact action buttons to assign new status

16.6.1 Educators' Tools in Doubtfire

The Task Inbox is the main educators' view in Doubtfire and is configured as three columns, as shown in Fig. 16.6. To the left is the list of tasks, this will include tasks that require feedback or have new comments. When a task is selected, the centre panel displays the associated submission with action button that can be used to assign new states to the selected task. To the right is then a panel for comments, which is used to communicate feedback.

Staff work through the tasks in their inbox, viewing the submissions, providing comments, and updating task statuses. Each of these interacts is designed to require as few interactions as possible, thereby minimising the overhead associated with accessing each submission (Renzella and Cain 2018).

16.6.2 Feedback Workload and Impact

The workload of educators can be very high when feedback is seen as a transmission of information from educator to student (Nicol and Macfarlane-Dick 2006). In Doubtfire, feedback is responsive to students, only students that want to use the feedback provided will submit their tasks for feedback. This reduces the educator's workload and makes their feedback more effective.

Feedback not only becomes more effective because it is responsive to students who are motivated to use it, but also because there is an open dialogue between educator and student, on the right panel for comments, that can be used by the student and the educator to ensure that the student understands and can apply the feedback (Nicol and Macfarlane-Dick 2006).

By facilitating timely and iterative feedback in this way, it can be said that Doubtfire 'gamifies' feedback. Games make feedback cycles fast, keeping the stakes low and thus maintain a positive relationship with failure (Pope 2003). We believe Doubtfire does the same thing. The investment educators and students put in their communication is focused on delivering a message fast, with the awareness that when that message is not understood the other party will come back and ask further questions thus reducing the time and effort educators need to put in the form of the feedback, rather than its substance.

16.6.3 Portfolio Marking Workload

Using this process, we have been able to scale this kind of portfolio assessment up to large student cohorts, increasing the number of portfolios marked by each teaching staff member in a certain amount of time. The iterative formative feedback process is key to this rapid turnaround, as the task statuses provide valuable information that speeds up the summative assessment process. In effect, by focusing on formative feedback, staff assessment efforts are directed to helping students develop appropriate understanding and skills, supporting student learning and aligning staff and student efforts.

Doubtfire also minimises the overhead associated with accessing each submission (Renzella and Cain 2018), which makes them much more easily and rapidly accessible, as compared to paper submission or other forms of digital submissions. When working with cohorts of hundreds of students, the time saved in Doubtfire on accessing submissions can add up to hours of work.

16.6.4 Using Data to Make Informed Decisions

Educators can use the information obtained from students' target-grade and submitted tasks to plan timely educational interventions. This can be personalised, addressing individual student's needs, or targeting the whole group. Figure 16.7 shows an example visualisation of the task statuses for each task. The proportions in the pie charts show how many students currently have this task in a particular status and can be used to monitor student progress through the unit tasks.

Once the unit is finalised, data and statistics on each task, grade level or learning outcome can be used to evaluate the educational resources in a detailed manner, knowing exactly what did and didn't work, and achieve better outcomes on the next iteration of the unit.

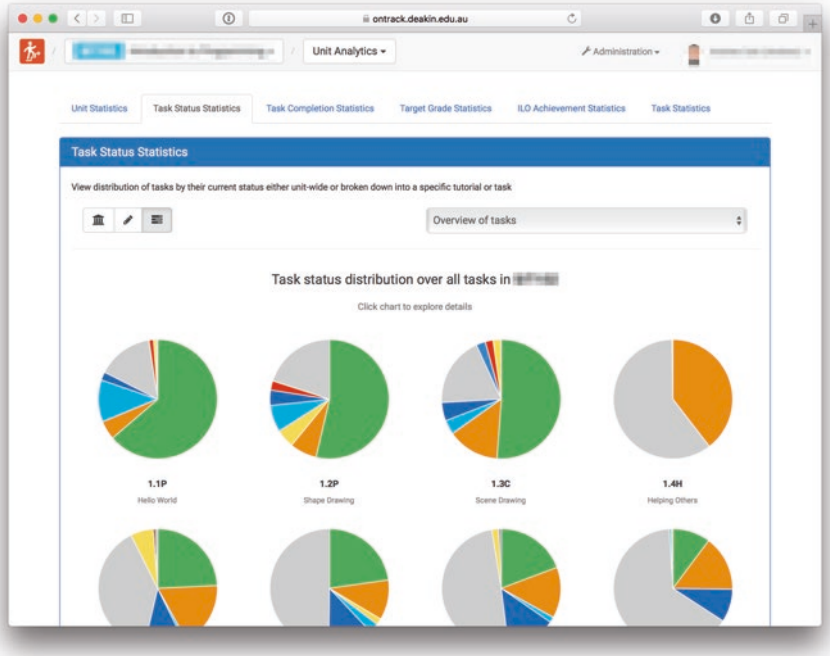


Fig. 16.7 Example visualisation showing task status proportions by task

16.7 Implications and Future Directions

As a system developed upon evidence-based learning design, Doubtfire has been effective in promoting feedback cycles, self-assessment and reflection, and in reducing the marking workload for teaching staff.

The main challenges in the implementation of Doubtfire have been around the design of tasks for the different grade levels. Although this model recommends the standards of achievement to be designed around the levels of the SOLO taxonomy, it has been noticed in some implementations that the different grade levels do not focus on increased complexity but quantity of work. Something similar is observed with feedback, which in many cases does not promote learning but is comprised mainly of instructions that need to be followed without further reflection. More clear instructions and guidance will soon be included in Doubtfire to further support educators with the design of feedback and assessment.

Further features to be developed in Doubtfire will support peer assessment, aligned with the idea of sustainable assessment. Peer assessment further develops evaluative judgement and in combination with self-assessment, is likely to establish a basis for students to become judges of their own learning (Boud and Falchikov 2007; Boud and Soler 2016; Tai et al. 2017).

To further understand the benefits of Doubtfire, the impact that Doubtfire has on educators' knowledge and application of the educational design principles should be investigated. This would also be fruitful for further refinement of Doubtfire's educator interface to support the professional development of university educators in their teaching and learning scholarship.

References

Bandura, A. (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Current theory and research in motivation* (Nebraska symposium on motivation 1990: Perspectives on motivation) (Vol. 38, pp. 69–164). Lincoln: University of Nebraska Press.

Biggs, J., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. New York: Academic.

Biggs, J. B. (2011). *Teaching for quality learning at university: What the student does*. UK: McGraw-Hill Education.

Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university*. Milton Keynes: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Boud, D. (1997). Enhancing learning through self-assessment. *Assessment in Education: Principles, Policy & Practice, 4*(2), 319–324. https://doi.org/10.1080/0969594970040209.

Boud, D., & Falchikov, N. (2007). Developing assessment for informing judgement. In D. Boud & N. Falkchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term*. London: Routledge. https://doi.org/10.4324/9780203964309.

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133.

Cain, A., Woodward, C. J., & Pace, S. (2013). Examining student progress in portfolio assessed introductory programming. In *Proceedings of the 2013 IEEE international conference on teaching, assessment and learning for engineering (TALE)* (pp. 67–72). Bali: IEEE.

Cain, A., Grundy, J., & Woodward, C. J. (2017). Focusing on learning through constructive alignment with task-oriented portfolio assessment. *European Journal of Engineering Education, 43*(4), 569–584. https://doi.org/10.1080/03043797.2017.1299693.

Cain, A., Renzella, J., Cummaudo, A., & Wilson, R. (2018). *Doubtfire LMS repository*[Computer software]. Retrieved from http://github.com/doubtfire-lms.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407.

Cowan, J. (2010). Developing the ability for making evaluative judgements. *Teaching in Higher Education, 15*(3), 323–334.

Dahlgren, L. O., Fejes, A., Abrandt-Dahlgren, M., & Towald, N. (2009). Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education, 14*(2), 185–194. https://doi.org/10.1080/13562510902757260.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17*(2). https://doi.org/10.5465/AMR.1992.4279530.

Law, C. Y., Grundy, J., Vasa, R., & Cain, A. (2016). An empirical study of user perceived usefulness and preference of open student model visualisations. In *Proceedings of the 2016 IEEE symposium on visual languages and human-centric computing (VL/HCC)* (pp. 49–53). Cambridge: IEEE. https://doi.org/10.1109/VLHCC.2016.7739663.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090.

O'Donovan, B., Rust, C., & Price, M. (2016). A scholarly approach to solving the feedback dilemma in practice. *Assessment and Evaluation in Higher Education, 41*(6), 938–949. https://doi.org/10.1080/02602938.2015.1052774.

Pope, D. C. (2003). *Doing school: How we are creating a generation of stressed out, materialistic, and miseducated students*. New Haven/London: Yale University Press.

Ramsden, P. (1992). *Learning to teach in higher education*. London/New York: Routledge.

Renzella, J., & Cain, A. (2018). Supporting better formative feedback in task-oriented portfolio assessment. In *Proceedings of the 2017 IEEE 6th international conference on teaching, assessment and learning for engineering (TALE)* (pp. 360–367). Hong Kong: IEEE. https://doi.org/10.1109/TALE.2017.8252362.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/00332909.124.2.240.

Sutherland, J., & Schwaber, K. (2007). *The scrum papers: Nuts, bolts, and origins of an agile framework*. Cambridge, MA: Scrum.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2017). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Thompson, D. G. (2016). Marks should not be the focus of assessment — But how can change be achieved? *Journal of Learning Analytics, 3*(2). https://doi.org/10.18608/jla.2016.32.9.

Tosi, H. L., Locke, E. A., & Latham, G. P. (1991). A theory of goal setting and task performance. *The Academy of Management Review, 16*(2), 480–483. https://doi.org/10.2307/258875.

Woodward, C. J., Cain, A., Pace, S., Jones, A., & Kupper, F. J. (2013). Helping students track learning progress using burn down charts. In *Proceedings of the 2013 IEEE international conference on teaching, assessment and learning for engineering (TALE)* (pp. 104–109). Bali: IEEE.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). New York: Academic. https://doi.org/10.1016/B978-012109890-2/50031-7.

Andrew Cain is an Associate Professor of Software Engineering at Deakin University and Associate Head of School (Curriculum) for the School of Information Technology. Andrew works with the Deakin Software and Technology Innovation Laboratory on software innovations associated with learning. In this role Andrew leads the development of the Doubtfire learning management system, and a number of other projects targeted at improving student learning outcomes with a focus on first year programming units.

Laura Tubino is a Lecturer in Curriculum Design in the Faculty of Science, Engineering and Built Environments at Deakin University. She has extensive experience in curriculum and assessment design.

Siva Krishnan is a Senior Lecturer in Learning Futures at Deakin University. He leads a multidisciplinary team responsible for supporting curriculum enhancement initiatives in the Faculty of Science, Engineering & Built Environment. He has published in the area of problem-based learning and learning and assessment design in higher education.

Chapter 17
Challenges with Designing Online Assessment to Support Student Task Understanding

Tiffani Apps, Karley Beckman, and Sue Bennett

Abstract While online assessments can provide new opportunities to redesign traditional assessments, they also present challenges related to design and communication to support students. Open-ended online assessment may place more onus on students to make choices, interpret assessment requirements and self-regulate their learning. This chapter presents some challenges of re-imagined online assessment design for teachers and students in one case study example, drawn from a collective case study. The chapter outlines student, teacher and contextual challenges that affect how a task may be interpreted and reflects on practical implications for the design of online assessment for learning.

17.1 Introduction

Online technologies offer opportunities for higher education to re-imagine assessment in a digital world through increased flexibility that allows students to progress at their own pace any time anywhere, with new possibilities for sharing work and collaborating outside formal class interactions (Bennett et al. 2017). A simultaneous trend towards exploring more student-centred assessment approaches invites us to conceive of new ways to realise 'assessment for learning' through open-ended online tasks. Open-ended online assessment often places more onus on students to make choices, interpret assessment requirements and self-regulate their learning. Thus, there are important implications for how these tasks are designed, communicated and supported.

This chapter presents some challenges of re-imagined online assessment design for teachers and students in one case study example. The case study presented is drawn from a collection of case studies conducted in four Australian universities

T. Apps (✉) · K. Beckman · S. Bennett
School of Education, University of Wollongong, Wollongong, Australia
e-mail: tiffani@uow.edu.au; karleymc@uow.edu.au; sbennett@uow.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_17

that explore students' self-regulated learning in online assessment tasks. The online assessment tasks were specifically chosen to allow us to explore designs that were sufficiently open-ended to permit student choice and interpretation, and that integrated online technologies in ways that were intended to support either new forms of assessment or better scaffold more familiar forms. Some tasks were fully online, while others were supported by a combination of online and face-to-face interactions.

All were relatively modest attempts to integrate digital technologies into assessment by regular university teaching academics, none of whom were particularly skilled with technology. Furthermore, none of these redesigns were conducted under the auspices of a project supported by specific funding. As such, all were examples of routine assessment design work, which was initiated by university academics. This is important because our goal was to neither generate or focus on 'innovations', but to explore real-life examples of technology integration that could inform us about the current state of technology integration – the 'state of the actual' rather than the 'state of the art' (Selwyn 2010). By asking what is occurring and why, rather than what 'could' or 'should' be, we move towards understanding what might new forms of digital assessment be realistically and sustainably achieved by the current workforce or through incremental change.

This chapter details one selected case study from the broader study to highlight how student, teacher and contextual factors affect students' engagement with online assessment in practice. The selected case study highlights the challenges experienced when re-imagining assessment for both teachers and students, particularly in interpretation of this online assessment, and considers the implications for the design of online assessments.

17.2 The Research Study

We undertook a large-scale study of student learning processes within open-ended online assessment tasks. The broader study comprised 11 embedded case studies across four Australian universities. Each case study focused on a formal online assessment task within a semester-long university unit. Data collected from unit coordinators, the teaching team and students, including learning management system (LMS) data logs, assignments, questionnaires and interview data, were used to explore the range of teacher, student and contextual factors that shaped students' engagement with the online assessments.

We conceptualised the interrelations between teaching and learning processes for a specific task using an overarching framework based on Biggs' (2001) generalised model from the student approaches to learning research tradition, adding an important new element of self-regulated learning (Pintrich 2004; Winne and Hadwin 1998). Adapted from these models, Fig. 17.1 outlines our conceptual framing, demonstrating the interrelations in the teaching and learning processes, as well as the

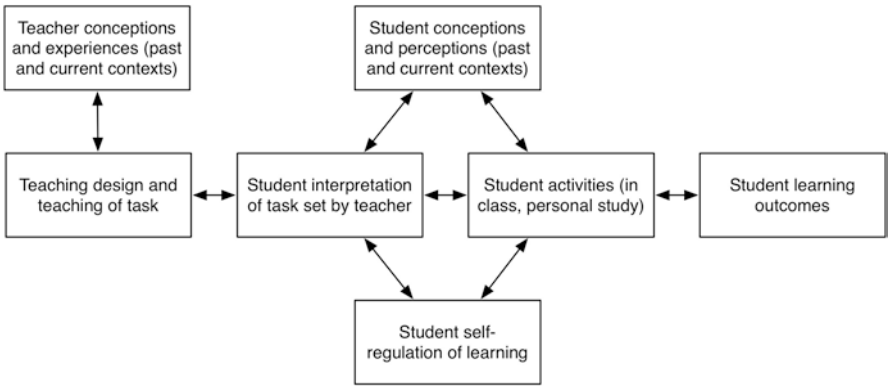


Fig. 17.1 Overarching framework for conceptualising teaching and learning in higher education

different stages of student engagement in learning and self-regulation, all situated within the broader teaching and learning context.

Approaches to learning are contextually shaped by the current teaching and learning context, including the influence of instructor, peers, the institution, department and discipline. Within this context, practices are shaped by how students conceive learning and studying, their understanding of the current task context (including the subject and course within which it sits, as well their current teachers and peers), and by their previous experiences of learning and assessment practices. Within the learning context students actively interpret the task designed by the teacher, and it is this interpretation that shapes their approach to learning and their learning outcomes (Fig. 17.1). Integrating self-regulated learning offers a further framework for conceptualising relevant processes and factors. In our research, we have used Winne and Hadwin's (1998) model of self-regulated learning as a recursive, weakly sequenced set of four stages whereby students: (1) develop their understanding of the task and its requirements, (2) create their own goals and plans based on that understanding, (3) adopt study tactics appropriate to their goals and plans, and (4) meta-cognitively monitor their progress and make any adaptations necessary to their understanding and/or approach. This detailed model draws our attention not only to students' activities, but also the factors that influence those activities.

The case study presented in this chapter was a first-year philosophy unit of 387 students and 6 teachers. The initial idea for the online assessment task was to create a number of brief sequential activities designed to support students' understanding of the particular conventions of essay writing in the philosophy discipline. The task consisted of three activities and was assigned a weighting of 15%. The intention of the task design was to support students to develop skills in composing a philosophical essay, prior to the subsequent assessment.

All students enrolled in the subject were invited to participate in the study. This chapter presents data collected from 313 students using an open-ended item from two questionnaires that asked students to outline their understanding of the task in their own words. The first questionnaire was administered after students had

received a description of the task but before they had started to work on it, and the second questionnaire was administered after completion of the assessment. The study also collected interview data from 12 of the students, who had been randomly selected from a pool of volunteers from the broader student participant group. The purpose of the interviews was to uncover contextual factors that contributed to students' understanding of the task across the assessment period.

The unit coordinator and four of the five tutors were interviewed twice, prior to and after the teaching session. The purpose of these interviews was to elicit the understandings of the task developed by the members of the teaching team, and in so doing consider how these shaped the design of the task and how the task was explained to students. The unit coordinator, Derek, described the purpose of the task in his initial interview as follows:

> So this is a kind of first year level course about how to write an essay but it's pitched not at the technical stuff; there's plenty of material on like how to do footnoting right or how to find resources in the library – all that kind of stuff I'm assuming is covered elsewhere. This is more around, try and understand the more abstract aspects, the harder to get your head around aspects of what an essay actually is, what its purpose is, what the parts of an essay are, what function they play in the essay, how to put them together in the right way, how to decide if something is relevant to your essay or not and trying to give students a slightly better picture of what I think an essay is, okay, and what criteria I'm actually using when I read their essays and mark them. (Derek, subject coordinator, initial interview)

In essence, Derek wanted his students to understand that the philosophy essay has particular conventions distinct from the academic essays his first-year students were used to. This was the first time the assessment task had been implemented in an attempt to support students to better understand the expectations of philosophical essay writing. Derek's aim was for students to complete the assessment task during their independent study, without impinging on class and staff time. His vision was to use online technology to provide self-paced, open-ended activities to support students to build a new understanding of the underlying conventions of the essay form as it is used within the discipline of philosophy.

17.3 The Design of the Task

Derek had quite sophisticated aspirations for the design of the task. The intention was for students to engage with a series of formative online assessment tasks that would lead them to think more deeply and reflect upon how they approached the subsequent assessment of writing an essay. Figure 17.2 outlines the online assessment task as it appeared in the unit outline at the beginning of semester.

To supplement this brief task description, further details about the task were communicated to students via a number of channels including on the LMS, in lectures and in class by tutors. The final design for the assessment task comprised three activities, two to be completed online and one in class. The first online activity featured an audio recording of the unit coordinator explaining the key features and

Assessment 1: Critical contribution

Marking:	Marked out of 100 - Percentage of total subject mark % 15
Description:	The first assignment consists of a short course on the basics of essay writing, consisting of short instructional videos and comprehension quizzes, culminating in preparing an essay plan for your essay.
Due Date:	Opens start of week 3, must be completed by end of week 5 – Sun 28 Aug.
Graduate Quality developed:	Effective communicators
Format:	Online quizzes and submission of an essay plan.
Assessment Criteria:	Quizzes will test comprehension of instructional material on essay writing.
Submission Method:	Online via Moodle

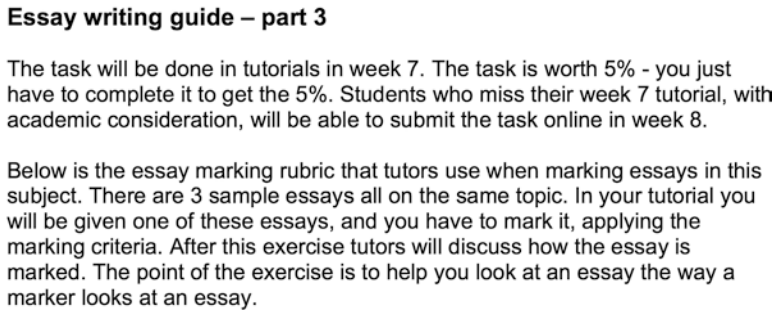
Fig. 17.2 Critical contribution task official description in the unit outline

conventions of an academic essay in philosophy. Students could listen to this recording as many times as they wished and then completed 10 multiple-choice questions. Students had one opportunity to attempt the quiz and were allocated 20 min. After completion, students received their score and the correct answers to the questions.

The second online activity required students to organise the paragraphs of a sample philosophy essay. Students received the task details on the learning management system, as follows: "A well-structured sample essay has been broken up into 10 sections and jumbled. The task is to reassemble the sections of the essay into the correct order." Students could download the essay parts to examine the structure and prepare their response. Students had 30 min to complete the online quiz by ordering the paragraphs. Students received an immediate score out of 10 for the number of paragraphs in the correct position. The answer key was uploaded to the learning management system in the following week.

The third activity as described in the unit outline (Fig. 17.2) required students to write an essay plan and submit online for feedback. Derek changed this task because, as explained to the researchers, he had run out of time to design and implement this final part of the task as he had planned. He particularly had not thought through how he would provide feedback on student essay plans in a timely fashion to support students with the subsequent essay task. As a result, the task was changed to students completing a mock assessment of sample philosophical essays in tutorial class, which had been assigned different grades. This change to the final component of the task was communicated to students via the LMS and in face-to-face tutorials. Students received new task details on the LMS, as presented in Fig. 17.3. Students were awarded a participation mark for completing this activity.

The assessment description, as it appeared in the unit outline, described a series of transactional tasks that involved comprehension of instructional materials. The assessment as implemented differed significantly from what Derek originally conceived due to a number of challenges experienced in the design of the online


Essay writing guide – part 3

The task will be done in tutorials in week 7. The task is worth 5% - you just have to complete it to get the 5%. Students who miss their week 7 tutorial, with academic consideration, will be able to submit the task online in week 8.

Below is the essay marking rubric that tutors use when marking essays in this subject. There are 3 sample essays all on the same topic. In your tutorial you will be given one of these essays, and you have to mark it, applying the marking criteria. After this exercise tutors will discuss how the essay is marked. The point of the exercise is to help you look at an essay the way a marker looks at an essay.

Fig. 17.3 Written description of the final part of the assessment task in the subject LMS

assessment. Specifically, he lacked the time needed to re-design the task and the digital skills and knowledge to implement his ideas fully. Nor did he seek help from the central support services available and, ultimately, he had to compromise on his design ideas. The result was that the task was much less open-ended than he had envisaged.

17.4 Student Perspectives of the Task

In this section we consider how the students in this case study engaged with the online assessment task. We know that students often undertake online learning tasks in ways that differ from the teacher's intentions, leading to poor learning processes and outcomes (e.g., Kennedy and Judd 2007; Waycott et al. 2012). Further, task interpretation and self-regulation of learning are key to student success in open-ended online assessments. In such assessments, students are required to interpret tasks and plan their engagement, and there are potentially fewer opportunities for teachers to monitor, intervene and address problems with student learning. To do this, we present task interpretation data from the case study of 313 students undertaking the philosophy unit of study. We then draw on qualitative interview data to uncover contextual factors that contributed to students' general difficulty understanding the re-imagined critical contribution task across the assessment period.

17.4.1 Task Interpretation

We applied Hadwin and Winne's (2012) three layers of task interpretation: explicit, implicit and socio-contextual, as a framework to distinguish between students' levels of task interpretation. Table 17.1 outlines Hadwin and Winne's layers of task interpretation.

Table 17.1 Layers of task interpretation (Hadwin and Winne 2012, p. 206)

Layer	Description
Explicit	Explicit task features refer to the overt aspects of the task including those explicitly outlined in the assignment task descriptions and criteria or verbal descriptions provided by teachers. These include features such as assignment genre, length, content, and task procedures.
Implicit	Implicit task features are inferred by the student and influenced by internal and external conditions. These include the assignment task purpose, alignment of the task with other assignments and unit learning outcomes, strategies and timings.
Socio-contextual	Socio-contextual task features are not specifically task related, but rather inferred/learned beliefs or conventions of the discipline or broader program of study in which the assignment is situated, including features such as disciplinary genres and conventions.

Table 17.2 Layers of task interpretation applied to the critical contribution task

Layer	Application to the critical contribution task
Explicit	Include features such as assignment genre, length, content, task procedures. *Form/procedures:* online quiz(zes), podcast
	Content: essay writing, essay structure, essay marking, essay plan
Implicit	Describes the assessment objectives: Preparation for writing an essay or philosophy essay (without indicating distinction between genres)
Socio-contextual	Describes philosophical essays as a distinct academic essay form

The task description in the unit outline (Fig. 17.2) provided students with an explanation of the explicit task features. Yet, the task did have a clear purpose within the philosophy discipline which remained implicit. Table 17.2 applies the layers of task understanding (Hadwin and Winne 2012) to this particular task using the written descriptions and unit coordinators explanation of the task.

Students were asked to "explain" what they were "required to do" for the task in an open-ended questionnaire item prior to commencing and post completion of the task. This data collection strategy was employed to understand how students developed an early interpretation of the task and whether and how that interpretation changed as they completed the task. As the online assessment could be described according to the three layers, students' descriptions too could contain one or more features. So, for example, if a student's task description contained only explicit features, it was assigned only one layer of understanding (explicit), but if it contained both explicit and implicit features it was assigned to both layers.

An overview of the layers of task understanding present in students' pre and post task open-ended questionnaire responses are provided in Table 17.3. Analysis of students' task interpretations revealed that many students did not clearly or accurately articulate features of the task. For this reason, another category of task interpretation was included in the reporting of these findings to represent responses that demonstrated an unclear or inaccurate task interpretation.

Table 17.3 Layers of students' task interpretation

Layer	Pre task (313)	Post task (244)
Unclear/inaccurate	138 (44%)	104 (42%)
Explicit	168 (53%)	117 (47%)
Implicit	19 (6%)	30 (12%)
Socio-contextual	12 (3%)	33 (14%)

After receiving the formal task description, 44% of students were unable to provide a clear or accurate description of the task prior to commencing work on the task. This finding was not surprising at this stage in the semester given the brevity of the task description. However, 42% of students were still unable to provide a clear or accurate task description after having completed the task. This finding indicates that a significant portion of students remained confused about the nature of the task, even after engaging in the prescribed activities.

At an explicit level, most students described details about task features, form and/or content prior to (53%) and post completion (47%) of the task. This is perhaps not unexpected, as the written communication about the task, including the unit outline description and communication on the unit LMS, detailed the explicit features.

Within this context, a much smaller portion of students described implicit aspects, with only 6% of students detailing the assessment objectives prior to completing the task and 12% post completion. While the percentage of students who described implicit aspects doubled across the assessment period, our analysis shows that a significant number of students (88%) did not demonstrate this level of understanding after completion of the task. In plain terms, this means that the majority of students did not correctly describe the task purpose at either point in time across the task period. Again, this finding points to the brevity of the official task description and raises questions about the broader context of the task, for example, the ways that the task was communicated to students in class and online and the change in task design.

At the socio-contextual level, only 3% of students connected the purpose of the task to appreciating the distinction between a traditional academic essay and a philosophical essay prior to completing the task. After completing the task, the percentage of students who described such an understanding increased to 14%. While this increase is heartening, the low percentage suggests that only a minority of students had been able to place the requirements of this assessment in the wider context of assessment in their discipline or in higher education more generally. This is a significant finding because a key motivator for Derek's design of the task was to support students to understand that the philosophy essay was distinct from the academic essays his first-year students were used to. Yet, a deep understanding of a task requires more than the explicit descriptions or instructions it includes students' sense-making of implicit and contextual cues as well (Hadwin 2006).

As a cohort, the philosophy students' task understanding was mostly unclear/inaccurate or focused on only explicit features of the task across the assessment

period. While a small portion of students demonstrated some growth in their understanding of implicit and socio-contextual features of the task over time, the majority of students did not describe these features. This suggests that many had difficulty in developing a complete and accurate task understanding. To better understand students' difficulty with task interpretation we present findings from 12 students who were interviewed below.

17.4.2 Student Interviews

In this section, we share students' explanations of their approaches to the assessment task, framed by the four recursive phases of self-regulation from Winne and Hadwin (1998): (1) developing accurate and complete task perceptions; (2) setting high quality goals during planning; (3) adopting and adapting strategies to achieve goals; and (4) continually evaluating and adapting studying during and across tasks.

17.4.2.1 Developing Accurate and Complete Task Perceptions

Most students' task interpretations focused on an explicit understanding of the task. Students described the task in terms of its discrete components and the process they had to undertake to complete the task. They also described value associated with the task when describing their interpretations, referring to a general perception of it being an easy task. For example, Harry provided an explicit description of the task along with his perception, which he later reconsidered:

> I think it's just there's three online things; we have to go on there and listen to a recording and then answer some questions afterwards. I did the first one the other day. It was pretty easy. (Harry, initial interview)

> I think when it was first explained, I did think it was going to be a lot easier than it was and I think I definitely under-estimated the amount of time I'd need to prepare for it. (Harry, post task interview)

Much like the larger cohort of students, few of these students described the implicit or socio-contextual layers of understanding as described by the unit coordinator. Though, this is not to say they did not have their own implicit understandings that differed from the intentions of the unit coordinator. Students' descriptions suggest that they were almost dismissive of the task form, that is "it's just three online things", perceiving them as almost inconsequential. Along with low weighting of allocated marks and the explicit instructions, it is plausible that students understood this as a simple gateway task to the subsequent 'real' assessment, the essay.

Nine of the twelve students described experiencing challenges understanding the task associated with a range of factors including past experiences, communication about the task, and task design. Two students made reference to their previous experiences with essay writing in high school when discussing challenges associated

with their task understanding. For these students their lack of awareness of implicit and socio-contextual layers of the task resulted in feelings of overestimation of their ability. As Saxon explained:

> Yeah, because like in high school I did a lot of essay-based subjects like Ancient History, Modern History, Advanced English, all that stuff, so I knew how to write essays like well and like since the first part of the quiz was how to write an essay then I just assumed "Oh yeah, I'll be fine" then, you know, got two questions wrong or something, I was like "Ooh, fudge" so maybe I'm not as great as I think I was. (Saxon, post task interview)

If Saxon had developed a more holistic task understanding that included the purpose of the task, which was to learn the conventions of a philosophical essay as distinctly different to a general academic essay, rather than rely on the explicit task understanding and his prior experiences, he may not have experienced such a challenge. These findings suggest that in cases where the purpose of the task is not clear, students may draw on their past experiences of similar tasks, which in this case did not align with Derek's design intentions.

Four students described limited communication from teaching staff about changes in the task as a key challenge in their task understanding. Students were unsure about changes to explicit task features (e.g., task criteria, due dates, and when to access tasks on the LMS), which affected their ability to engage with the task and, in two cases, students missed the opportunity to complete one of the task components.

Seven students described confusion directly associated with the task design of activity two, about how to arrange the essay paragraphs into the correct order. The design of the task added to this confusion as the LMS assigned marks to the position of each essay paragraph in a pre-determined placeholder rather than acknowledging pieces in the correct order. Thus, if a student had one piece in the wrong position and the following five pieces were in the correct order, they would all be marked as incorrect. In this activity, the task design impacted on students' confidence to complete the task and many students discussed the challenge of completing the activity rather than the structure of the essay, as Tom explained:

> It was probably the essay puzzle task. That one was really difficult. It's more because there's no real right or wrong; for an essay it's more just like some are better than others so I was really confused just how they were assessing that … it was really confusing. (Tom, post task interview)

Comments like this one indicate that task design contributed to students feeling less confident about writing a philosophy essay after completing a task that was intended to support their understanding.

17.4.2.2 Setting High Quality Goals During Planning

In the case of the assessment task an optimal goal for students would have been to understand the features and structure of a philosophy essay as distinct from more general forms of essay writing and apply this knowledge in their own essay (the subsequent assessment task). Instead, many students' goals were focused on simply

completing the task or achieving a certain grade. The quote from Harry above demonstrates how a poor initial interpretation can lead to a low-quality goal (just listen and answer some questions), and his further reflection shows how this in turn led to him being under prepared. Given that many students had difficulty in understanding the task it is unlikely they could have set high quality goals. Previous research demonstrates that students with an accurate and more complete understanding of a task are more likely to achieve a higher learning outcome as they can set high quality goals and set standards that help them to monitor and evaluate their progress (Beckman et al. 2019; Miller 2003; Oshige 2009).

17.4.2.3 Adopting and Adapting Strategies to Achieve Goals

The students described a range of strategies for completing each component of the task. Given students' focus on the mechanics of the task when describing their task understanding, most strategies described by the students were specific to each component of the task rather than more broadly associated with understanding a philosophical essay. This is perhaps not surprising given the transactional nature of the tasks, which raises important questions about the suitability of the task design for achieving the intended outcomes. The first online activity featured an audio recording of the unit coordinator explaining the key features and conventions of an essay. Students could listen to this recording as many times as they wished and then complete 10 multiple choice comprehension questions. Strategies discussed by students when completing this activity included: note taking while listening to the audio, answering questions while reviewing notes, and taking time to complete the quiz to achieve the highest score. Collectively, students described a rote learning strategy that focused on answering the questions rather than developing a deep understanding of the philosophical essay as a genre, which was the intended outcome of the reimagined assessment design. In addition, all case students discussed the ease with which they completed the quiz. For example:

> There's an audio file and as I listened to the audio file I just typed main points and all like the main things that the teacher was saying and I typed that all up and studied that and then took the quiz. (Ralph, post task interview)

The second online activity required students to arrange the paragraphs of a sample philosophy essay in the correct order. Students initially perceived this activity as "easy" but had difficulty in clearly explaining how they went about completing it. Students described a number of strategies including: changing approach, taking more time, working with their peers who were equally confused, locating connecting sentences, reviewing tutorial content, and referring to a worked example. Of all the strategies discussed none of the students made reference to Part A of the task, suggesting a disconnection in the way they viewed the component activities. For example:

> Yeah, that was a bit difficult. I don't really know how I prepared for that. I think I just kind of went into it, like I just kind of found like the introduction, the reference list and then tried

> to fit everything in between but it was a bit difficult. I probably should have prepared a bit more for that. There wasn't anything that I remember that was given to help us prepare. (Gabby, post task interview)

The third activity involved students in 'marking' sample philosophical essays. Students described the way the task had evolved, but did not express any concerns about changes. They reported the class discussion associated with the marking exercise as instrumental in their understanding of the purpose of the activity. This suggests that this activity supported case students to engage in a collective reflective process. For example:

> We sat in groups and discussed the essays and what mark to give them. Yeah, that was fine. That was really good in helping us understand what they're expecting. I didn't really have to prepare for that I guess. It was more just like collaborating with other students and their thoughts and everything. Yeah, I think that was really the best task of them to really understand. (Gabby, post task interview)

17.4.2.4 Continually Evaluating and Adapting Studying During and Across Tasks

Effective learning is as much about responding to challenges, shortcomings and failures as it is about choosing the right strategy (Hadwin 2006). Components of the task that were valuable in supporting students' self-regulation included face-to-face class discussions. Class discussion was particularly important in supporting students to develop a deeper task understanding and, in turn, select more appropriate strategies. Students discussed using scores from each component of the task as an evaluation tool. They described a sense of confidence if they scored well and, conversely, confusion and low confidence if they scored poorly. When students scored poorly they did not seek feedback about their performance on the task. Several explained their lack of motivation to seek feedback by the low value assigned to the task, with each component worth 5%. The second activity included feedback uploaded to the unit's LMS space, but only one of the 12 students had accessed this resource. Conversely, all students discussed seeking help from their peers. However, discussing the task with peers served to reinforce students' generally low motivation to access support, as there was a general sense that because others found the task difficult, the confusion was acceptable. Research has shown that the goals students set for themselves are based on the task interpretations together with a range of contextual factors (Butler and Cartier 2004). In this case, students' behaviour was associated with a collective low task value and motivation to achieve, resulting in 'giving up' during the monitoring and control phases of self-regulation.

Overall, the students' capacity to self-regulate their learning across the task period was inhibited by their limited task understanding. Many students who described inaccurate and/or explicit task features adopted surface strategies (like learning by rote), collaborated with peers who were equally unclear about task requirements, or completed each component discretely. This was most evident when students described strategies such as spending more time on the task when faced

with a difficulty, which was unlikely to be effective, rather than trying to develop a better understanding of what they were being asked to do. In addition, students did not seek feedback about poor task performance, rather they "chatted" to their peers and accepted the task as "difficult". The most valuable experience for students in refining task understanding occurred during the third component of the task. This component was an in-class activity that provided students with the chance to analyse a philosophical essay and receive immediate feedback about their judgements during a class discussion. The structure of the task provided students with the opportunity to monitor their task understanding and refine standards within and across tasks, which would inform their monitoring in the subsequent assessment task for this unit, composing their own philosophy essay.

This analysis of our students' experiences raises several important questions about the design of online assessment tasks including: how can we promote understanding of the explicit, implicit and socio-contextual aspects of a task to ensure students better appreciate the broader purpose of the assessment design; how can we embed meaningful, connected and timely feedback in online spaces; and perhaps most importantly, what types of assessment tasks are most effective for online delivery?

17.5 Factors Shaping Task Engagement

The study enabled us to identify a range of personal and contextual factors that shaped students' engagement with the online task, including factors such as knowledge of the task, perceptions of and prior experience with online learning and assessments, study strategies, and motivation. Overall, we found that student interpretations of the assessment activities were largely incomplete and focused mainly on explicit features (e.g., word count or form), rather than seeking out the purpose or underlying logic of the task. This is perhaps not surprising for two key reasons: (1) the reimagined task design and its enactment were not well aligned with the intended purpose; and (2) the task purpose was not clearly communicated to students in any official documentation nor were they clarified sufficiently by the teaching team.

In this case study, many students underestimated the difficulty of the task. This may have been due to the form it took. For example, the quiz and ordering tasks may have seemed like inherently simple forms, but in fact the content of the task was more challenging. This was problematic as it limited students' capacity to effectively engage in the assessment. Even when confused, the students did not tend to seek help from their teachers, turning instead to their peers, whose own confusion then compounded the problem.

It is important to also acknowledge that the ambitions of the unit coordinator, who designed the task, outstripped his capacity to realise an effective assessment design. His plans for an open-ended online task to support students to develop their understandings of how to write a philosophy essay were undermined by his own under-estimation of the time and effort to create successful online activities.

Ultimately, it was the final, in-class activity that provided a useful contrast. This was the most effective activity from a student perspective, even though it could have been the most ambiguous considering the last minute changes. The students' comments suggest that the more interactive nature of the task, with discussion and feedback, were critical to advancing their understanding.

The online component was designed to support student learning without encroaching on valuable class time and to better allocate staff time to grading student work. Despite the advantages the online learning environment afforded, this case study highlighted design challenges of the online assessment. A key theme in this case was the explanation and communication of the online assessment to the students. For students, interpretation of assessments tasks involves more than understanding explicit instructions, but also the intended task purpose in the context of the broader disciplinary ways of knowing, all of which may be implied or expected of students (Hadwin and Winne 2012). In this case, the brevity of the formal written description, the varied nature of communication about assessment across the teaching team, the methods of communication with the students (including the formal description), and the additional information provided online and in class by the teachers shaped students' understanding and, ultimately, their engagement in the online assessment and learning outcomes.

The findings support previous suggestions that task design and communication may be particularly significant for online assignment tasks that are complex and involve task structures and components that may be less familiar or less prescribed than traditional assignments (see Oshige 2009). This may be so for both students and teachers, presenting particular new challenges in the design, communication and interpretations of online assignment tasks that warrant further detailed investigation. Thus, while online assessments can provide new opportunities to redesign traditional assessments, they also require careful design and communication to support students.

17.6 Conclusion

In closing, we come back to the theme of this edited collection, the reimagining of assessment in a digital world. The case was ultimately a modest attempt to leverage digital technologies to create a flexible independent learning experience, but one which challenged students' capacity to self-regulate their learning because most did not develop a clear understanding of what the task required of them. Derek was an experienced teacher with a sound understanding of assessment strategies who found himself ill-equipped to realise his own ambitions for a new design. The outcome was a poorly designed task, which did not take advantage of digital technology, that was confusing to students and not sufficiently clarified as it was being implemented. This case of realisation falling short of the reimagining is likely to be a common experience, with implications for students, academics and institutions. It reminds us

that the 'state of actual' is often a long way from the 'state of the art' in technology integration, and there is still much we do not understand about how best to close that gap.

References

Beckman, K., Apps, T., Bennett, S., Dalgarno, B., Kennedy, G., & Lockyer, L. (2019). Self-regulation in open-ended online assignment tasks: The importance of initial task interpretation and goal setting. *Studies in Higher Education*, 1–15. https://doi.org/10.1080/03075079.2019.1654450.

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology, 48*, 672–682. https://doi.org/10.1111/bjet.12439.

Biggs, J. (2001). Enhancing learning: A matter of style or approach. In *Perspectives on thinking, learning, and cognitive styles* (pp. 73–102). London and New York: Routledge.

Butler, D., & Cartier, S. (2004). Promoting effective task interpretation as an important work habit: A key to successful teaching and learning. *The Teachers College Record, 106*(9), 1729–1758.

Hadwin, A. (2006, August). Do your students really understand your assignment? *LTC Currents, 2*(3), 1–9.

Hadwin, A., & Winne, P. (2012). Promoting learning skills in undergraduate students. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning* (pp. 201–227). New York: Cambridge University Press.

Kennedy, G. E., & Judd, T. S. (2007). Expectations and reality: Evaluating patterns of learning behaviour using audit trails. *Computers & Education, 49*(3), 840–855. https://doi.org/10.1016/j.compedu.2005.11.023.

Miller, M. (2003). *Predicting university students' performance of a complex task: Does task understanding moderate the influence of self-efficacy?* (Master of Arts thesis). Retrieved from https://dspace.library.uvic.ca

Oshige, M. (2009). *Exploring task understanding in self-regulated learning: Task understanding as a predictor of academic success in undergraduate students* (Master of Arts thesis). Retrieved from https://dspace.library.uvic.ca

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407.

Selwyn, N. (2010). Looking beyond learning: Notes towards the critical study of educational technology. *Journal of Computer Assisted Learning, 26*(1), 65–73. https://doi.org/10.1111/j.1365-2729.2009.00338.x.

Waycott, J., Dalgarno, B., Kennedy, G., & Bishop, A. (2012). Making science real: Photo-sharing in biology and chemistry. *Research in Learning Technology, 20*(2), 16151. https://doi.org/10.3402/rlt.v20i0.16151.

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah: Lawrence Erlbaum.

Tiffani Apps is a Lecturer in the School of Education at the University of Wollongong. Her research interests focus on educational technology, digital literacy and pedagogy. Tiffani's current research focuses on university students' self regulated learning online and children and young people's digital literacy practices across early years, school and higher education settings.

Karley Beckman is a Lecturer in the School of Education at the University of Wollongong. Her research focuses on technology in education spanning across primary, secondary and higher education settings with a particular interest in how people engage with technology across their everyday lives and educational contexts.

Sue Bennett is a Professor in the University of Wollongong's School of Education. Sue's research investigates how students and teachers engage with technology in their everyday lives and in educational settings.

Chapter 18
Reimagining Assessment Through Play: A Case Study of MetaRubric

Yoon Jeon Kim and Louisa Rosenheck

Abstract Younger generations of learners want to have a sense of ownership and play active roles in their own learning processes. While many innovative pedagogical approaches in higher education aim to provide learning experiences that empower students as partners in their educational experience, assessment can fail to offer that ownership. In this chapter, we describe how bringing play into assessment could transform both educators' and students' roles in assessment processes, and how a playful take on assessment could introduce social, iterative, and democratised assessment processes. We further illustrate how play could be integrated with assessment using the case of MetaRubric, a playful assessment design tool that educators can play to learn about the features of good rubrics, or have students play to co-create rubrics in a playful way. We conclude the chapter by discussing the potential benefits of playful assessment for assessment of and with younger generations.

18.1 Introduction

Advances in technology are replacing routine cognitive tasks and manual labor and jobs increasingly require creative thinking and complex communication skills (Autor et al. 2003; Levy and Murnane 2004). In many countries, including the United States, there is a growing societal concern as to how the education system can prepare younger generations for the twenty-first century global economy (Bybee and Fuchs 2006; Florida 2002; Freidman 2005). The consensus is that society must prepare learners with knowledge and skills that are fundamentally different from those needed in previous centuries. It is clear that the competitive global economy will require the skills that underlie innovation and entrepreneurship, such as creativity, problem-solving, leadership, communication and collaboration, and social and emotional skills (OECD 2014).

Y. J. Kim (✉) · L. Rosenheck
MIT Playful Journey Lab, Cambridge, MA, USA
e-mail: yjk7@mit.edu; louisa@mit.edu

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_18

These societal changes have been driving pedagogical innovations in higher education, where educators shift teaching from instructor- and content-centric methods of *teaching* to authentic, student-centered, and problem-based approaches of *learning*, in order to adequately prepare graduates for future jobs (Christensen and Eyring 2011). For example, "flipped" classroom approaches are increasingly common. In a flipped classroom, students study content outside of the classroom and then come together during class time to ask the instructor questions or to apply what they've learned as a team on projects or authentic problems (Bishop and Verleger 2013). Similarly, project-based learning (Helle et al. 2006) is popular, where students participate in an extensive project for 6–8 weeks, working with corporate or non-governmental organization (NGO) partners on issues that reflect real problems of practice, while the instructor facilitates and supports learning by providing content and strategies on demand (Jiusto and DiBiasio 2006). Project-based learning is an example of a more student-centered, open-ended approach to learning in higher education meant to support lifelong learner skills (e.g., collaboration and meta-cognition).

These approaches, to varying degrees, view students as partners by empowering them to make decisions and actively contribute to their own learning (Bryson 2014). Simultaneously, the digital generation, who grew up using social media where participation and production are highly valued (Pearson Report 2018), might be reluctant to take a passive role in their own learning process (Zimmer 2017). Therefore, student empowerment and the notion of the student as a partner in learning needs to be seriously considered. However, even with the aforementioned student-centered approaches, truly collaborative and empowering assessment opportunities are very limited. Unless we change how students are assessed, we will not see significant changes in the system as a whole. For this reason, assessment is a key leverage point that can have a significant impact on how students learn.

Educators need ways to productively engage students as partners in the process of assessment—from creation to implementation—to fully honour students' agency in ways that are aligned with pedagogical approaches that position students as active partners of their learning process. Assessment in higher education, however, has not caught up to pedagogical trends, and fails to assess students' learning in self-regulated and collaborative ways using authentic and engaging problems. We therefore need new types of assessment and tools in higher education (López-Pastor and Sicilia-Camacho 2017) that can measure, and more importantly support, the skills students require to become reflective practitioners in their technical domains and beyond, while also inviting them to be co-producers of their learning and assessment.

In this chapter, we present an assessment design and learning tool called MetaRubric to demonstrate how educators can use a collaborative and playful process to engage peers and students in the creation of rubrics or assessment criteria to capture what various stakeholders—not just instructors—value in open-ended student work. We first review the existing literature of playful learning to describe what factors might contribute to playful assessment experiences and how embodying the principles of playful learning for assessment could be beneficial. We describe our

research investigating how innovative educators design rubrics. We then introduce MetaRubric, which is designed to guide educators and students in the co-creation of rubrics and authentically demonstrate what makes rubrics good. We then discuss how the design principles of MetaRubric could be incorporated in digital learning environments to foster playful interactions between learners that can lead to more empowered and democratic ways of assessing learning. Finally, we discuss how tools like MetaRubric could be used to help educators to create their own authentic assessment tools that empower learners as co-creators in the digital world.

18.2 The Potential for Play in Assessment

We assert that assessment needs to embody the principles of *playful learning* as a way to change how we approach assessment in both digital and non-digital learning environments. The addition of playfulness could lead to better assessment processes that engage learners and educators in an empowered, transparent, and collaborative way. In particular, we see an opportunity to integrate playfulness in processes of co-design.

The benefits of playful pedagogies have been widely documented at all levels of education, from early childhood and K-12 education to adult and higher education (e.g., Lillard et al. 2013). Defining fun, play, or the conditions that lead to playful experiences is difficult (Juul 2005), and not surprisingly, there is no consensual theory or terminology of playfulness (Deterding et al. 2011). The goal of our work is neither to define what playfulness is nor to establish a taxonomy of playfulness, but rather to apply the design principles of playful learning and processes of game design and playful user experience design (Deterding et al. 2011) for the purpose of transforming how assessment is conducted in education.

When we think of "playfulness," we are describing a fun experience that is not necessarily easy. Papert (1980) describes "hard fun", where hard or challenging experiences are still engaging enough to be fun, and it is our aim to create assessments that embody this concept. Similarly, learners can be "pleasantly frustrated" when they persist through challenging problems in well-designed games (Gee 2003), unlike the unpleasantly frustrating and anxious experiences associated with most assessments. We believe playfulness in assessment can mean that assessment provides an opportunity or space for learners to explore and experiment within the assessment activity (Resnick 2014), both individually and socially. Therefore, we can achieve the goal of playfulness in assessment activities by thoughtfully designing activities so they are "flexible enough for players (learners or teachers) to inhabit and explore through meaningful play" (Salen and Zimmerman 2004, p. 165). Similarly, a playful approach in assessment should allow learners or educators to have more freedom related to how much effort they choose to expend and how often they fail and try again (Osterweil and Klopfer 2011).

An opportunity for bringing *play* into assessments in higher education is presented by the increased use of co-creation of rubrics or criteria, whereby educators and students collaboratively develop the assessment tools used in class (Fraile et al. 2017). The co-creation of assessment criteria or rubrics can help students to make meaning of the concepts that underlie the assessment, and it also shifts the ownership of the assessment from purely teacher driven to one that is shared. It is generally a participatory process, based on discussion and interaction. However, we contend that this co-creation process can also be *playful*. The core principle of our work is that creating and implementing assessments should be playful for both students and teachers. If learning is fun, active, and authentic, then there is no reason the experience should suddenly change for rubrics or other assessment products, which are traditionally seen as intimidating or dull. A well-designed rubric should be seen as a tool to help students learn and progress, rather than as a threat or ultimatum. By using play, we can help students in higher education engage with rubrics and assessment more fully as they create pedagogically-sound and well-balanced rubrics.

We present our work on playful assessment in schools, drawing out useful implications for assessment in higher education, particularly in a digital world where learning occurs across both digital and non-digital enviroments. We start by describing the development of the MetaRubric.

18.3 The Origin of MetaRubric: Investigating How Innovative Educators Design Rubrics

Our work started with the goal of creating a playful tool that can guide educators through the process of (a) creating a good rubric for open-ended student work, (b) learning about the qualities of good rubrics from the process, and (c) understanding innovative assessment design practices. To inform the design of this tool, we aimed to concretely understand what practitioners do related to assessing student open-ended work. The first step was a literature review with the goal of finding out what classroom assessment used to assess students' learning with more open-ended learning approaches. From the literature review, we concluded that rubrics are the "default" practice when educators employ learning activities that are intended to support skills and dispositions beyond just content knowledge. Yet little was found detailing on-the-ground practices with respect to what educators actually do with rubrics, the kinds of processes they use, and what motivates them to go beyond simply using overly rigidly defined rubrics.

Being dissatisfied with the literature review, we then followed a two-step approach to identify innovative educators who could provide more insights about rubric use. The goal was not finding a representative population, but rather finding specific cases that could provide emergent understanding for the theory (Charmaz 2014). Using this approach, we first identified a few teachers from the literature

Table 18.1 Teachers interviewed to inform the design of the tool

Teacher ID	Teacher's role	Brief description
1	Middle school STEAM teacher	Teacher of digital shop class, a blended makerspace course, in a public middle school
2	Senior university lecturer	Lecturer of education and public policy at a Scottish university
3	High school English teacher	Performance assessment teacher trainer, and current teacher at a charter school focused on college and workplace readiness
4	Middle school STEAM lab and makerspace facilitator	Curriculum developer and teacher of programming, robotics; runs a drop-in makerspace in the school
5	Formative assessment coach	Middle school special ed. teacher and formative assessment coach
6	Elementary school team teacher	1st/2nd grade combined age team teacher at a private school with a project-based, inquiry learning curriculum
7	Elementary school team teacher	1st/2nd grade combined age team teacher at a private school with a project-based, inquiry learning curriculum
8	High school science and engineering teacher	Teacher of physics, biology, robotics, and engineering, including an engineering, manufacturing, and design course in a traditional public high school
9	Fifth grade teacher	Teacher in a high-achieving charter school network focused on teacher development
10	Director of professional development	Teacher educator supporting a game-based integrated learning model at an urban school

review (e.g., blog posts, teacher professional development portals) that were known to experiment with interesting assessment techniques and were known for their interest and passion for classroom assessment (e.g., the teacher published a blog post about his or her experiment of assessing learning in maker-centered curriculum). Second, we conducted snowball sampling by asking the initially identified teachers to suggest other teachers who have interesting ideas about the use of rubrics. Through these methods, the team recruited 10 teachers, including one working in higher education (Table 18.1).

Ten educators participated in 30-minute semi-structured interviews. Eight sessions were phone interviews recorded using teleconferencing software, one session of which included a pair of team teachers that were interviewed together. One educator opted to participate by answering the interview questions over email. The interviews were then analyzed for common themes and contrasting approaches based on notes and recordings from the sessions to answer the guiding question of *"what makes good rubrics?"*

The research team conducted a thematic analysis using the transcriptions of the interviews (Braun and Clarke 2006). The analysis yielded three broad themes that these educators apply in their assessment work by going beyond simply listing specific features of a task, which are (a) intentionally breaking the common perceptions

of what constitute good rubrics, (b) empowering students, both in the process and implementation of rubrics, and (c) incorporating social aspects for rubric creation and use processes. These insights are just as pertinent to rubric development in higher education. Better designed, and just as importantly, co-designed, rubrics can develop the graduate skills like creativity, problem-solving, communication and collaboration that are in such high demand in our digital world. We explore the themes raised by our interviewed educators, highlighting issues relevant to higher education before describing our response to "scale up" these innovative practices in the form of the MetaRubric.

While the literature overwhelmingly looked to rubrics as the main way to assess more performance-based and open-ended student work (e.g., Jonsson and Svingby 2007), our interviewed educators expressed their concerns about poorly designed rubrics. They shared how they move their assessments away from rigid and overly prescriptive rubrics that limit students' learning and agency. While the interviewed educators don't necessarily disagree with the concept of a rubric, they take extra steps that go beyond standard rubric creation to ensure that their rubrics are flexible enough for students to communicate the skills and values they demonstrated through the activity. These educators want students to have the freedom to decide for themselves what they can do to show their mastery, without the requirements of a known rubric hanging over their heads as the primary goal. For example, Teacher 8 teaches in a school that requires grades and therefore must map a point system onto some kind of rubric. He therefore does use a rubric, but keeps it as flexible as possible and incorporates performance criteria that motivate his students to explore and learn. For example, his rubric includes "cool factor", which could be very subjective. By intentionally choosing criteria that can be provocative and difficult to agree upon, he challenges his students to further explore different aspects of their work instead of encouraging the features of the project to conform through completely objective standards.

This teacher also emphasized the importance of the rubric as a tool that help students with their own learning processes rather than meeting all the requirements that the teacher specified. He says, "If I did not have oversight from administration I would not use rubrics or rubrics with point systems at all. Administration wants a clear path to a grade even if that path is not the best for learning. That is why I get the students involved in the process of designing and writing their own grading rubrics."

The interviewed educators emphasized student autonomy in the assessment process as one of their main goals. They find that when students have freedom to pursue their interests and set their own goals, they achieve far beyond expectations. Because of this, many instructors are constantly looking for ways to make students feel heard and to take ownership in the assessment process. Some of the key ways teachers in this study achieve the goal of empowering students include inviting students to engage in self-assessment, portfolio defenses, and co-creation of rubrics or learning goals.

To return to Teacher 8, this teacher asks his students to self-assess their work before a teacher conference in which student and teacher compare their

assessments. His rubric has two columns to assign numerical value—one done by the student and the other by the teacher. In this form, students have the opportunity to discuss what their thought process was and why they assigned a certain value for performance criteria. While Teacher 8 takes his role in assessing student work seriously, he says he often gives students the points they advocate for, mainly because using evidence to construct an argument, as well as having the confidence to question a teacher, are such important skills for them to be practicing. However, he also finds that, especially in his engineering class, "It's really surprising how much I've seen kids not really care about grades, they care about just building and designing. I'm the one defending the kid's grade, saying you don't understand how impressive this is."

The portfolio defenses in Teacher 1's school are a good example of empowering students in demonstrating understanding, not only in a formative stage but also as summative assessment. During the defence presentation, each student is judged by a panel made up of their own teacher, another teacher, and a peer, all of whom must first go through training to reliably evaluate the presentation according to the rubric. By including students in the portfolio review process, students learn the importance of taking each other seriously and that a peer's opinions are as valuable as a superior's. But the key piece of this assessment method is that the work the student has produced has already been evaluated. The portfolio defense rubric targets the student's ability to describe their process, explain the significance of their work, and apply their skills to new contexts. This metacognitive element means that no one but the student can know what must be included in the portfolio and at what level, empowering students to be responsible for showcasing their own competencies.

Lastly, co-creation of rubrics or learning goals emerged as a common practice that gives students real ownership of their learning as well as providing a social dimension to learning. So much school work is done simply because the teacher assigns it, without students having any input into what is considered important in the classroom. Teacher 5 starts many assignments by having her students discuss what components should go into the rubric and how they should be prioritized. She feels that this is an invaluable exercise for students on multiple levels. It gives them the perspective of weighing everything that goes into a work and determining what makes it good. In addition, once the class has come to a consensus, it provides an artifact to refer to throughout the project that reminds students why they are working so hard: because they themselves felt these criteria were important. She says, "I use these methods to help students to become more active in, and to take more responsibility for their learning. I find these methods valuable in that they can encourage students to be motivated and learn at a deep level. They also help them in understanding academic standards, and assessment and feedback literacies." Interestingly, Teacher 5 and many other teachers reported that although it may take a leap of faith to give up some control over what the assessment criteria will be, in the end the rubrics the students come up with never stray far from what the teacher would have used without student input. This may well be because years of traditional schooling have influenced students' views of what is important to learn. Regardless, it is an observation that can encourage new teachers—even those in

higher education—to put more control in the hands of their students and enable students to feel ownership of their own learning process.

The ideas raised by the interviewed educators were incredibly influential in helping us design a playful assessment tool. In its original conception, MetaRubric was a tool developed for K-12 educators, and we will describe it as such before we move on to outlining how it might be used in higher education.

18.3.1 MetaRubric: A Playful Assessment Design Tool for Better Rubric Making

Based on the interviews and literature review, we created a design and thinking tool that can guide educators' and students' creation and use of rubrics beyond what assessment textbooks suggest as good practices. MetaRubric is a playful assessment design and learning tool where a group of 3–5 players follows the instructions on a sequence of cards (as shown in Fig. 18.1) to walk themselves through the experience of creating a good rubric. To start, the group chooses a movie that they have all seen and like. Then they are each asked to draw a movie poster for that movie—these drawings serve as the basis for the subsequent rubric construction, and set players up to have both the experience of assessing with a rubric and of being

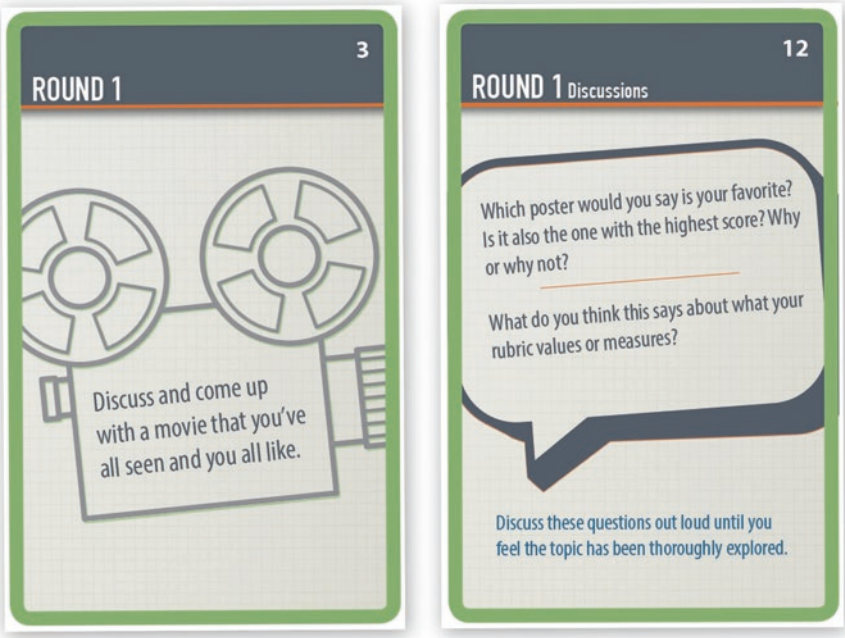


Fig. 18.1 Sample cards from the MetaRubric game

assessed by one. The next step is for players to individually write a list of criteria they think a good movie poster should meet, which starts the process of identifying what elements and skills are most important in a piece of work. From there, they are asked to share and discuss their criteria in order to come up with one group rubric with a set of criteria they all agree on. This activity involves a lot of group discussion, through which many of the key parts of rubric creation emerge organically.

The group typically discovers that some people have thought of concrete criteria such as "includes image of main character" or "has title of the movie", while others have more subjective criteria like "evokes strong emotion" or "depicts tragedy". Players have to thoroughly explain why they think each criterion is important, and moreover, what exactly that criterion means and how it would be evaluated. Once they have a group rubric, players use that rubric to score each other's movie poster drawings and compare scores to identify the round 1 winner. Here the group will usually run into challenges related to assigning point values, something left intentionally vague, as how point values are assigned is another key aspect to consider when designing a rubric. Allowing players to discover that each has approached scoring differently allows for fruitful discussions to take place. At the end of the round, the group is given discussion prompts to reflect on whether scoring was reliable or fair, and how evaluators may have differed. For example, one prompt asks if the most highly scored poster was one of the group's favourites. Often times, players discover through this comparison that there is a mismatch between the scores and what they value.

Next, players move on to round 2, which is intended to help players to understand what the qualities of good rubrics are. They essentially go through the same process as in round 1, but instead of evaluating movie posters, they are evaluating their group rubric with a new rubric: a "meta" rubric. They start by individually writing down a list of criteria that a good rubric should meet. Then, once again they share and discuss their ideas to come up with one group "metarubric." These often contain items such as "has a balance of concrete and creative criteria," "items are specific and distinct," and "rubric is easy to read and easy to use." This part of the activity really helps players reflect on what they think makes a rubric work well and what they even mean by a "good" rubric. As before, players then use the group metarubric, this time to evaluate the individual rubrics they wrote in round 1, and again they add up the scores to compare both the rubrics' results as well as the assessment methods used. This leads into a discussion of possible rubric formats and how they might use rubrics themselves in the future. By the end of the experience, players have usually realized that the rubric they created early on was not particularly good. But this is by design, as they undoubtedly will have grappled with authentic problems in the process, and begun building their understanding of what rubrics can do and how they can do it.

Fig. 18.2 Educators playing MetaRubric

18.3.2 What Educators Understand from Playing MetaRubric

The entire process of playing MetaRubric is fun and engaging. At the workshops we have hosted with educators, we witnessed a lot of laughter and engaging conversations (Fig. 18.2). But at the end of the session, what kinds of new insights and heuristics have they obtained from playing MetaRubric?

To answer this question, we conducted a survey with 56 pre- and in-service teachers who used MetaRubric. In the survey, we asked open-ended questions regarding insights teachers gained from playing MetaRubric. We identified three broad themes from the responses.

Firstly, participants came to appreciate the assessment design process that was embodied in MetaRubric, where you create an actual artifact you want to assess, collaboratively come up with criteria through discussion, use the agreed upon criteria, identify discrepancies in the scores, and iterate. The teachers commented, "The MetaRubric helped clarify a framework for approaching assessment"; "I realized that it is important to evaluate rubrics before just using them;" and "[I realized] how helpful it is to talk with others about how we design helpful and useful rubric items." Secondly, educators recognized how the same process can be used to engage students in the co-creation of assessment and rubrics, and that this process can foster student autonomy and agency. For example, a teacher said, "I had this idea of inviting students to work on this process of making rubrics. I thought that would be a great approach to help students internalize and build a connection to the content." Many similarly responded with comments like, "Students need a role in creating assessment criteria and student agency can be taken into account. The thinking process of working through the rubric criteria was powerful. I could see this being valuable in the classroom with students as well". This insight came out of teachers' reflection of their own experience as they "take on the role" of both tester and tested in MetaRubric when their work is scored by others. Thirdly, teachers also mentioned that they realized the importance of deeply understanding the learning goals that they intend to measure and clearly communicate them with students.

Finally, educators also reported how MetaRubric provided them an opportunity to reflect on common assumptions about what a standard rubric should be like and how those qualities can limit what you can achieve with rubrics. For example, many described how they noticed the tension between subjectivity and objectivity in assessment after playing MetaRubric, e.g., "I realized that people can evaluate the same things using the same criteria in different ways, and how this could make a rubric subjective based on who's evaluating and their interpretation." They also pointed out the danger of focusing on one dimension as potentially being limiting, and how hard it is to make a well-balanced rubric, e.g., **"**It's hard to make an authentic and useful rubric."

In conclusion, while limited, we found that a tool like MetaRubric can help educators to experience an iterative and collaborative process of assessment design and gain perspectives about how they can engage students in the process of assessment tool development. In this way, teachers gained insights into the qualities of good rubrics through a playful experience that took less than two hours.

18.4 Implications of Play for Assessment in the Digital World

There are several lessons that can be learned from the MetaRubric work to reimagine assessment through the lens of play in this ever-connected world. Firstly, bringing playfulness into assessment can offer educators and students an opportunity to engage with tasks that are generally seen as dry. For example, the mechanics of MetaRubric could be applied in e-portfolio assessments where students and the instructor iteratively refine the criteria of good work in the process of collecting and documenting artifacts to make the whole assessment process more engaging and game-like. This would also help both students and instructors to understand how standards and criteria are captured in artifacts, and how scoring takes place.

Secondly, a playful assessment tool such as MetaRubric can be used to assist in developing rubrics that support students to develop twenty-first century skills (e.g., critical thinking and communication skills) through self-expression and social interactions. For example, instructors could consider adopting a playful assessment activity that leverages the use of social media and online tools (e.g., Twitter or Instagram) by younger generations, where the students demonstrate evidence for their learning through diverse forms of self-expression.

Finally, we can imagine more tools like MetaRubric implemented in not just face-to-face learning environments, but online learning environments as well to create playful and social interactions between students and educators, most obviously synchronously but potentially asynchronously. For instance, the idea of playful and collaborative creation of assessment rules could be applied in a purely online learning environment to foster social aspects of assessment that would otherwise be limited unless intentionally facilitated. For example, at the MIT Media Lab, the facilitators of an online professional development course delivered via a digital platform called Unhangout used MetaRubric to facilitate playful interactions among

learners who had never met in person to engage in an intimate discussion of what defines good learning, a conversation which would be difficult if the only available medium was an online forum.

We speculate that the underlying principles of MetaRubric could be applied in a wide range of contexts for diverse learners, from K-12 to postgraduate, as a means of developing young people's understanding of how assessment works more broadly in the digital world, what Price calls assessment literacy (Price et al. 2011). This type of literacy is a crucial competency for young people to become literate citizens as they spend enormous hours online, for example, one survey reports that teenagers spend nearly nine hours a day online, a space where people need to critically evaluate credibility of different information in the internet (McGrew et al. 2017). As MetaRubric is inherently playful, it can build assessment literacy in an engaging but still persuasive way. Students could even assess rubrics being used in other classes using the rubric they have developed. In this way, the empowerment agenda described by the innovative teachers can be perpetuated.

It is worth also sounding a note of caution. While playfulness is closely connected to learning, attempts to connect assessment with playfulness have been limited thus far, and, therefore, more research needs to be done to show how playfulness can broaden assessment practices in general. One challenge that we anticipate with this work is the fundamental tension that underlies the assessment development process. Allowing learners to experiment, explore, and socially engage with learning experiences on the surface seems to conflict with the common practices of assessment, where the assessor (i.e., teachers, assessment developers) predetermines what is or is not acceptable work, practices which are rarely collaborative. Additionally, traditional psychometric evaluation techniques have heavily focused on single measures of validity and reliability, and, therefore, on constraining aspects of tasks or contexts to be able to establish rigor (DiCerbo et al. 2017).

For future work, we will need to directly investigate students' experiences using tools like MetaRubric, how they collaboratively create performance rubrics, and what the drawbacks and benefits of such an approach are. Additionally, we need to develop a better understanding of why some educators might be reluctant to embrace the idea of a playful and collaborative approach to assessment, and determine what additional tools and supports those educators might need to fully adopt tools like MetaRubric.

18.5 Conclusion

In this chapter, we discuss limitations of the standard rubric creation process and illustrate how the playful assessment tool, MetaRubric, has been used to help educators learn about better rubric design principles. Based on feedback we also propose that it can help engage students as partners in the assessment process. Through our approach, we intend to address the contrast between learner-centered, authentic, and playful ways of learning and rigid assessment tools that often lack these

characteristics. By doing so, we hope that playful assessment can increase the overall literacy of young people who are living in this very connected digital world where they need to critically participate in the creation and evaluation of information. While there are many innovative efforts to improve higher education to better cater to the learning desires of the young generations, assessment has been slow and difficult to change. We propose that playful tools such as MetaRubric can be used to transform assessment, making it more meaningful, relevant, empowering, and even enjoyable for the generations who grow up in this global and digital era.

References

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics, 118*(4), 1279–1333.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. In *Proceedings of the 120th ASEE Annual Conference & Exposition 2013* (pp. 1–18). Atlanta: ASEE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Bryson, C. (2014). *Understanding and developing student engagement*. Abingdon: Routledge.

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching, 43*(4), 349–352.

Charmaz, K. (2014). *Constructing grounded theory*. Los Angeles & London: SAGE.

Christensen, C. M., & Eyring, H. J. (2011). *The innovative university: Changing the DNA of higher education from the inside out*. Somerset: John Wiley & Sons.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (MindTrek '11)* (pp. 9–15). Tampere: ACM.

DiCerbo, K., Shute, V. J., & Kim, Y. J. (2017). The future of assessment in technology rich environments: Psychometric considerations. In J. M. Spector, B. Lockee, & M. Childress (Eds.), *Learning, design, and technology: An international compendium of theory, research, practice, and policy*. New York: Springer.

Florida, R. (2002). *The rise of the creative class and how it's transforming leisure, community and everyday life*. New York: Basic Books.

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53*, 69–76.

Freidman, T. L. (2005). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.

Gee, J. P. (2003). *What video games have to teach us about literacy and learning*. New York: Palgrave Macmillan.

Helle, L., Tynjälä, P., & Olkinuora, E. (2006). Project-based learning in post-secondary education–theory, practice and rubber sling shots. *Higher Education, 51*(2), 287–314.

Jiusto, S., & DiBiasio, D. (2006). Experiential learning environments: Do they prepare our students to be self-directed, life-long learners? *Journal of Engineering Education, 95*(3), 195–204.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144.

Juul, J. (2005). *Half-real: video games between real rules and fictional worlds*. Cambridge, MA: MIT press.

Levy, F., & Murnane, R. J. (2004). Education and the changing job market. *Educational Leadership, 62*(2), 80–83.

Lillard, A. S., Lerner, M. D., Hopkins, l. E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin, 139*(1), 1–34.

López-Pastor, V., & Sicilia-Camacho, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education, 42*(1), 77–97.

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator, 41*(3), 4–9.

OECD. (2014). *What it means for higher education: The impact of skills on early labour market outcomes – Insights from the OECD employment outlook 2014*. Paris: OECD. Retrieved from http://www.oecd.org/education/imhe/HE-Quarterly-December-2014-EmploymentOutlook.pdf.

Osterweil, S., & Klopfer, E. (2011). Are games all child's play? In S. de Freitas & P. Maharg (Eds.), *Digital games and learning* (pp. 152–171). London: Continuum International.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Pearson Report. (2018). *The next generation of learners*. Retrieved from https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/news/news-annoucements/2018/The-Next-Generation-of-Learners_final.pdf.

Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education, 36*(4), 479–492.

Resnick, M. (2014, August). Give P's a chance: Projects, peers, passion, play. In *Constructionism and creativity: Proceedings of the third international constructionism conference* (pp. 13–20). Vienna: Austrian Computer Society.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Zimmer, C. (2017, July 24). *Getting to know gen Z – Exploring middle and high schoolers' expectations for higher education*. Retrieved from https://next.bncollege.com/wp-content/uploads/2015/10/Gen-Z-Research-Report-Final.pdf.

Yoon Jeon Kim is the executive director of the MIT Playful Journey Lab. Her work is centered on the topic of innovative assessment and how technological advancement influences what we are measuring about student learning and how we are measuring it. For over 10 years, she has applied evidence-centered design in combination with educational data mining techniques in a wide range of game- and simulation-based assessment development and evaluation projects. The core of her work is close collaboration with practitioners, empowering teachers to innovate around classroom assessment and use playful assessment tools that can truly impact student learning.

Louisa Rosenheck is Associate Director of the MIT Playful Journey Lab, and a designer and researcher of educational technologies in the MIT Education Arcade. She manages the design, content, and development of educational games and simulations, and oversees the research on how games can be effectively used in both formal and informal educational settings. She also brings the playfulness of games into her assessment design work, creating digital and non-digital tools to help both students and teachers assess creative work and soft skills. In addition to these areas, she supports other designers and organizations in creating innovative ed tech and curriculum. Through teaching at the graduate level and various capacity building partnerships, she aims to increase the deep learning in games and tools being produced for education.

Chapter 19
Sharing Achievement Through Digital Credentials: Are Universities Ready for the Transparency Afforded by a Digital World?

Trina Jorre de St Jorre

Abstract Social and professional platforms have changed the ways in which we connect and share, but most universities still evidence and communicate achievement in the same ways as they have for decades. Digital credentials have emerged as a way of sharing more detailed achievement than is possible through a grade or academic transcript including the circumstances of achievement (standards and criteria) and the evidence that justified it (student work). Achievement can also easily shared through social and professional platforms. Where digital credentials are associated with meaningful assessment they can be used to provide graduates with rich evidence of achievement that might help them to gain opportunities. However, public sharing also means that the content and quality of student work, and the consistency of judgements made, are open to scrutiny by a much broader audience. This has implications beyond the assessment task: for students (whose work and achievement is shared); for teachers (who design assessment and make judgements); and for institutions (who are held accountable for quality and consistency). This chapter will consider the transparency associated with current assessment practices, the benefits and risks associated with use of digital credentials to denote student achievement, and the readiness of universities for transparent assessment.

19.1 Introduction

Social and professional platforms have changed the ways in which we connect and share, so much so that many of us now remain digitally connected almost constantly through portable devices and platforms such as Facebook, LinkedIn and Twitter, amongst others (Sensis 2017). Anyone who cares to, can easily produce, reproduce

T. Jorre de St Jorre (✉)
Deakin University, Melbourne, VIC, Australia
e-mail: trina.j@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_19

and share information across multiple channels using diverse multimedia: why use words when you can use a picture, and why use a picture when you can convert an image to an animated gif[1]? Over 562 million people are using the professional networking platform 'LinkedIn' alone (https://about.linkedin.com/), and social and professional platforms are increasingly being used in job recruitment and selection processes, including the screening of candidates (Caers and Castelyns 2011; Priyadarshini et al. 2017).

In contrast, universities have done very little to utilise the connectivity offered by digital technologies, and still communicate and evidence achievement in the same ways as they have for decades. Grades are used almost universally to represent levels of academic achievement, and an academic transcript is provided at graduation as evidence of achievement that might be used to gain employment or access to advanced study (Sadler 2009). However, academic transcripts provide poor evidence of what students actually know and can do, and rarely capture transferable skills required in the workplace, such as teamwork, communication and critical thinking (Boud and Associates 2010; Flynn 2004). Some universities have digitised academic transcripts, but these are by no means widespread, and achievement is still presented in the same format and limited detail as with the paper-based tradition.

Digital credentials, utilising 'badging technologies', have emerged as a way of representing more detailed evidence of achievement than is possible through a grade or academic transcript (Oliver 2016). Digital badges are "electronic symbols used to document performance and achievement" (Carey and Stefaniak 2018). They consist of a digital image that links to additional information, such as: the context of achievement, the criteria for assessment, the identity of the recipient and the organisation that issued the credential, and examples of the work submitted for assessment. Where issued through a compatible platform, they can be used to easily share achievement with a broad audience through social and professional platforms. Thus digital credentials could be used in addition to or in replacement of grades, to convey more meaningful evidence of achievement, such as the mastery of transferable skills or students' capacity to integrate and apply learning from a course to other contexts.

As with any credential, value of a digital credential is dependent on the integrity of the assessment task and judgements made, but public sharing means that the content and quality of student work, and the consistency of judgements made, are open to scrutiny by a much broader audience. This has implications beyond the assessment task: for students (whose work and achievement is shared); for teachers (who design assessment and make judgements); and for institutions (who are held accountable for quality and consistency). This chapter will consider the transparency associated with current assessment practices, the benefits and risks associated with using digital credentials to denote student achievement, and the readiness of universities to deliver more transparent assessment in a digital world.

[1] A gif is a digital image file type that supports both static and animated images.

19.2 Transparency in Assessment

Whereas assessment was once considered 'secret teacher's business' (Boud 2014), it is increasingly being recognised that we must open up the 'black-box of tertiary assessment' (Hattie 2009). "We implicitly trust our academics to know what they value in their subjects, to set examinations and assignments, to mark reliably and validly, and then to report these marks" (Hattie 2009, p. 259). However, it is students who must interpret and respond to the guidelines and feedback provided (Carless 2017) and it is institutions that are held accountable for academic standards (Sadler 2011). Furthermore, judgements about academic achievement, in the form of grades and academic transcripts, are used by third parties to make decisions that affect the employment, career and life prospects of graduates (Sadler 2011). Thus, transparency in the communication of what is expected and what has been achieved, has implications for the learning and future prospects of students, the comparability of graduates and the quality assurance of institutions.

Assessment practices have become more transparent than they once were (Boud 2014). This increase in transparency has been driven by a shift from assessment of learning, to assessment for learning, which is now a well-entrenched pedagogy in higher education: as evidenced by a rapidly expanding body of literature and research related to its principles over the past two decades (Carless 2017; Hattie 2009). Assessment for learning recognises the influence of assessment over where students direct their effort, and formative feedback as a mechanism through which students can identify opportunities for targeted improvement (Carless 2017; Gibbs 2010). A key role of teachers in delivering assessment for learning (and student-focussed learning more generally) is to develop students' capacity to understand and judge their own work. To do so, students need to understand the learning outcomes and quality of work required (Sadler 2010a). Thus there is now an expectation that students be provided with transparent learning outcomes, the standards and criteria upon which judgements will be made, and feedback on their performance.

However, the ways in which student achievement is judged and shared are still not transparent. For example, while learning outcomes are now more often shared publicly, the standards and criteria upon which judgements are made are usually shared only with students in relation to a specific assessment task, and are rarely made public or explicit at a course level (Boud 2017). Furthermore, achievement across assessment tasks is averaged, subject to variable weighting, and summed to produce a single subject grade that provides little detail about what students actually know or can do (Boud 2017; Jackel et al. 2017). In principle, standards-based approaches to assessment assume students will be able to achieve learning outcomes as a result of a sequence of learning activities and that assessment demonstrates evidence of students having met those requirements (Boud 2017). However, in practice, assessment usually fails to differentiate between learning outcomes, so whether an individual excels or fails in a particular area is not captured in their overall grade. Those that fail to achieve a learning outcome can still pass or even do well if they achieve well in other outcomes.

19.3 The Impetus for Greater Transparency

Quality assurance agencies are increasingly requiring universities to communicate and evidence learning outcomes beyond inputs and objectives (Australian Qualifications Framework Council 2013; El-Khawas 2014; New Zealand Qualifications Authority 2011; Tuck 2007). There is also an increasing expectation that in addition to discipline specific outcomes, graduates should achieve broader learning outcomes of importance to lifelong learning and employability (Commonwealth of Australia 2015). These requirements have implications not only for what is assessed (capabilities beyond discipline knowledge and skills) but also for how universities evidence, aggregate and represent achievement. However, as yet there has been little change in how achievement is assessed or represented at a course level (Boud 2017).

The academic transcript is the most detailed record of achievement provided by universities, and reveals patterns of high and low performance across a degree programme (Flynn 2004). However, transcripts only communicate learning that is for credit, and focus on subjects rather than capabilities, so they capture knowledge (vaguely) rather than skills. Academic transcripts provide little or no information about the context, criteria and standards associated with assessment and mean little to stakeholders outside the issuing institution (Flynn 2004). This lack of transparency makes it difficult to compare graduates completing different degrees at the same institution, let alone those from different institutions or jurisdictions. Comparability of graduates is important to their mobility, especially given the context of globalisation and the internationalisation of education (Sadler 2010b).

To provide more context about achievement, universities in some countries now provide statements to supplement the academic transcript. For example, the European Diploma Supplement (European Commission Education and Training 2010) and the Australian Higher Education Graduation Statement (Australian Government 2015) provide descriptions of the degree or award and the institution that issued it, and can include short descriptions of extra-curricular achievement such as university awards or programs. However, these documents still do little to identify the criteria or standards of achievement that were met to achieve the degree. If these are made more explicit, and students are judged by what they can do, rather than how they compare to other students in a specific cohort, outcomes could be more effectively compared across courses, institutions or countries (Boud 2017).

Whilst comparability is important for those who would judge graduates, opportunities for differentiation are important for graduates themselves. Graduate competition and unemployment have increased and academic achievement or attainment of a degree is no longer the differentiator it once was (Callaghan 2011; Graduate Careers Australia 2013, 2014). To be competitive, graduates need to be able to evidence mastery of important transferable skills such as teamwork, communication and critical thinking (Brown et al. 2003; Burning Glass Technologies 2014; Tomlinson 2008). Concerns have been raised worldwide, over the suitability of higher education for teaching these capabilities and fulfilling the rapidly evolving

needs of the future workforce (AlphaBeta 2015, 2016; Bridgstock 2009; Jackson 2013; Rae 2007). Some employers have acted on these concerns by dropping degree requirements altogether and instead assess job applicants on their ability to demonstrate the capabilities required, regardless of where they have been gained (Coughlan 2016; Sherriff 2015). Thus, graduates need more meaningful evidence of their capabilities than that conveyed by grades or academic transcripts, and need to understand how to integrate knowledge and experience from all aspects of their lives.

Institution-wide practices to purposefully scaffold the development of important capabilities across entire degrees have become more prevalent (Lawson et al. 2013). Numerous approaches to curriculum mapping specifically focussed on the development of graduate capabilities at course level have been reported in the literature and have been attributed with variable impact (Bath et al. 2004; Oliver 2013, 2015; Spencer et al. 2012). However, even where capabilities are embedded and assessed within the curriculum, they are rarely made explicit to students (Jorre de St Jorre and Oliver 2018), let alone to other stakeholders who might offer graduates opportunities. Assessment for employability, should provide students with meaningful evidence of their achievement whilst also developing students' capacity to articulate and make judgements about their own capabilities (Jorre de St Jorre and Oliver 2018). Thus universities need to make their expectations and student achievement more transparent – to employers, students and those who teach them.

19.4 Affordances (and Limitations) of Digital Credentials

Digital micro-credentials have emerged as a way of representing more detailed evidence of achievement than is possible through a grade or academic transcript, such as the circumstances of and the evidence that justified it (Oliver 2016). They are similar to digital portfolios in that they can be used by individuals share digital evidence of achievement; however, digital credentials have additional functionality, in that they can be used to succinctly verify how and by whom achievements were judged, whilst also linking to artefacts such as a portfolio or other assessment evidence. Where issued through a compatible platform, 'digital badging' technologies allow credentials to be easily shared through professional platforms or social media applications such as LinkedIn, Facebook and Twitter. Thus digital credentials can be used to provide concise and validated evidence of achievement with a broad external audience.

Where digital credentials are associated with meaningful assessment and judged against holistic criteria and standards they can provide rich evidence of achievement (Miller et al. 2017). Furthermore, where the assessment associated with a digital credential requires students to integrate achievement from their course with achievement from other aspects of their lives, or to transfer learning from the course to other diverse contexts, they can be used to share personalised achievement, which differentiates them from other graduates. The case study in Box 19.1 provides an example of how digital credentials have been used by an Australian University to incentivise and recognise achievement relevant to employers.

Box 19.1 Digital Credentials Case Study
Case study: Employability Credentials in Practice.

Deakin Hallmarks provide an example of how digital credentials might be used to incentivise and recognise achievement that is relevant to employers (Jorre de St Jorre et al. 2016; Miller et al. 2017). Deakin Hallmarks are extra-curricular university awards that are contextualised to degree programs and aligned with Deakin's graduate learning outcomes. The awards are developed and assessed in collaboration with industry or professional partners and recognise outstanding achievement of transferable capabilities, including communication, teamwork, global citizenship, digital literacy, critical thinking and problem solving.

Students must apply for the award by drawing together evidence that meet the standards and criteria associated with the award, often through documentation in a digital portfolio. Students who gain the award receive a digital credential which provides a description of the context and rationale for the award, the standards and criteria achieved, the organisations involved in the development of those criteria, the identity of the panel of experts who assessed and endorsed their achievement, and the work submitted to gain the award (Fig. 19.1).

The perceptions of students and employers who participated in a pilot of the above credentialing program, suggest that use of digital credentials to recognise capabilities at a degree program level, can be used to engage students in reflection on their capabilities, career and personal identity, and provide them with evidence that is meaningful to employers (Miller et al. 2017).

While digital credentials can be used to recognise and share meaningful achievement, the design of digital credentials is ultimately at the discretion of the issuer, within constraints (functionality and conventions) imposed by the issuing platform – of which there are many (Open Badge Network 2016). Through those platforms, digital credentials can now be easily be designed, issued and shared (for free

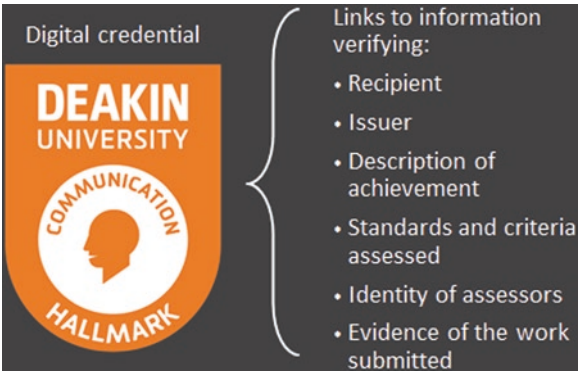


Fig. 19.1 Employability credentials in practice

or for a price), making it relatively simple for digital credentials to be issued to anyone, by anyone, for anything. Consequently, the achievement required and information communicated through credentials is highly variable, as is their value to different users and for different purposes (Abramovich et al. 2013; Carey and Stefaniak 2018). Thus there is still much debate as to what pedagogical value they have and what purpose they are best suited e.g., as rewards, incentives or to recognise achievement (Carey and Stefaniak 2018; Farmer and West 2016; Gibson et al. 2015).

To some, digital badging technologies are associated with negative preconceptions, because they have often been used in association with low level tasks and competencies, or issued for participation (Carey and Stefaniak 2018). The term 'badge' itself also has negative connotations for some. For example, physical badges are used to indicate competencies in childrens programs, such as the Boy Scouts (Anderson and Staub 2015). In a gaming context, badges are often used to indicate status or progression through levels of achievement along a storyline, and collecting badges often permits access to new levels, tasks or rewards (Gibson et al. 2015). Similarly, digital badges have also been used by commercial companies to indicate status, promote desired online behaviours, foster brand loyalty and encourage customer retention (Gibson et al. 2015). Even in the context of higher education, digital badges are often used to recognise participation in an activity or completion of a task, rather than to recognise learning, understanding or skills (Carey and Stefaniak 2018). Thus use of digital badging technology in this manner, has led some critics to dismiss the concept because they associate the use of this technology with the reduction of learning to overly discrete or lower level skills (Carey and Stefaniak 2018).

In contrast, some advocates of digital badging technologies have made broad and ambitious claims about the superiority of 'badges' to traditional assessment, but have failed to differentiate between assessment and the technology itself (Anderson and Staub 2015). Digital badging technologies provide an alternative way of representing achievement. However, the assessment associated with a credential can be just as flawed as that represented by a grade, and the information conveyed through the credential can still be vague or even misleading. Where assessment is flawed, the transparency afforded by digital credentials can actually be detrimental, both to the student who receives the credential and the institution that confers it.

19.5 The Readiness of Universities for Public Sharing in a Digital World

The value of any credential is determined by the credibility of the assessment process and the reputation of the assessor. However, the transparency afforded by digital credentials makes the legitimacy of the assessment process and the judgements made even more important because they can be judged by a broad audience. There is an abundance of literature on what constitutes good practice in assessment (e.g.,

Boud and Falchikov 2007; Carless 2017). However, the staff responsible for the design and delivery of assessment are rarely experts and implementation often fails to meet these aspirations (Bearman et al. 2016, 2017; Price et al. 2011). Evidence suggests that current assessment practices are not always fair (Hailikari et al. 2014). Indeed, many of the practices routinely employed in higher education institutions compromise grade integrity and it is largely the lack of transparency associated with grades that allows these practices to go unchallenged (Sadler 2009). These flaws in assessment are not new or unique to assessment associated with digital credentials, but they must be a consideration for those who design assessment, make judgements and communicate achievement through digital credentials.

Performance standards can help to improve the transparency of assessment and students learning, however, they are often poorly articulated, unclear and confusing to students (Hendry et al. 2012). Where credentials are to be shared with a broad audience for the purpose of warranting employability, it is also imperative that they represent what students are able to do at the completion of their course (are summative) and reflect the skills and knowledge valued by employers or other stakeholders (Boud 2017). Achievement is also more meaningful, where it is individualised and requires students to conceptualise and integrate complex and abstract concepts (Boud 2017). This can be achieved where assessment is sufficiently open to allow for multiple solutions and student work is judged against holistic standards and criteria. However, judgements made against holistic criteria and standards are subjective and may be open to greater scrutiny, especially if the context of those judgements is not well communicated or understood. Thus the credibility and experience of assessors is important to ensuring that the judgements made are fair and valid. Assessment by a panel rather than an individual is one way of assuring the quality of the judgements made, as per the Deakin Hallmarks case study provided above.

Another consideration for the design and sharing of achievement through digital credentials, is the nature and legitimacy of the assessment artefacts themselves. Assessment is traditionally a private communication between a student and a teacher, but when assessment artefacts (examples of student work) are embedded in a credential as evidence of what was achieved, this communication becomes public. This has direct implications for students in terms of their graduate and digital identity, as well as for any third parties to whom they refer in that discourse.

The ways in which identity is constructed by ourselves and others is multifaceted and has become more complex with digitisation and the proliferation of social and professional networks (Cover 2016). Digital media provides new opportunities for self-promotion and for individuals to fashion their own representations through selection of what is shared. However, every time we engage with digital spaces or networks, we leave traces of ourselves that are open to interpretation by others, not just in that moment, but also into the future (Cover 2016). Interactivity (the ways in which people engage and communicate) in digital spaces can also have negative repercussions where it is at odds with the identity that an individual intends to promote at another time or within a specific context, such as how they might want to represent themselves as professionals. Thus, institutions that require or encourage

students to share their work through digital credentials, must also consider how they develop digital literacy around what is appropriate to share, not only in terms of copyright and confidentiality, but also in terms of digital and graduate identity.

19.6 Balancing Risk with Opportunity

In summary, digital micro-credentials offer an alternative way of denoting learner achievement: they have affordances that can be used to make assessment more transparent, and to provide graduates with more meaningful evidence of their achievements, especially where they are used to evidence summative assessment at the end of programs of study. However, there are significant risks associated with doing so without addressing current flaws in assessment practices and considering the implications of public sharing for institutions, teachers and students. That is not to say that universities should not utilise the affordances of digital technologies to warrant their graduates' achievements. Rather, the sector and individual institutions need to consider how they will uphold academic standards and protect their reputations, whilst also serving the interests of their graduates and adapting to a rapidly changing and increasingly digital world. Amongst other measures, institutions must move forward on assessment reform and develop and support the educators responsible. While most of the literature on assessment has focussed on the students' experience of assessment, to improvement assessment in practice, we also need to understand how academics design, implement and make judgements about assessment (Bearman et al. 2016). Team approaches to assessment design, governance and quality assurance across degree programs will also be important.

References

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education?: It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development, 61*(2), 217–232.

AlphaBeta. (2015). *The New Work Order. Ensuring young Australians have skills and experience for the jobs of the future, not the past*. Melbourne: Foundation for Young Australians.

AlphaBeta. (2016). *The New Basics: Big data reveals the skills young people need for the New Work Order*. Melbourne: Foundation for Young Australians.

Anderson, D. M., & Staub, S. (2015). Postgraduate digital badges in higher education: Transforming advanced programs using authentic online instruction and assessment to meet the demands of a global marketplace. *Procedia – Social and Behavioral Sciences, 195*, 18–23. https://doi.org/10.1016/j.sbspro.2015.06.165.

Australian Government. (2015). *Australian higher education graduation statement*. Retrieved from https://www.education.gov.au/australian-higher-education-graduation-statement

Australian Qualifications Framework Council. (2013). *Australian qualifications framework* (2nd ed.). Adelaide: Author.

Bath, D., Smith, C., Stein, S., & Swann, R. (2004). Beyond mapping and embedding graduate attributes: Bringing together quality assurance and action learning to create a validated and living curriculum. *Higher Education Research & Development, 23*(3), 313–328.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the Assessment Design Decisions Framework. *Teaching in Higher Education, 21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: A cross-disciplinary study. *Higher Education, 74*(1), 49–64. https://doi.org/10.1007/s10734-016-0027-7.

Boud, D. (2014). Shifting views of assessment: From secret teachers' business to sustaining learning. In J. McArthur, D. Hounsell, N. J. Entwistle, C. Kreber, & C. Anderson (Eds.), *Advances and innovations in university assessment and feedback: A Festchrift in honour of professor Dai Hounsell*. Edinburgh: Edinburgh University Press.

Boud, D. (2017). Standards-based assessment for an era of increasing transparency. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education*. Dordrecht: Springer.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. London: Taylor & Francis.

Bridgstock, R. (2009). The graduate attributes we've overlooked: Enhancing graduate employability through career management skills. *Higher Education Research & Development, 28*(1), 31–44. https://doi.org/10.1080/07294360802444347.

Brown, P., Hesketh, A., & Williams, S. (2003). Employability in a knowledge-driven economy. *Journal of Education and Work, 16*(2), 107–126.

Burning Glass Technologies. (2014). *Moving the goalposts: How demand for a Bachelor's degree is reshaping the workforce*. Boston: Burning Glass Technologies.

Caers, R., & Castelyns, V. (2011). LinkedIn and Facebook in Belgium: The influences and biases of social network sites in recruitment and selection procedures. *Social Science Computer Review, 29*(4), 437–448. https://doi.org/10.1177/0894439310386567.

Callaghan, R. (2011). *Selling the dream: Are we offering employability or making a vocational offer?* In paper presented at the 20th Annual Teaching Learning Forum, Perth, WA. Retrieved from http://otl.curtin.edu.au/tlf/tlf2011/refereed/callaghan.html

Carey, K. L., & Stefaniak, J. E. (2018). An exploration of the utility of digital badging in higher education settings. *Educational Technology Research and Development, 66*(5), 1211–1229. https://doi.org/10.1007/s11423-018-9602-1.

Carless, D. (2017). Scaling up assessment for learning: Progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higer education*. Dordrecht: Springer.

Commonwealth of Australia. Department of Education and Training. (2015). *Higher education standards framework (threshold standards)*. Canberra: Australian Government.

Coughlan, S. (2016, 18 January). Penguin scraps degree requirement. *BBC News*. Retrieved from http://www.bbc.com/news/education-35343680

Cover, R. (2016). *Digital identities*. San Diego: Academic.

El-Khawas, E. (2014). Inside quality reform: Early results on using outcomes for improvement. *Quality in Higher Education, 20*(2), 183–194. https://doi.org/10.1080/13538322.2014.926083.

European Commission Education and Training. (2010). *The diploma supplement*. Retrieved from https://ec.europa.eu/education/diploma-supplement_en

Farmer, T., & West, R. E. (2016). Opportunities and challenges with digital open badges. *Educational Technology, 56*(5), 45–48.

Flynn, W. J. (2004). The case for revitalizing the traditional academic transcript. *Community College Journal, 74*(5), 27–32.

Gibbs, G. (2010). *Using assessment to support student learning*. Leeds: Leeds Met Press.

Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., & Knight, E. (2015). Digital badges in education. *Education and Information Technologies, 20*(2), 403–410. https://doi.org/10.1007/s10639-013-9291-7.

Graduate Careers Australia. (2013). *GradStats: Employment and salary outcomes of recent higher education graduates*. Retrieved from http://www.graduatecareers.com.au/wp-content/uploads/2013/12/GCAGradStats2013.pdf

Graduate Careers Australia. (2014). *GradStats: Employment and salary outcomes of recent higher education graduates*. Retrieved from http://www.graduatecareers.com.au/wp-content/uploads/2014/12/GCA_GradStats_2014.pdf

Hailikari, T., Postareff, L., Tuononen, T., Räisänen, M., & Lindblom-Ylänne, S. (2014). Students' and teachers' perceptions of fairness of assessment. In J. McArthur, D. Hounsell, N. J. Entwistle, C. Kreber, & C. Anderson (Eds.), *Advances and innovations in university assessment and feedback: A Festchrift in honour of professor Dai Hounsell*. Edinburgh: Edinburgh University Press.

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 & higher education student outcomes: Policy, practice & research*. Ako Aotearoa: Wellington.

Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education, 37*(2), 149–161. https://doi.org/10.1080/02602938.2010.515014.

Jackel, B., Pearce, J., Radloff, A., & Edwards, D. (2017). *Assessment and feedback in higher education: A review of literature for the Higher Education Academy.* Retrieved from https://www.heacademy.ac.uk/system/files/hub/download/acer_assessment.pdf

Jackson, D. (2013). Business graduate employability – Where are we going wrong? *Higher Education Research & Development, 32*(5), 776–790. https://doi.org/10.1080/07294360.2012.709832.

Jorre de St Jorre, T., & Oliver, B. (2018). Want students to engage? Contextualise graduate learning outcomes and assess for employability. *Higher Education Research & Development, 37*(1), 44–57. https://doi.org/10.1080/07294360.2017.1339183.

Jorre de St Jorre, T., Johnson, L., & Oliver, B. (2016). Deakin hallmarks: Principles for employability credentials. In *ASCILITE 2016: Shaping the future of tertiary education: Proceedings of the Australasian Society for Computers in Learning in Tertiary Education 2016 conference* (pp. 305–311). Adelaide: ASCILITE.

Lawson, R., Taylor, T., Herbert, J., Fallshaw, E., & French, E. (2013). *Hunters and Gatherers: Strategies for curriculum mapping and data collection for assuring learning – Final report 2013 (SP10-1862)*. Sydney: Office for Learning and Teaching, Department of Industry, Innovation, Science, Research and Tertiary Education.

Miller, K. K., Jorre de St Jorre, T., West, J. M., & Johnson, E. D. (2017). The potential of digital credentials to engage students with capabilities of importance to scholars and citizens. *Active Learning in Higher Education.* https://doi.org/10.1177/1469787417742021.

New Zealand Qualifications Authority. (2011). *The New Zealand qualifications framework*. Retrieved from https://www.nzqa.govt.nz/assets/Studying-in-NZ/New-Zealand-Qualification-Framework/requirements-nzqf.pdf

Oliver, B. (2013). Graduate attributes as a focus for institution-wide curriculum renewal: Innovations and challenges. *Higher Education Research & Development, 32*(3), 450–463. https://doi.org/10.1080/07294360.2012.682052.

Oliver, B. (2015). *Assuring graduate capabilities: Evidencing levels of achievement for graduate employability*. Sydney: Office for Learning and Teaching.

Oliver, B. (2016). *Better 21C credentials: Evaluating the promise, perils and disruptive potential of digital credentials*. Melbourne: Deakin University.

Open Badge Network. (2016). *Open Badge network discussion paper on Open Badges at policy levels* (O5-A1). Retrieved from http://www.openbadgenetwork.com/wp-content/uploads/2017/09/O5-A1_Policy_Discussion_Paper.pdf

Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education, 36*(4), 479–492. https://doi.org/10.1080/02602930903512883.

Priyadarshini, C., Kumar, Y., & Jha, R. R. (2017). Employer attractiveness through social media: A phenomenological study. *The Qualitative Report, 22*(4), 969–983.

Rae, D. (2007). Connecting enterprise and graduate employability: Challenges to the higher education culture and curriculum? *Education + Training, 49*(8/9), 605–619.

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education, 34*(7), 807–826. https://doi.org/10.1080/03075070802706553.

Sadler, D. R. (2010a). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015.

Sadler, D. R. (2010b). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education, 35*(6), 727–743.

Sadler, D. R. (2011). Academic freedom, achievement standards and professional identity. *Quality in Higher Education, 17*(1), 85–100. https://doi.org/10.1080/13538322.2011.554639.

Sensis. (2017). *Sensis social media report 2017*. Retrieved from https://irp-cdn.multiscreensite.com/535ef142/files/uploaded/Sensis-Social-Media-Report-2017.pdf

Sherriff, L. (2015, August 4). Ernst & Young removes degree classification from entry criteria as there's 'no evidence' university equals success. *The Huffington Post UK*. Retrieved from http://www.huffingtonpost.co.uk/2015/08/04/ernst-and-young-removes-degree-classification-entry-criteria_n_7932590.html

Spencer, D., Riddle, M., & Knewstubb, B. (2012). Curriculum mapping to embed graduate capabilities. *Higher Education Research & Development, 31*(2), 217–231. https://doi.org/10.1080/07294360.2011.554387.

Tomlinson, M. (2008). The degree is not enough: students' perceptions of the role of higher education credentials for graduate work and employability. *British Journal of Sociology of Education, 29*(1), 49–61. https://doi.org/10.1080/01425690701737457.

Tuck, R. (2007). *An introductory guide to National Qualifications Frameworks: Conceptual and practical issues for policy makers*. Geneva: International Labour Organization (ILO).

Trina Jorre de St Jorre is a Senior Lecturer at Deakin University, where her research and practice is focused on strategies for enhancing student achievement and graduate employability. She is interested in strategies that develop graduate identity and engage students in the conscious development of attributes of importance to their employability. She leads the development of Deakin Hallmarks: digital micro-credentials designed to develop and promote graduate employability.

Part V
Conclusion

Chapter 20
Concluding Comments: Reimagining University Assessment in a Digital World

Phillip Dawson and Margaret Bearman

Abstract This chapter outlines three possible implications of the ideas explored in *Reimagining University Assessment in a Digital World*. These are: a renewed focus on the future; the changing nature of teachers' assessment work in a digital world; and new ways of thinking about scalability.

As we developed this book we knew what we wanted: a fundamental reimagining of assessment for our new digital world. We also knew what we did not want: a book showcasing interesting e-assessment approaches. Other books have done that before us, and it didn't address our fundamental concern of the need to rethink what assessment should be. The previous chapters in this book provide several new reimaginings of assessment. This concluding chapter builds on some common threads raised in Parts II, III and IV, to identify three new directions for assessment in a digital world: a renewed focus on the future; the changing nature of teachers' assessment work in a digital world; and new ways of thinking about scalability.

20.1 Futures

Part II: The changing role of assessment in the digital world put forward a range of arguments that collectively aim to persuade us that (a) the world is changing, and (b) assessment should prepare learners for this world. If the digital world of work will involve different types of work (O'Donnell, Chap. 9) and different types of self-presentation (Ajjawi et al., Chap. 6), then assessment needs to prepare learners for

P. Dawson · M. Bearman (✉)
Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, VIC, Australia
e-mail: p.dawson@deakin.edu.au; margaret.bearman@deakin.edu.au

M. Bearman et al. (eds.), *Re-imagining University Assessment in a Digital World*, The Enabling Power of Assessment 7,
https://doi.org/10.1007/978-3-030-41956-1_20

that work. If machines can do much of the work people used to do, we need to reconsider what we assess (Bearman and Luckin, Chap. 5), and how we incorporate machines into assessment (Dawson, Chap. 4). This focus within the book raises the demands put on assessment and assessment designers. We need to go beyond assessment being merely authentic for the present (Gulikers et al. 2004). A new type of 'future-authentic' assessment may be required: assessment that faithfully represents not just the current realities of the discipline in practice, but the likely future realities of that discipline.

Reimagining assessment so it is future-authentic presents a much greater challenge than present-authentic assessment. It requires an attempt at predicting the future, and imagining the circumstances that work will be undertaken within. The futures most touted in this book relate to the rise of artificial intelligence and big data. The argument is made here and elsewhere that learners should be developed to do things that machines cannot do, and this book discusses the role assessment can play in this development.

The danger in a future-oriented approach is that we might pick the wrong future or the wrong timescale for the future. Throughout history, technological progress has not usually been as fast as it is now. There have even been generations-long periods where technological progress stagnated or went backwards. While it is tempting to abandon the assessment of lower-level knowledge because in the future only machines will need this knowledge, this couples the utility of courses very closely to the continued existence of particular technology.

Future-authentic assessment is thus a useful goal, and a very hard one to implement, evaluate and research. However, authentic assessment designers could still benefit from adding a concern for the future to their design process: what will work look like in the years to come? What aspects of work are likely to be retained even when particular knowledge and skills change?

20.2 Teachers

Part III: The role of big data in reimagining assessment identified ways in which the digital world can provide educators with data to change their assessments (Rogaten et al., Chap. 11). It also provided ways in which computers can augment or even take on some of the roles traditionally performed by educators, such as Knight's chapter on formative feedback on written tasks (Chap. 10). However, elsewhere in the book there are challenges identified for educators in navigating this space. Bearman et al. raise concerns about how the lack of transparency in 'black box' analytics and algorithms may make educators' lives harder (Chap. 3). Although not the focus of any one chapter, as a whole the book paints a picture of the future of teachers' assessment work in higher education. In this picture, teachers will be doing more planning and more design work leading to more production of feedback information (albeit not necessarily by teachers) and more analysis about the impact of assessment upon students. Teachers will be focusing on those things they do that have most impact on learning and students' self-monitoring. This may mean that

other tasks – such as marking – could diminish, or possibly become outsourced. This change in the nature of academic work is happening alongside the unbundling of traditional holistic academic roles into more specialist positions focused on individual aspects of academic work (Macfarlane 2011). For example, in the Australian higher education sector, an increasing amount of staff are employed with the sole duty of marking student work in online courses. Assessment work for teachers in a digital world is and will be very different to what it once was.

The nature of assessment work in the digital world may be very attractive to some. Repetitive marking remains a great burden on many teachers, and the adoption of assessment technologies is often intertwined with a desire to contain assessment workloads (Bennett et al. 2017). However, Pitt and Winstone's chapter raises concerns that processes like feedback should not lose their human dialogic elements (Chap. 7). The question remains, which aspects of assessment and feedback need to be personalised against the shift towards massified university education?

As teachers hand more of their assessment work over to digital systems, the nature of the assessment work that is left is likely to become more specialised and bespoke. As discussed in Bearman and Luckin's chapter, a key challenge for assessment in a digital world is identifying what this work will be (Chap. 5). There are fundamental questions to be addressed about the boundary between various human roles and computers in assessment. Automated marking systems are becoming more accurate, and in *some* circumstances they exceed the accuracy of human markers. Automated feedback systems provide richer information about performance for students to make sense of in *some* contexts. Both of these domains are progressing at a pace that suggests better-than-human performance may be the norm in *some* topics in the coming years. However, this is likely not the case for *all* circumstances, contexts and topics. Having a better sense of where technology is most or least useful, may enable teachers to focus on uniquely human relationships and dialogue as in the educational alliance (Telio et al. 2015).

We suggest an important new role for teachers in the digital world of assessment. Teachers may increasingly need to work as advocates for ethical practice. The digital world of assessment is increasingly corporatized, and technology vendors have their own objectives (as discussed in Bearman et al's. chapter (Chap. 3)). Students have the difficult task of negotiating and presenting themselves in this new digital world (as discussed in Ajjawi et al's. chapter (Chap. 6)). As the remaining powerful humans in an education system that is increasingly automated, educators need to be crucial advocates for equity, fairness and justice. We see this as particularly important as university assessment continues to scale and oversight of individual student experiences becomes increasingly diffuse.

20.3 Scale

Part IV: Practical exemplars showcases what is possible for assessment in a digital world, including new ways to present student achievement (Jorre de st Jorre, Chap. 19), assessment designs to assure completion of learning outcomes (Cain et al.,

Chap. 16)), and technology to facilitate self and peer assessment (Tai and Adachi, Chap. 15)). Common across this section is an interest not just in designs that can make assessment better, but in designs that work given the pragmatics of the digital world. Of all these pragmatics, scale is perhaps most pressing. Higher education is now provided at massive scale in some contexts, with thousands of students enrolled in any given course module. This shift is inextricably linked to the digital world and therefore prompts the need to reimagine assessment. The necessity and affordances of scale are consequences of the digital world, through new types of courses like MOOCs (as discussed by Corrin and Bakharia, Chap. 14), new tools that are intended to improve efficiency (as critiqued by Pitt and Winstone, Chap. 7) and new approaches to degree enrolments that target global online markets.

Particular approaches to addressing scale are addressed in the chapters here and elsewhere in the literature, for example, in David Carless' book *Scaling up Assessment for Learning in Higher Education* (Carless 2017). Our book has added several chapters to this conversation that directly address the intersection of assessment, technology and scale, both scaling up of good practice, and the challenges that come with assessing at scale. However, there remains a gap in terms of metrics, concepts, and language to discuss differences in the ways that assessments scale.

One helpful means of understanding the challenge may be the adaptation of scalability concepts from the computer sciences. When programmers are designing software, they think about scale in terms of a rough relationship between the size of a dataset and the amount of resources required to process that dataset. This gives them a rough metric to compare the resources required for different approaches to solving a problem. Adapted to an assessment context, marking individual student essays scales *linearly*; it takes roughly twice as long to mark twice as many essays. The provision of useful exemplars to a cohort of students might scale *better than linearly* because as the number of students increases it is less likely that new ways to present a task will be identified beyond what is identified in a smaller cohort. Finally, programs that scale as a *constant* require no more resources to run for a large dataset compared to a small dataset; the most obvious parallel here is computer-marked multiple-choice questions, which consume as much assessor time for a

Table 20.1 Application of computer science notions of scalability to the domain of assessment

Scale	Assessment example
Constant	Computer marked MCQs use no additional human time per student
Linear	Marking individual student essays
Worse than linear	Comparing every student assignment against every other student assignment
Better than linear	Building a database of common exemplars for different cohorts of learners

student population of 1000 as they are for 10,000. Table 20.1 below summarises the ways that some potential approaches might scale.

Massive Open Online Courses (MOOCs) provide evidence that it is possible to develop assessment regimes that scale better than linearly. MOOCs are often funded with a set amount of resources and require assessment approaches that consume the same resourcing regardless of the numbers of students. Often these approaches rely heavily on computer-marked work, and computer-arranged peer- and self-assessment (as discussed in Corrin and Bakharia's chapter (Chap. 14)). The MOOC literature thus documents the successes and failures of attempts to make assessment scale as a constant (Dawson and Henderson 2017).

For an assessment regime to be feasible it needs to scale the same way (or better) than the available resources scale. Since resourcing models for education have traditionally tended to scale linearly, with more student fees providing for more educators, most traditional assessment approaches have also scaled linearly. This means educators have needed to make hard decisions about where to invest their resources. Using digital understandings of scale like those in Table 20.1 can help educators re-think deployment of assessment resources to where they can have greatest effect. Efficiencies in one area can be leveraged so that less efficient but more educationally important approaches can be used elsewhere. For example, approaches that scale better than linearly (e.g., MCQs) can be used in part of a unit so that more intensive dialogic feedback approaches can be deployed in another part.

The promise of automated marking and feedback systems is enabling workloads to scale better than linearly – that is, not requiring more resources for 1000 students than 10 students. Unfortunately, not all approaches that scale well achieve the intended purposes of assessment well. For example, high-stakes summative multiple choice assessment is very scalable, however it can have negative consequences for learning, even going as far as students learning false facts (Marsh et al. 2007). As many chapters within this book touch on, a key challenge for reimagining assessment in a digital world is to identify approaches that scale well while simultaneously enhancing learning and certifying achievement.

20.4 Reimagining

This was not an easy book to produce, because reimagining is difficult. In the absence of reimagining, much of what has happened at the intersection of assessment and technology has been reproduction of old tasks in new technologies.

This ideas in this book have prompted the challenges raised in this chapter of future-authenticity, teacher work and scalability. There are many other challenges that warrant further research.

The digital world means more for assessment than simply prepending an 'e-'. Without a thorough *reimagining* of assessment, university education will fail to fully prepare students for a newly digital world.

References

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology, 48*(2), 672–682. https://doi.org/10.1111/bjet.12439.

Carless, D. (2017). Scaling up assessment for learning: Progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 3–17). Singapore: Springer.

Dawson, P., & Henderson, M. (2017). How does technology enable scaling up assessment for learning? In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 209–222). Singapore: Springer.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67–85. https://doi.org/10.1007/bf02504676.

Macfarlane, B. (2011). The morphing of academic practice: Unbundling and the rise of the Para-academic. *Higher Education Quarterly, 65*(1), 59–73. https://doi.org/10.1111/j.1468-2273.2010.00467.x.

Marsh, E., Roediger, H., Bjork, R., & Bjork, E. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199. https://doi.org/10.3758/bf03194051.

Telio, S., Ajjawi, R., & Regehr, G. (2015). The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine, 90*(5), 609–614. https://doi.org/10.1097/acm.0000000000000560.

Phillip Dawson is an Associate Professor and Associate Director of the Centre for Research in Assessment and Digital Learning at Deakin University. He holds degrees in education, artificial intelligence and cybersecurity. His research focuses on assessment and learning matters like feedback, rubrics and academic integrity, often with influences from computer science.

Margaret Bearman is a Research Professor within the Centre for Research in Assessment and Digital Learning, Deakin University. Margaret's research focusses on learning and teaching in higher and professional education. She is particularly interested in ways of improving feedback, the sociomateriality of standards, and what it means to be an online learner in a digital world.