The Enabling Power of Assessment 5
Series Editor: Claire M. Wyatt-Smith

David Carless
Susan M. Bridges
Cecilia Ka Yuk Chan
Rick Glofcheski *Editors*

Scaling up Assessment for Learning in Higher Education

Springer

The Enabling Power of Assessment

Volume 5

Series editor
Claire M. 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

David Carless • Susan M. Bridges
Cecilia Ka Yuk Chan • Rick Glofcheski
Editors

Scaling up Assessment for Learning in Higher Education

Springer

Editors
David Carless
Faculty of Education
University of Hong Kong
Hong Kong, Hong Kong

Susan M. Bridges
Faculty of Education
University of Hong Kong
Hong Kong, Hong Kong

Cecilia Ka Yuk Chan
Ctr. Enhancement of Teaching & Learning
University of Hong Kong
Hong Kong, Hong Kong

Rick Glofcheski
Faculty of Law
University of Hong Kong
Hong Kong, Hong Kong

ISSN 2198-2643 ISSN 2198-2651 (electronic)
The Enabling Power of Assessment
ISBN 978-981-10-3043-7 ISBN 978-981-10-3045-1 (eBook)
DOI 10.1007/978-981-10-3045-1

Library of Congress Control Number: 2016957515

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This edited collection emanates from, but is not exclusive to, an international conference held at the University of Hong Kong, May 14–15, 2015. The conference was entitled Assessment for Learning in Higher Education and attracted 318 delegates from 27 countries. The keynote speakers were John Biggs, David Boud, Royce Sadler, David Carless and Rick Glofcheski.

The genesis of the conference was Rick Glofcheski's success in 2011 in being one of the two inaugural winners of a teaching excellence award organized by the University Grants Committee of Hong Kong and spanning all the universities in Hong Kong. This success built on Rick's numerous previous awards at the University of Hong Kong. Rick teaches in the Faculty of Law and is well known for his innovative approaches to assessment. To support the promotion of good practice in teaching, learning and assessment, Rick was awarded a substantial grant to support dissemination. Part of this funding was used to support a 1-day event in December 2012, involving speakers local to Hong Kong and David Boud as an overseas guest contributor.

The event in May 2015 which gave rise to this collection was an extension of the event in 2012. It involved collaboration between Rick, the Centre for Enhancing Learning and Teaching (CETL) and other colleagues at the University of Hong Kong. The organizing committee for the conference was led by Cecilia Ka Yuk Chan, Head of Professional Development in CETL, and included Grahame Bilbow, Director of CETL, Suki Ekaratne, Susan M. Bridges and David Carless.

It has been well known for many decades that students' learning behaviours are deeply influenced by their perceptions of the assessment tasks that they are tackling. Assessment for learning seeks to prioritize the learning function of assessment over its generally more dominant role of grading and certification. The chapters in this volume engage with different elements of assessment for learning and discuss how these might be implemented more widely.

The title of the collection is *Scaling Up Assessment for Learning in Higher Education*. The idea of scaling up denotes the quality, quantity and depth of implementation of particular pedagogic strategies. We are indebted to Phillip Dawson for suggesting this title over a lunch in Melbourne in September 2015. We should also acknowledge the contributions of Betty Pok-Yee Lee and Qiyun Zhu in formatting the chapters.

Approximately half of the chapters in the volume arise from presentations at the May 2015 conference, whilst we also invited a number of international experts who were not present at this event to contribute chapters. We are delighted with their response.

University of Hong Kong, Hong Kong, China
June 2016

David Carless
Susan M. Bridges
Cecilia Ka Yuk Chan
Rick Glofcheski

Contents

Contributors

Rola Ajjawi Deakin University, Burwood, VIC, Australia

Margaret Bearman Monash University, Melbourne, VIC, Australia

Chris Beaumont Edge Hill University, Ormskirk, Lancashire, UK

Michael G. Botelho University of Hong Kong, Hong Kong, China

David Boud Deakin University, Burwood, VIC, Australia

Susan M. Bridges Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong

David Carless Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong

Cecilia Ka Yuk Chan University of Hong Kong, Hong Kong, China

Phillip Dawson Deakin University, Burwood, VIC, Australia

Rick Glofcheski Faculty of Law, University of Hong Kong, Hong Kong, Hong Kong

Michael Henderson Monash University, Melbourne, VIC, Australia

Dai Hounsell University of Hong Kong, Hong Kong, China

Tansy Jessop Southampton Solent University, Southampton, UK

Anders Jönsson Kristianstad University, Kristianstad, Sweden

Elizabeth Molloy Monash University, Melbourne, VIC, Australia

Claire Moscrop Edge Hill University, Ormskirk, Lancashire, UK

Susanne Narciss Technical University of Dresden, Dresden, Germany

Ernesto Panadero Autonomous University of Madrid, Madrid, Spain

Edd Pitt University of Kent, Canterbury, UK

Charlotte E. Rees Monash University, Melbourne, VIC, Australia

Anna D. Rowe Macquarie University, Sydney, NSW, Australia

Claire M. Wyatt-Smith Australian Catholic University, Brisbane, Australia

Tracy X.P. Zou University of Hong Kong, Hong Kong, China

About the Editors and Authors

About the Editors

David Carless is Professor of Educational Assessment and Associate Dean (Learning and Teaching) in the Faculty of Education, University of Hong Kong. His research is focused on learning-oriented assessment and sustainable feedback in higher education. His most recent sole-authored book, published by Routledge in 2015, is entitled *Excellence in University Assessment: Learning from Award-Winning Practice*.

Susan M. Bridges is Associate Professor and Assistant Dean (Curriculum Innovation) with the Faculty of Education/Centre for the Enhancement of Teaching and Learning, University of Hong Kong. Her research interests are interactional and ethnographic, exploring the 'how' of effective pedagogy in higher education. Her most recent co-edited book published by Springer in 2016 is entitled *Educational Technologies in Medical and Health Sciences Education*.

Cecilia Ka Yuk Chan is the Head of Professional Development and an Associate Professor in the Centre for the Enhancement of Teaching and Learning at the University of Hong Kong. She is an expert in enhancing student learning particularly in the area of engineering, business and science education. Her research focuses on the development, assessment and evidence of generic skills.

Rick Glofcheski is Professor in the Faculty of Law, University of Hong Kong, teaching and researching tort law and labour law. His research interests include assessment for learning. Rick's teaching innovations have been recognized across the university sector. He was selected twice for the University Distinguished Teaching Award (2009 and 2015), and in 2011, he was selected for the inaugural University Grants Committee Teaching Award.

About the Authors

Rola Ajjawi is Senior Research Fellow in the Centre for Research in Assessment and Digital Learning at Deakin University, Australia. She has over a decade of experience in health professions education and has published over 60 peer-reviewed journal articles, book chapters and books. Her research focuses on workplace learning, assessment for learning and feedback design.

Margaret Bearman is Associate Professor in Health Professions Education and Educational Research (HealthPEER), in the Faculty of Medicine, Nursing and Health Sciences, Monash University. She has research interests in assessment and feedback within clinical and simulated environments. Her book, *Simulated Patient Methodology: Theory, Evidence and Practice*, was published in 2015 by Wiley-Blackwell.

Chris Beaumont is Senior Lecturer in the Faculty of Education, Edge Hill University, UK, and a National Teaching Fellow. He previously taught in industry, schools, FE and the Open University and was Deputy Director of the *Write Now* CETL. His research is focused on problem-based learning, employability and assessment feedback, particularly the application of artificial intelligence and technology to enhance dialogic feedback and promote self-regulated learning.

Michael G. Botelho is an Associate Professor at the Dental Faculty, University of Hong Kong. He has been involved with faculty educational reform or innovation for almost 20 years and has a special interest in problem-based learning, clinical assessments and e-learning. He is actively engaged in evaluating current educational practices using a student-centred approach to improve and redesign clinical curricula.

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

Phillip Dawson is Associate Professor and Associate Director of the Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Australia. He researches at the intersection of assessment, learning and technology. His most recent completed project explored how university educators make decisions in assessment design and produced extensive resources available at http://assessmentdecisions.org

Michael Henderson is an Associate Professor of Educational Technologies in the Faculty of Education at Monash University. He is the Programme Director for

Graduate Studies and leads the Learning with New Media group. His research is focused on the role of digital technologies in enhancing learning, including assessment feedback.

Dai Hounsell was Visiting Professor in the Centre for the Enhancement of Teaching and Learning at the University of Hong Kong from 2013 to 2016. He is Professor Emeritus of Higher Education and a former Vice Principal at the University of Edinburgh. His main interests are in understanding and improving the quality of university learning and teaching, particularly through the lenses of assessment and feedback.

Tansy Jessop is Professor of Research-Informed Teaching at Southampton Solent University. Her research interests are assessment and feedback, student engagement and curriculum design. Tansy has led the 'Transforming the Experience of Students Through Assessment' (TESTA) project, which has reached more than 50 universities in the UK, fostering more holistic design of assessment and feedback to benefit student learning.

Anders Jönsson is Professor of Education at Kristianstad University, Sweden. His main research interest is in classroom assessment, both for summative and formative purposes, but he has also been responsible for the development of the Swedish National Assessments in biology, chemistry and physics for year 6. Currently, Anders is working on a research project funded by the Swedish Research Council, investigating how teachers handle the needs of students who are at risk of failing certain school subjects.

Elizabeth Molloy is an Associate Professor in the Health Professions Education and Educational Research Unit at Monash University, Australia. Her areas of research include feedback, workplace-based learning and professional transitions. In 2009, she co-edited *Clinical Education in the Health Professions* (Elsevier) and, in 2013, co-edited *Feedback in Higher and Professional Education* (Routledge).

Claire Moscrop is Senior Lecturer and Programme Leader for the IT Management for Business Degree at Edge Hill University. Her research is focused on technology-enhanced learning, specifically in the enhancement of the assessment and feedback process and student engagement. Claire's current research projects include using eye trackers to assess student interaction with feedback and a project assessing the shift in feedback engagement across levels of study.

Susanne Narciss is Professor of Psychology of Learning and Instruction at the Technische Universitaet, Dresden. Her research interests include technology-enhanced self-regulated learning, the role of motivation and metacognition in instructional contexts and conditions and effects of informative tutoring feedback. Her chapter on feedback strategies was awarded the prestigious Distinguished Development Award 2007 by the American Association on Educational Communications and Technology.

Ernesto Panadero is a Researcher in the Faculty of Psychology at Universidad Autónoma de Madrid, Spain. His research studies the relationship between self-regulated learning and the use of formative assessment and aims to understand self-regulated learning at both individual and collaborative levels. He is one of the coordinators of the Special Interest Group 'Assessment and Evaluation' of the European Association for Research on Learning and Instruction and a member of the Continental Europe team in the Assessment for Learning International Network.

Edd Pitt is Lecturer in Higher Education and Academic Practice at the University of Kent, Canterbury, UK, where he is also Programme Director of the Postgraduate Certificate in Higher Education. His principal research interest is in assessment and feedback in higher education, with a particular interest in staff and students' emotional processing of feedback and its implications for pedagogy.

Charlotte E. Rees is Professor, Director of Medicine Curriculum and Director of Health Professions Education and Education Research in the Faculty of Medicine, Nursing and Health Sciences at Monash University, Australia. Her research focuses on patient-centred professionalism and workplace learning in healthcare education. Her most recent co-authored book is titled *Healthcare Professionalism* and will be published by Wiley-Blackwell in 2017.

Anna D. Rowe is an Early Career Researcher and Programme Research and Development Officer for Professional and Community Engagement (PACE) at Macquarie University, Australia. Her research focuses on the emotional and relational dimensions of feedback, assessment and other areas of learning and teaching in higher education. Other research interests include work-integrated learning and graduate attributes.

Claire M. Wyatt-Smith is Professor of Educational Assessment and Evaluation, Australian Catholic University. Claire is the inaugural Director of the *Learning Sciences Institute Australia* with offices in Brisbane, Sydney and Melbourne and serves as the founding Editor of the Springer series, *The Enabling Power of Assessment*. Her current research examines teacher assessment and the use of standards for improving learning and teaching, fairness and student diversity.

Tracy X.P. Zou is an Assistant Professor in the Centre for the Enhancement of Teaching and Learning at the University of Hong Kong. She is currently committed to developing a community of practice on teaching and learning across the University. Her research interests include assessment and feedback, experiential learning and communities of practice in higher education.

Part I
Enabling Assessment Change

Chapter 1
Scaling Up Assessment for Learning: Progress and Prospects

David Carless

Abstract A definition of assessment for learning (AfL) is provided. From a synthesis of relevant literature, I outline four main AfL strategies: productive assessment task design, effective feedback processes, the development of student understanding of quality and activities where students make judgments. I explore the notion of scaling up in relation to spread, depth, sustainability and shifts in ownership. Then I present a rationale for the scaling up of AfL following from dissatisfaction with current practices and persuasive research evidence on practices congruent with AfL. I relate the notion of scaling up to the geographical spread of AfL research activity, its somewhat modest impact on university assessment policies and in relation to the expansion of feedback research. I then consider what conditions might facilitate deeper and broader implementation of AfL, including the role of quality assurance, the importance of leadership and incentives, the development of assessment literacy through professional development activities and the potential of technology to act as a lever for enabling AfL strategies.

Introduction

Assessment for learning (AfL) is now reasonably well-entrenched as part of higher education (HE) pedagogy. It is well-recognized that assessment is a crucial driver of student learning and that well-implemented assessment processes provide positive prospects for meaningful learning, whereas flawed assessment risks leading student learning in unproductive directions. There has been a wide range of research activity and projects in HE influenced explicitly or implicitly by AfL principles over the last 20 years or so. There is also a rapidly expanding related literature, including various book length treatments (e.g. Carless, 2015; Knight, 1995; Sambell, McDowell, & Montgomery, 2013). This range of evidence and reports of practice provide tentative indication that AfL has reached a stage of maturity. In the terminology

D. Carless (✉)
Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong
e-mail: dcarless@hku.hk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_1

of educational change, it seems to have become institutionalized (Fullan, 2001) in that the practices are embedded within the pedagogy of a wide number of teachers in HE.

It is, however, difficult to gauge precisely the extent to which interest in AfL has led to widespread implementation at course levels (Boud, 2014). Despite the arguments for AfL, there remain powerful imperatives surrounding summative assessment and grading which risk overpowering learning-oriented approaches to assessment. These include: fairness and reliability of grading, grade inflation and honours classification and student malpractices, such as plagiarism or other forms of cheating. Middle and senior managers are usually preoccupied with quality assurance aspects of assessment, including preventing and managing malpractice rather than encouraging diverse or innovative approaches to assessment (Meyer et al., 2010).

The main aims of this opening chapter are to make a case for scaling up AfL, discuss the extent of implementation of AfL over time and across geographical locations and frame the collection by charting some key issues in relation to the potentials and challenges for scaling up of AfL. I develop the arguments in the following stages. First, I define what AfL is and synthesize its main implementation strategies. Next, I propose a framework for scaling up and propose key rationales for the scaling up of AfL. I analyse the breadth and depth of AfL implementation through a discussion of AfL research and development in different contexts, its modest but increasing impact on university assessment policies and in relation to the key issue of feedback. I conclude with a discussion of drivers and factors impinging on the scaling up of AfL and analyze some of the barriers arising.

AfL and Its Main Strategies

At the outset it is important to define what AfL is. There are various terminologies associated with approaches to assessment focused on enhancing student learning: formative assessment, assessment for learning, assessment as learning and learning-oriented assessment. The term AfL came into common parlance in the early 2000s to emphasize the purpose for which assessment is carried out in contrast to formative and summative assessment which relate to the functions which are served (Wiliam, 2011). In the HE literature, AfL is often not defined explicitly. Accordingly, I adopt the following definition from the literature related to schooling:

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning (Black et al., 2004, p. 10)

In relation to the school sector, the King's College London group led by Black and Wiliam did much to promote and encourage scaling up of AfL practices, and I turn now to a discussion of the main AfL strategies for schooling and in HE.

AfL strategies in relation to schooling seem somewhat more clearly defined and agreed upon than those in HE. The following list of five key strategies (Wiliam & Thompson, 2008) is relatively authoritative:

1. Clarifying learning intentions and success criteria
2. Engineering effective questioning and classroom discussions
3. Providing feedback that moves learners forward
4. Activating students as owners of learning
5. Activating students as instructional resources for one another

The AfL in HE literature carries some resonance with these strategies. Only the second of these is under-explored with a need for further investigation of apprenticing undergraduates into academic discourse through sensitive challenge and induction into academic practices (Black & McCormick, 2010). The relevant HE literature to date seems to lack a definitive statement of key AfL strategies, and in order to trace the development of the ideas, I first discuss three key perspectives to help me work towards a synthesis.

In a review of conditions under which assessment supports student learning, Gibbs (2006) elaborates a number of issues in relation to the design of assessment and the development of effective feedback processes. He suggests that assessment tasks should capture student time and effort, distribute this effort evenly over the duration of a course and engage students in productive learning activity. He considers a number of issues in relation to feedback, including its frequency and timeliness, linkages with assessment criteria and the impact of feedback on student future learning (Gibbs, 2006).

In a vision of assessment reform, Boud and associates (2010) make a number of points relevant to the current discussion: assessment should engage students in learning that is productive, feedback needs to be used actively to improve student learning, students and teachers should become responsible partners in learning and assessment and AfL should be placed at the centre of course design.

Sambell, McDowell and Montgomery (2013) suggest six features of AfL: appropriate balance of summative and formative assessment, authentic complex assessment tasks, self-evaluation activities, rich in informal feedback, rich in formal feedback and offering confidence-building opportunities and practice.

Synthesizing these works and other relevant literature, Table 1.1 summarizes on the left what I see as the main AfL strategies and on the right-hand side of the table suggests some means of operationalizing them. These examples of implementation processes are illustrative and not intended to be exhaustive. Technology can act as an enabler of AfL, and the fourth example in each category is an online or technology-related strategy.

The first strategy is productive assessment task design: the development of tasks which carry potential to stimulate meaningful learning processes amongst students. This includes designing tasks which encourage students to sustain deep approaches to learning aligned with the learning outcomes sought. This kind of assessment may often mirror real-life elements of the discipline. For example, Glofcheski

Table 1.1 Synthesis of main AfL strategies and processes

AfL strategies	Illustrative implementation processes
Productive assessment task design	Alignment with intended learning outcomes
	Authentic assessment
	Integrated and coherent assessment
	Collaborative writing through wikis
Effective feedback processes	Integrated guidance and feedback
	Students generating and seeking feedback
	Closing feedback loops
	Technology-enabled feedback dialogues
Developing student understanding of the nature of quality	Students generating and/or decoding criteria
	Applying criteria
	Analysing and discussing exemplars
	Online dialogue about exemplars
Students practising making judgments	Providing peer feedback
	Receiving peer feedback
	Self-monitoring work in progress
	Online facilitation of peer interaction

(this volume) discusses assessment in law focused on authentic assessment which facilitates a wide range of learning outcomes relevant to future professional life.

The second strategy is represented by the development of effective processes as a central factor in curriculum and assessment planning. A trend in recent work is to examine how feedback designs can promote student uptake of feedback (e.g. Boud & Molloy, 2013). This can involve, for example, the integration of guidance and feedback and emphasis on students seeking, generating and using feedback. Moscrop and Beaumont (this volume) illustrate the potential of technology, such as a learning coach via an intelligent tutoring system to enhance feedback dialogues and scaffold student self-regulated learning.

The third strategy relates to student understanding of the nature of quality work and its relationship with transparent criteria or rubrics. A key role of the teacher is to support students in developing capacities to discern quality and make sound evaluative judgments (Sadler, 2010). Dialogue around exemplars, for example, contributes to the development of student expertise in making judgments. Students' enthusiasm for exemplars is a key sub-theme in the analysis of students' experiences of assessment (Carless, this volume).

The fourth strategy follows from the third in that it focuses on making judgments about quality in relation to the work of a peer or one's own work in progress. Giving peer feedback is often even more beneficial than receiving comments because it is more cognitively engaging: involving higher-order processes, such as diagnosing problems and suggesting solutions (Nicol, Thomson, & Breslin, 2014). Peer review processes also help students to calibrate their own judgments and enhance their own self-evaluative capacities.

AfL involves partnership between teachers and learners. Assessment task design is largely in the hands of the teacher but is interpreted, and responded to, by students. Effective feedback processes can be facilitated by teachers but it is only students who can act on feedback. Understanding quality and making judgments also place the student at the centre of their learning with the teacher playing an important guiding and facilitating role. The central role of students in AfL is an undercurrent throughout the volume and a particular focus of the research reported in the chapters by Carless and Jessop.

Scaling Up Educational Change

In the terminology of educational change, what are key issues in relation to scaling up of AfL? A starting point for scale relates to quantity: the number of teachers and institutions which are carrying out a specific pedagogic strategy or innovation. A more comprehensive conceptualization of scale comprises four interrelated dimensions: spread, depth, sustainability and shifts in ownership (Coburn, 2003). Spread involves implementation of pedagogic innovation at additional sites or in more groups within existing sites. Depth involves refining pedagogic practice in deep and meaningful ways that influence student learning. Depth also needs to impact the beliefs of teachers and their underlying assumptions about pedagogy (Kezar, 2011). Sustainability relates to longevity, requiring policy and infrastructure systems in place to support continued improvement in pedagogy over time with potential transfers of ownership to encourage continuous refinement and further scaling up (Coburn, 2003).

In relation to AfL in schools, Wiliam (2007) suggests that teacher communities of practice are a productive strategy for scaling up. He has developed five scaling up principles which carry potential wider relevance (Leahy & Wiliam, 2012; Wiliam, 2007). First, gradualism in that generally teachers take small incremental steps in implementing change. Second, flexibility is required in order to facilitate teacher adjustment to techniques to make them work in their context. Third, there needs to be a degree of choice so as to enable teachers to select which AfL techniques they are going to implement. Fourth, a certain amount of accountability is desirable so that teachers are accountable to the teacher learning community for implementing changes. Fifth, support occurs through the building of trust amongst participants in the learning community (Leahy & Wiliam, 2012; Wiliam, 2007). All of these issues seem to resonate with HE, including the fourth principle of accountability carrying additional quality assurance dimensions discussed later in the chapter.

Why Do We Need to Scale Up AfL?

Two key elements of a case for scaling up AfL are dissatisfaction with existing assessment practices and research evidence suggesting the power of well-implemented AfL strategies. I discuss these in turn below.

First, there has been considerable airing over the last 20 years or so of dissatisfaction from both teachers and students about assessment practices. From the staff perspective, assessment is sometimes seen as a pernicious influence on the learning process, tending to direct students towards grades and instrumentalism rather than a wider learning experience, emphasizing summative assessment to the detriment of more formative approaches, and failing to encourage the higher-order learning outcomes to which university education aspires, and seen as time-consuming and implicated in unwelcome auditing and quality assurance procedures.

From a student perspective, there is plenty of evidence from institutional surveys both in the UK and other parts of the world that assessment is one of the least satisfying aspects of their student experience. Students' concerns include fairness; lack of clarity about what they are expected to achieve; disappointment if marks do not meet their expectations; emotional challenges, such as pressure, anxiety and discouraging experiences; and concerns about feedback processes, particularly their timeliness and usefulness.

Whilst it cannot be assumed that all of these staff and students' concerns are fully justified and reasonable, they are suggestive of considerable misgivings about aspects of how assessment is currently organized and implemented. These challenges are compounded by relatively low assessment literacy of staff and students (Norton, Norton, & Shannon, 2013; Price, Rust, Donovan, and Handley 2012). A recent paper (Bevitt, 2015) sums up well a number of imperatives for assessment change: to enhance the student experience, to harness technological developments, to encourage AfL and to respond to the needs of increasingly diverse student populations in the context of massified HE.

Second, there is a range of research evidence which indicates that approaches associated with AfL are powerful means of enhancing student learning. The landmark Black and Wiliam (1998) research synthesis captured attention by accumulating evidence that well-implemented formative assessment improves student performance in schooling and in HE across a variety of contexts and settings.

The influential meta-analysis of meta-analyses (Hattie, 2009) indicates the visible learning attributed to practices congruent with AfL: student self-evaluation and metacognitive strategies, formative evaluation and feedback and collaborative learning through reciprocal teaching. Of the 138 practices reviewed by Hattie, many of the most effective practices resonate with AfL. Self-report grades (ranked no. 1) and metacognitive strategies (no. 13) involve students making judgments, providing formative evaluation (no. 3) and feedback (no. 10) are closely aligned with effective feedback processes and reciprocal teaching (no. 9) shares facets with peer review and peer feedback.

To sum up, dissatisfaction with existing assessment practices and the research evidence in favour of AfL strategies provide a rationale for in-depth, sustainable attempts at encouraging and supporting more widespread implementation of AfL practices.

Breadth and Depth of Influence of AfL

The next section attempts the difficult task of gauging the breadth and depth of implementation of AfL. I review three possible indicators: geographical spread of research activity, influence on university assessment policies and the expansion of academic attention to the key AfL strategy of effective feedback processes.

Geographical Spread

A key scaling up factor is geographical spread, the extent to which there appears to be deep and sustained AfL activity in multiple significant settings. As it is impossible to gauge how teaching, learning and assessment are implemented around the world, I discuss research and development activity as one of the indicators of scaling up of AfL. There are a number of examples of positive sustained implementation of practices congruent with AfL in selected international settings.

The UK seems to be a leading context for the implementation of AfL concepts. For example, Oxford Brookes University and Northumbria University both achieved prestigious Centre for Excellence in Teaching and Learning funding for sustained good work related to AfL. Colleagues from both of these institutions have been prominent figures in the AfL-related literature since the 1990s: Margaret Price and Chris Rust at Brookes and Liz McDowell and Kay Sambell at Northumbria. Research from the UK is also well-represented in this volume (see chapters by Jessop, Pitt, Moscrop & Beaumont).

Turning to Australia, there appears to be plenty of research activity in relation to the strategies discussed in Table 1.1. Two eminent scholars, David Boud and Royce Sadler have produced sustained research on AfL in HE over a period of more than 30 years. Boud's contributions include his early work on self-assessment (Boud, 1995), his equally influential championing of sustainable assessment for lifelong learning (Boud, 2000) and his analysis of the implications for assessment of increasing focus on learning outcomes and standards-based approaches (Boud, this volume). Sadler's work includes his seminal paper on formative assessment (Sadler, 1989) and his analysis of feedback in relation to the development of student understanding of quality (Sadler, 2010).

There is also a continental European school of AfL research, a significant strand of it stimulated by sustained work involving Filip Dochy and his collaborators (e.g. Dochy, Segers, Gijbels, & Struyven, 2007; Dochy, Segers, & Sluijsmans, 1999). A parallel European development is an expanding literature related to the key AfL concepts introduced in Table 1.1. Assessment design issues are investigated, for example, in relation to the assessment of professional competencies (van der Vleuten & Schuwirth, 2005) and a framework for quality assessment in competence-based education (Baartman, Bastiaens, Kirschner, & van der Vleuten, 2007). Analysis of effective feedback processes includes the interactive tutoring feedback

model (Narciss, this volume) and feedback in online environments (Alvarez, Espasa, & Guasch, 2012). Developing student understanding of the nature of quality can be facilitated by well-designed and well-used rubrics (e.g. Jonsson & Panadero, this volume). Peer assessment is one of the most favoured means of enabling students to practise making judgments (e.g. Strijbos, Narciss, & Dunnebier, 2010).

In sum, this brief geographically based synopsis is suggestive of considerable research and development work congruent with AfL in certain key settings, but little is known about various other parts of the world. In China, for example, the long history of competitive examinations represents a challenge to AfL, although there are some nascent initiatives to introduce a more formative orientation to assessment at the university level (Chen, Kettle, Klenowski, & May, 2013).

Assessment Policy Documents

A further indicator of how deeply AfL might be embedded within the fabric of HE pedagogy arises from an examination of university assessment policy documents which are generally readily available on university websites. An earlier synopsis of assessment policies in the UK and Australia (Boud, 2007) found that quality assurance aspects of assessment were predominant. A study of assessment policies in New Zealand (Meyer et al., 2010) reinforces this picture, suggesting that discussion of AfL is largely lacking in institutional policy documentation. In order to scrutinize these findings further, I have undertaken some preliminary analysis of assessment policy documents at a number of universities.

At King's College London, the home of formative assessment research in schooling, the assessment policy document focuses particularly on marking frameworks: different models of marking, including procedures for blind double marking and conversion of marks from studying abroad. The King's feedback policy emphasizes timeliness and the return of feedback within 4 weeks. Potentially more illuminating from an AfL point of view is a parallel King's education strategy (2013–2016) which highlights student dissatisfaction with assessment and feedback. The document sets out an aim to reduce the burden of assessment through a more considered, flexible AfL regime. It outlines a number of assessment reviews being conducted in various disciplines, including mapping the student assessment journey, reviewing assessment at programme levels, ensuring that assessments are calibrated to encourage lessons from one assessment to be applied to the next and considering new forms of synoptic assessment above and beyond modules.

The University of Melbourne assessment procedure document focuses on 17 procedures including the operation of Boards of Examiners, compliance, penalties, release of results, supplementary assessment and appeals. There is an additional Coursework Assessment Policy, suggesting that assessment should be balanced so as to provide diagnostic, timely and meaningful formative feedback, as well as summative judgments. There is some reference to feedback which is viewed as involving comments indicating to students how they have performed against

assessment criteria and how they can improve their performance. Discussion of feedback includes warnings against students communicating with examiners and how students may request access to examination scripts.

The University of Bristol assessment policy documentation contains both procedural and AfL elements. The main regulations and code of practice document focuses on progression, awards and the conduct of assessment in a roughly similar way to the King's and Melbourne procedures. More pertinent from an AfL perspective are the institutional principles for assessment and feedback which highlight the promotion of effective student learning. They include a principle that all assessment is for learning and suggestions for: a range of assessment methods, assessment mapping. the imaginative design of assessment and feedback, research-informed practices and encouragement for staff to improve their assessment and feedback literacy. Assessment and feedback are viewed as a conversation which provides students with opportunities to engage in continuing dialogues about their learning.

This brief and selective review of assessment policies in three major universities provides tentative support for the positions of Boud (2007) and Meyer et al. (2010) that university assessment policies generally emphasize rules and procedures and an emphasis on quality assurance aspects of assessment, such as grading and moderation procedures. There is evidence in some of the policy statements, particularly at the University of Bristol, of thoughtful treatment of AfL elements, such as feedback. I turn next to examine how and why feedback has generated considerable recent attention.

An AfL Priority Area: The Case of Feedback

As a case of scaling up of research and development interest in an AfL area, I now analyse how in the space of 20 years feedback processes for students have gone from being a neglected research niche to a relatively high-profile topic. In the late 1990s, feedback was an under-researched area (Higgins, Hartley, & Skelton, 2002), yet there has been a remarkable growth in articles focused on feedback in the last 15 years or so. For example, in the main journal of our subfield, *Assessment and Evaluation in Higher Education*, during the 10-year period from 1996 to 2005, there were 11 articles with feedback in the title (3 % of the total articles), whereas from 2006 to 2015 there were 65 (representing 11 % of the articles). Of these 76 articles, 38 appeared in the 3-year period (2013–2015).

Probably the most urgent and persuasive driver for the expansion of feedback research is the consistent finding in National Student Surveys in England and Wales that feedback is perceived by students as one of the least satisfactory elements of their university experience (HEFCE, 2014; Williams & Kane, 2009). Jessop (this volume), for example, reports students' perceptions of episodic and haphazard feedback not connected to the next task or across modules. Student misgivings about feedback are also reported in other jurisdictions, Australia (e.g. ACER, 2010)

and Hong Kong (Carless, 2006), so it seems to represent a widespread challenge. In the UK, the 'feedback issue' generated considerable media attention and was firmly on the radar of university senior management (Williams & Kane, 2009). This attention generated a host of initiatives designed to tackle the perceived problems. Many of these, such as focusing on feedback turnaround times (exemplified by the King's College feedback stipulations alluded to above), tend to be seen as 'quick fixes' rather than more considered to be scholarly attempts at reforming feedback processes.

A repercussion was that numerous funded projects on feedback were spawned. A well-known example is REAP (Re-engineering Assessment Practices) which was well-anchored conceptually in relation to the aspiration to promote self-regulated learning (Nicol & Macfarlane-Dick, 2006) and in relation to AfL strategies, e.g. developing student understanding of quality and students making judgments (Nicol, Thomson, & Breslin, 2014). A scaling up element of REAP involved linkages to strategic institutional developments. First, a new university policy for assessment and feedback consolidated REAP principles at an institutional level (Nicol & Draper, 2009). This is important because it moves from the potentially ephemeral work of a project to a more long-term legacy. The second institution-wide initiative was a 'feedback is a dialogue campaign' in association with the Student Union. This brought in different stakeholders through separate campaigns for staff and students on feedback principles and practices supported by leaflets and posters of advice. The extent of long-term strategic commitment to REAP ideas from senior management was, however, a moot point. Once funding dries up or key personnel depart, it is often difficult to sustain project legacies representing a barrier to scaling up.

A potentially productive mini-trend in relation to scaling up good practices in feedback is that a number of British universities have now instigated feedback awards to reward and encourage good practices. Some of these feedback awards are student-initiated or developed. The processes of these awards can surface and celebrate good practice (cf. Hounsell & Zou, this volume). Award schemes stimulate attention to a specific topic, providing rewards and incentives which can encourage the scaling up of good practice.

A further repercussion of increased attention to feedback processes is the expansion of previously under-explored research sub-strands, such as students' affective responses to feedback. The complex interplay between emotions and feedback is analyzed in two chapters in this volume. On the basis of phenomenographic research, Pitt (this volume) uncovered a range of emotional reactions to feedback, some of which were maladaptive and hindered action on feedback. Through synthesizing a wide range of literature, Rowe (this volume) brings out some of the complexities of emotions in relation to feedback and illustrates how a deeper understanding of emotions can play a role in the scaling up of AfL practices. The emotional legacy of feedback is also discussed in the chapter by Ajjawi and colleagues.

To sum up, I am suggesting that the scaling up of attention to the AfL element of feedback arose largely from student survey data which indicated dissatisfaction.

This caught the attention of various stakeholders, including senior management, middle managers and staff of various levels leading to a wide range of research and development initiatives.

Drivers and Challenges in Scaling Up AfL

I now discuss some drivers which might encourage wider and deeper implementation of AfL. I also consider some of the facilitating and inhibiting factors impacting on possible scaling up.

An inference I draw from the case of feedback is that a potential driver for assessment reform is evidence from quality assurance and quality enhancement processes. Although the continuous auditing agenda has its drawbacks, it may bring to light practices which are unpopular with students or do not stand up to quality assurance scrutiny. For example, programme reviews, stakeholder feedback or external examiner reports may identify suboptimal practices, and these can provide opportunities for middle management overseeing teaching and learning to follow up with action plans. Embedding the improvement of assessment and feedback within quality assurance processes is a key feature of the chapter by Jessop (this volume).

Leadership, especially at middle management levels such as Deans, Associate Deans, Heads of Department and programme leaders, is a potential lever for assessment change. Commitment from leaders to an AfL agenda might support the scaling up of related practices. Staff involvement is rarely sustained without visible support from institutional leaders. Middle management might develop strategies to encourage AfL, including prioritization of resources, rewards and incentives. Congruent with the expansion of teaching award schemes to include feedback awards, there could be similar additions of awards for best AfL practice or best assessment innovation.

There is a danger that adjustments arising from quality assurance or the priorities of academic leaders may reflect conservative approaches rather than AfL. It is important for institutions to develop climates where innovation in assessment is encouraged. The role of trust, or at least minimizing distrust, is a central issue in the encouragement of assessment reform (Carless, 2009). Trust would probably be more forthcoming if there were higher levels of staff assessment literacy and I turn to this issue next.

The development of staff assessment literacy carries potential to contribute to the scaling up of AfL. Building on the AfL strategies summarized earlier in Table 1.1, I suggest that teacher AfL literacy involves a sound grasp of principles and practices in assessment task design, effective feedback designs and developing student capacities in understanding and applying criteria through making judgments. The development of assessment literacy resonates with the scaling up concept of teachers being better able to respond to contextual challenges when they possess deep understandings of pedagogical principles. Assessment literacy would enhance teachers' capacities to adapt AfL practices to the needs of their students in particular

institutional and disciplinary settings. Discipline-specific implementation of AfL practices is well-represented in this volume, including dentistry (Bridges et al.), health professions (Ajjawi et al.) and law (Glofcheski).

Professional development activities are an obvious starting point for enhancing staff assessment literacy. Seminars and sharing sessions can be useful in exemplifying and disseminating good assessment practice which in the hands of enthusiasts may play a role in scaling up. Encouraging good practice is admirable, but what about reducing bad practice in assessment? Perhaps the most promising strategy to tackle this difficult issue is leadership, allied with a judicious balance of pressure and support. As suggested earlier, quality assurance might be used as a lever to tackle and reduce unsophisticated assessment practices. Mentoring and peer review of assessment practice is also worth scaling up.

An alternative or possibly complementary means of developing assessment literacy is through communities of practice in which 'accounts of practice' are surfaced and shared (Hounsell & Zou, this volume). Such activities resonate with the earlier discussion of communities of practice in schools as a means of enhancing ownership of AfL strategies. Supportive professional communities of practice facilitate collegial support and promote sustainability (Coburn, 2003). In such ways, AfL strategies could be shared, developed and refined amongst groups of colleagues.

Another form of a community of practice is represented by programme teams. Embedding AfL at programme levels is a useful strategy for scaling up. Programme-wide initiatives are a site for embedded professional development of university teachers in that they involve a range of colleagues discussing practice in context. For example, the TESTA (Transforming the Experience of Students Through Assessment) project methodology promotes a programme enhancement approach to assessment through the careful triangulation of data from the Assessment Experience Questionnaire and focus group interviews (Jessop, this volume).

The use of technology to enable innovative approaches to assessment and feedback represents a further possible driver for the scaling up of AfL practice. For this potential to be fulfilled there may need to be synergies between assessment literacy, technological literacy and professional development. Related issues are taken up by other chapters in this volume. Moscrop and Beaumont illustrate how technology can facilitate dialogic feedback cycles as a means of encouraging student uptake of feedback. The chapter by Dawson and Henderson takes both a critical perspective on technology-enabled AfL and suggests some possibilities for scaling up both practice and the related research base.

The above discussion is suggestive of some avenues for future research and development activity, extending or going beyond some of the discussion in this volume. What are effective ways of developing staff assessment and feedback literacy? How does staff assessment literacy help to seed student assessment literacy? What forms of leadership and support are most conducive to developing AfL? Under what circumstances is quality assurance a barrier to AfL and when might it support its further development? What AfL practices can be scaled up to operate effectively with large classes and multiple tutors, and how can technology effectively enable these processes?

The volume is arranged in four parts. Part I, 'Enabling Assessment Change', contains this chapter and the contributions of Boud, Hounsell and Zou and Jessop. Part II focuses on AfL strategies and implementation with chapters by Glofcheski, Bridges and colleagues, Jonsson and Panadero and Carless. Part III is entitled 'Feedback for Learning' and contains chapters by Ajjawi and colleagues, Pitt, Rowe and Narciss. The final section, 'Using Technology to Facilitate AfL', involves chapters by Moscrop and Beaumont and Dawson and Henderson. Most chapters explicitly address the scaling up theme, whereas others focus more on the specific AfL issue which their chapter addresses.

References

Alvarez, I., Espasa, A., & Guasch, T. (2012). The value of feedback in improving collaborative writing assignments in an online environment. *Studies in Higher Education, 37*(4), 387–400.

Australian Council for Educational Research (ACER). (2010). *Doing more for learning: Enhancing engagement and outcomes. Australasian student engagement report*. Camberwell, Australia: Australian Council for Educational Research.

Baartman, L., Bastiaens, T., Kirschner, P., & van der Vleuten, C. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review, 2*(2), 114–129.

Bevitt, S. (2015). Assessment innovation and student experience: A new assessment challenge and call for a multi-perspective approach to assessment research. *Assessment & Evaluation in Higher Education, 40*(1), 103–119.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8–21.

Black, P., & McCormick, R. (2010). Reflections and new directions. *Assessment & Evaluation in Higher Education, 35*(5), 493–499.

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

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

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

Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 181–197). London: Routledge.

Boud, D. (2014). Shifting views of assessment: From secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 13–31). Edinburgh, Scotland: Edinburgh University Press.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Retrieved from https://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf

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.

Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79–89.

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.

Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education, 38*(7), 831–846.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12.

Dochy, F., Segers, M., Gijbels, D., & Struyven, K. (2007). Assessment engineering: Breaking down barriers between teaching and learning, and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 87–100). London: Routledge.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

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

Hattie, J. (2009). *Visible learning*. London: Routledge.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education, 27*(1), 53–64.

Higher Education Funding Council for England (HEFCE). (2014). *National student survey*. Retrieved from http://www.hefce.ac.uk/whatwedo/lt/publicinfo/nss/

Kezar, A. (2011). What is the best way to achieve broader reach of improved practices in education? *Innovative Higher Education, 36*(4), 235–247.

Knight, P. (1995). *Assessment for learning in higher education*. London: Kogan Page.

Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 49–71). Thousand Oaks, CA: SAGE.

Meyer, L., Davidson, S., McKenzie, L., Rees, M., Fletcher, R., & Johnston, P. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly, 64*(3), 331–350.

Nicol, D., & Draper, S. (2009). A blueprint for transformational organizational change in higher education: REAP as a case study. In T. Mayes, D. Morrison, H. Mellar, P. Bullen, & M. Oliver (Eds.), *Transforming higher education through technology-enhanced learning* (pp. 191–207). York, UK: Higher Education Academy.

Nicol, D., & 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.

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.

Norton, L., Norton, B., & Shannon, L. (2013). Revitalising assessment design: What is holding new lecturers back? *Higher Education, 66*(2), 233–251.

Price, M., Rust, C., O'Donovan, B., & Handley, K. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford: Oxford Centre for Staff and Learning Development.

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

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550.

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.

Strijbos, J., Narciss, S., & Dunnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291–303.

Van der Vleuten, C., & Schuwirth, L. (2005). Assessing professional competence: From methods to programmes. *Medical Education, 39*(3), 309–317.

Wiliam, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 183–206). Bloomington, IN: Solution Tree.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3–14.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–84). New York: Lawrence Erlbaum Associates.

Williams, J., & Kane, D. (2009). Assessment and feedback: Institutional experiences of student feedback, 1996–2007. *Higher Education Quarterly, 63*(3), 264–286.

Chapter 2
Standards-Based Assessment for an Era of Increasing Transparency

David Boud

Abstract The macro policy context for assessment in higher education has changed to focus on explicit standards and learning outcomes. While different countries and institutions are at different stages of the process of reorienting assessment to become more directly standards based, the implications for assessment and learning are substantial. Assessment becomes transparent in multiple ways: it is possible to report on what students can actually do, rather than how they stand vis-à-vis others (norm-referenced assessment). Outcomes can be compared across courses, institutions and countries. Students can progressively track their achievement of outcomes when these are explicit. Assessment becomes open to scrutiny as never before as standards-based assessment requires a scaling up of transparency. This chapter explores the new context of assessment and what opportunities it affords. It considers the implications for assessment practice and identifies ways in which the new framework directly conflicts with familiar taken-for-granted assessment practices, such as conventional grading. It concludes by pointing to new opportunities offered and what needs to be done to realize them.

Introduction

The global refocusing of higher education on outcomes is prompting considerable curriculum innovation and the rethinking of teaching and learning practices. It has changed the discourse around degree programmes to foreground explicit learning outcomes, the development of programme-wide attributes and threshold standards. However, with a few exceptions it has had less impact on assessment whether for certification or learning purposes. This is surprising as the implications for assessment are profound. In a standards-based approach, what is central is that standards can be assured and appropriate criteria addressed by all students. Such an approach implies that it needs to be made clear what students should be able to do as a result of particular episodes of study (the learning outcomes) and that assessment

D. Boud (✉)
Deakin University, Burwood, VIC, Australia
e-mail: david.boud@deakin.edu.au

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_2

demonstrates and reports that students have met these requirements. This can be at odds with conventional marking and grading systems that aggregate information by subjects or course units and elide differences of outcome. These systems allowed students effectively to pass by overachieving in some areas and underachieving in others without ensuring that basic standards have been met. Baik and James (2014) started to identify some of the issues that arise for assessment in the context of learning outcomes. This chapter takes some of these further and works through the necessary implications of what this involves.

As participation in higher education increases throughout the world, and as students are expected to contribute increasingly to the cost of their studies, new expectations are arising about what universities should do and how they should do it. In particular, transparency of purpose and process is being required. If students choose to enrol in a given programme or course unit, it should be of the highest quality and do what it claims to do. This has been manifest by a greater focus on student satisfaction surveys, on explicit statements of learning outcomes and on the use of quality assurance processes. Shortcomings are no longer apparent to a limited group of specialists, but are paraded in national media. International league tables are much consulted for research ratings, which provide an invalid account of the quality of programmes, but ratings which focus on students' learning experiences are already available in many countries and are being used for national league tables.

Publicly available learning outcomes are a necessary prerequisite for transparency. These take the form of statements about what students will be able to do as a result of successful completion of their studies. They are more explicit than previous lists of subjects or syllabuses and enable judgments to be made of the institution and programmes as well as of their graduates. Assessment judgments about students only have meaning in the context of what a course seeks to do and the extent to which a student has met its requirements. Lists of course content or a traditional syllabus does not sufficiently communicate what completion of a programme or course unit signifies. The adoption of learning outcomes has now become a widespread practice in higher education. These describe what a student should be able to do as a result of successful completion of a period of study. While they were introduced at the level of the unit of study, programme learning outcomes describing entire degrees are becoming more common.

While the assessment of students was not the first area to be the focus of moves towards transparency, it has now become a key consideration. After all, if institutions are to be judged on the quality of their graduates, it is the assessment system that guarantees this quality, and it must be fit for this purpose. And the assessment system is what focuses students' attention on that which is most important, whatever other claims might be made about a programme.

Scaling up assessment or scaling up good practice in assessment is not only a matter of dealing with increasing student numbers, but addressing the wider context of where a programme is situated within a discipline, an institution, a country and globally. It must be located within a discussion of the macro changes that are occurring globally. Assessment has conventionally been seen within the context of an individual course unit in which assessment decisions commonly take place.

Marks and grades may for convenience be aggregated across a programme, but such a process says little about how a student meets the outcomes for a programme as a whole. Scaling up assessment at the course level means looking at the implications for assessment decisions on a wider front than the course unit and examining what the implications are when the learning outcomes for a programme are considered in a global context.

This chapter focuses on the tensions created for institutional assessment policy and disciplinary practice of a standards-based approach. It examines why and how conventional assessment assumptions and practices need to change so that assessment becomes more transparent and defensible in the context of global scrutiny of curriculum provision. It questions whether an attachment to conventional assessment practices is conceptually compatible with a standards-based approach to higher education or perhaps acts to undermine it. It suggests that a new focus on assessment is needed that places programme outcomes as central and the fostering of long term learning as a key feature of programmes. In addition to discussing what requirements assessment practices now need to meet, it also considers how assessment can be used in the fostering of student judgment, how students can be involved in the curation of outcomes and the meaningful portrayals of achievements for different audiences.

Academic Standards in a Global Context

The iconic discourse of education focuses on 'academic standards'. These have been taken-for-granted as the *sine qua non* of what it means to be educated. High academic standards are unquestioningly a good thing that students and educational institutions should strive for, but it is less clear to what they specifically refer. Are they general aspirations for excellence or do they mean something in particular? In the slowly moving shifts towards transparency, we are seeing academic standards now less as positional rhetoric—'we are committed to maintaining high academic standards'. Rather, they are something we plan to meet—'are our assessment practices up to standard?'—and intend to judge our students by: 'have you met the standard yet?' The international discourse of higher education has become that of learning outcomes which map on to academic standards. They are not rhetorical constructions but a tangible element of the design and planning process for courses represented in concrete terms.

Academic standards are generated from many sources. They are intrinsic to the structure of disciplinary knowledge. The nature of disciplines and the frameworks they have built to hold knowledge represent a key reference point. They are supplemented by the consensus views of experts. Most competency standards for professions or occupations are generated from consultation with those who practise in a particular area and who can judge what constitutes competence or capability in that area. Less commonly, standards can arise from empirical analysis of actual professional practices—what can it be observed that practitioners do when they perform in the domain of their expertise?

The New Landscapes of Academic Standards

Interest in standards is widening and transcending disciplines, professions and indeed educational institutions. It has taken on a dual focus. Firstly, assurance processes that enable standards to be monitored have been introduced, and secondly, alignment processes to generate common standards across jurisdictions have been mooted. While there is some overlap in the organizations responsible for these, the former mostly occur at a national level through national qualifications frameworks and national quality assurance agencies (e.g. the Tertiary Education Quality Standards Agency in Australia, the Quality Assurance Agency in the UK). Each country or jurisdiction has its own framework of qualifications and standards and agencies that ensure that standards of programmes are maintained across institutions.

The second focus mostly takes place in cross-jurisdictional spaces through accrediting agencies, international organizations and specific cross-country projects. There are international accrediting agencies for particular disciplines that operate worldwide. The longest standing is the *Association to Advance Collegiate Schools of Business* (AACSB). It provides 'internationally recognized, specialized accreditation for business and accounting programmes at the bachelor's, master's, and doctoral level' (http://www.aacsb.edu/accreditation). University business schools seek accreditation from the AACSB as it provides a well-established and high-status benchmark for quality through its rigorous assessments of programmes.

Other governmental and quasi-governmental bodies are also increasing players on the international scene. The project on 'Tuning Educational Structures in Europe' sponsored by the European Commission:

> seeks to identify generic and subject-specific competences for first-cycle degrees within the European Higher Education Area. .... These reference points, which address workload as well as learning outcomes, support the objectives of the Bologna Process to establish compatibility of qualifications across Europe.
>
> Tuning focuses not on educational systems, but on educational structures with emphasis on the subject area level, that is the content of studies. Whereas educational systems are primarily the responsibility of governments, educational structures and content are that of higher education institutions and their academic staff. (http://www.unideusto.org/tuningeu/)

In recent years, the OECD (2016) has been carrying out a feasibility study for the Assessment of Higher Education Learning Outcomes (AHELO). The purpose is to judge if it is practically and scientifically feasible to assess what students in higher education *know and can do* upon graduation. AHELO aims to be a direct evaluation of student performance at the global level, valid across diverse cultures, languages and different types of institutions. This approach has been trialled in two disciplines, economics and engineering, but has been stalled further by lack of agreement by some key governments. The OECD claim is that:

> a full scale AHELO would be a 'low stakes' voluntary international comparative assessment designed to provide higher education institutions with feedback on

> the learning outcomes of their students which they can use to foster improvement in student learning outcomes. (http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdsahelo.htm)

However, institutions that have high success in attracting students worldwide are vulnerable to having their claims exposed by empirical investigation.

An important part of a standards framework is the learning outcomes established in various countries. One of the first of these was the development of the UK Subject Benchmark Statements that:

> set out expectations about standards of degrees in a range of subject areas. They describe what gives a discipline its coherence and identity, and define what can be expected of a graduate in terms of the abilities and skills needed to develop understanding or competence in the subject. (http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements)

Originally developed by the Higher Education Academy, they now form part of the UK Quality Code for Higher Education (QAA, 2016) to which all higher education institutions are expected to subscribe.

A substantial development on academic standards for the disciplines was undertaken by the Australian Learning and Teaching Council in 2010–2011 (ALTC, 2011). Threshold learning outcomes were established through lengthy consultations within each discipline or discipline grouping. These were standards that indicate the threshold or minimum level expected of all graduates of an Australian programme in the given discipline. These have been used by the national regulatory body (TEQSA) as an indicator that can form part of the statutory requirements that all higher education institutions need to meet to gain recognition of their awards. The Australian government has established a Higher Education Standards Framework for Students. These standards represent the minimum acceptable requirements for the provision of higher education in or from Australia by higher education providers registered under the TEQSA Act 2011. The statement states that:

> The Standards also serve other broader purposes in Australian higher education including:
>
> 1. an articulation of the expectations for provision of higher education in Australia as:
>
>   a. a guide to the quality of educational experiences that students should expect
>   b. a reference for international comparisons of the provision of higher education
>   c. a reference for other interested parties, and
>
> 2. a model framework which higher education providers can themselves apply for the internal monitoring, quality assurance and quality improvement of their higher education activities. (Higher Education Standards Framework 2015 Explanatory statement)

Relevant to our discussion, it requires that:

> Methods of assessment are consistent with the learning outcomes being assessed, are capable of confirming that all specified learning outcomes are achieved and that grades awarded reflect the level of student attainment (Higher Education Standards Framework 2015, 1.5.7)

The Changing Ground of Assessment: Focus on Multiple Purposes

Before turning to the implications of this new architecture of standards for the scaling up of assessment, we need to examine what it is that assessment seeks to do. While its public image of certifying student achievements is often the most prominent, assessment has several distinct tasks to perform, and some of these are in tension with each other. Each purpose influences what students do, what they learn and how they approach their learning whether or not it intends to do so.

The three main functions of assessment are, firstly, to contribute to certifying student performance. This is often known as summative assessment. It represents how an institution judges student performance. It is formally recorded and validated as representing a set of achievements. Secondly, assessment provides students with useful information to aid their learning during their courses. They undertake a number of tasks, and feedback processes enable students to receive information that helps shape their subsequent study within the unit of study. It is commonly known as formative assessment. Finally, assessment builds students' capacity to make judgments about their own learning. To be effective as a learner, and subsequently as a practitioner in society, students need to be able to determine what they know and don't know and what they can do and can't do. If students only experience the judgment of others, then their ability to judge their work for themselves may not be developed. This purpose of assessment is known as sustainable assessment, as it is assessment that helps sustain learning over time and beyond the end of the programme (Boud & Soler, 2016).

Unfortunately, these functions can rarely be pursued simultaneously in any given assessment task. The intrinsic demands made of different purposes of assessment tasks and the information provided to students about their conduct of them differ in time and in content. If we take summative assessment, this needs to occur when students have *completed* their study of what is being assessed. Such assessment records what a student knows and can do at a particular point of time. To make this judgment while the outcomes being judged are not complete and learning of that matter is still occurring is invalid. It doesn't produce a worthwhile representation of what the graduate can do, only what they were able to do at some point in time prior to completion of the programme or course unit. This suggests that summative assessment needs to occur in the later stages of units or programmes rather than earlier. This is in conflict with the requirements of formative assessment. If this is left until after a student has completed their studies in a particular area, then it can't influence their performance as a student. They can't improve as a result of feedback because the timing of the task is too late in the sequence of study. The purpose of formative assessment needs to be pursued therefore at points in time when students can use useful information about their work when they still have an opportunity to apply it to their learning. That is, formative assessment needs to occur earlier in a unit or programme rather than later.

There is another tension between summative and formative assessment that needs to be considered: the kind of information each generates. Typically, assessment for certificatory purposes generates marks or grades. This may be sufficient as a summary of performance over a series of tasks, but the specific information it communicates about what a student can do, or more importantly for formative purposes, what a student needs to do to improve their work, is minimal. What information do marks like 72 % or 'C' provide to enable a student to do better? The information content of a mark is very low. To improve their work students need rich information about what they could do better, exemplars and illustrations of good work. If they are provided only as a supplement to marks and grades, they are often too late for students to act, as the task is typically timed too late for formative purposes. In summary, summative purposes need tasks that are loaded towards the end of the semester or the year; formative purposes need tasks loaded early in the semester or year.

The purpose of sustainable assessment is important throughout. Only by practising making judgments of their work on multiple occasions and being given assistance on calibrating their judgments will students develop the capacity to judge different kinds of work (Boud, Lawson, & Thompson, 2013). So, sustainable assessment activities need to be spread across the semester and, indeed, the programme. Success in this process can be determined if students are able to make good judgments about their performance in later summative tasks. If they cannot do this, then the outcome of sustainable assessment has not been met.

We can see from this discussion that one set of tasks cannot meet all the requirements of assessment, no matter how well they are constructively aligned (Biggs, 1996). This is not a simple matter of having a diversity of assessment methods, but of the timing of tasks, the provision of different types of information and the kinds of activity that surround tasks (feedback processes, self-judgments, etc.). That is, assessment be designed in the context of the overall educational design of programmes. In any given instance one purpose may need to be dominant, and this needs to change over the progress of a programme or course unit. Some tasks might be used predominantly for feedback, some predominantly for grading and others predominantly to promote self-regulation. Because grading has become such an all-consuming concern of students, careful design is needed to ensure that all the purposes of assessment fit together and can be successfully and compatibly pursued.

What Are the Implications of the Standards Agenda for Assessment?

Assessment today needs to be seen against the backdrop of the adoption of explicit academic standards and discipline learning outcomes. Within this context of stated learning outcomes, assessment is an activity that judges whether students can demonstrate attainment of these learning outcomes to a given standard. This is

needed to assure that the learning outcomes have been met by the time a student graduates, to enable a student to judge how they are progressing in meeting the learning outcomes and to provide useful information to students to enable them to meet these outcomes. Assessment now is necessarily focused on what students can do in relation to these yardsticks.

Addressing Outcomes Is Fundamental

This new focus has powerful implications that have not yet been fully realized in many higher education institutions. The most important is that assessment in the context of learning outcomes needs to start from considerations of how well the outcomes have been met. Emphasis here is relentlessly on how well has a specific standard been addressed: has it been met for purposes of certification, how much of a gap is there between what a student can do now and what they need to do to meet the standard and how will the student be able to judge if they are meeting the standard? The discourse of teachers and assessors needs to be about the characteristics of student work in the language of outcomes and standards, not about marks or doing better than other students.

In terms of the construction of assessment tasks that contribute in any meaningful way to final certification, they need, firstly, to identify the appropriate standards for the tasks students undertake and how they will be applied to the work at hand, that is, the criteria to be used to make judgments. This is not just the notion of criterion-referenced assessment of old (which tended to focus on criteria at the level of a particular unit of assessment or assessment task), but of standards-based assessment (which focuses on programme-wide suites of tasks). Secondly, there need to be assessment methods suitable for judging the particular learning outcomes being pursued. A range of approaches is needed beyond conventional tests and examinations. Thirdly, the balance of assessment approaches must reflect the range of learning outcomes. Overuse of particular methods has to be avoided, so, for example, if examinations are involved, then they must be restricted to judging the learning outcomes that can be best assessed by the particular kind of examination used. Finally, and most important of all, all necessary learning outcomes must be met by all students. The threshold for achievement for each outcome must be reached by all students for them to be able to complete the unit or be awarded the qualification. As Sadler (2015) puts it:

> The definitive measure of the adequacy of an institution's standards is whether the lowest performing students who gain credit for a course achieve higher order objectives to a sufficient degree. (p. 7)

No compensation through overachievement with respect to other outcomes is permitted. Of course, many students will also meet outcomes beyond the minimum and may be recognized for so doing.

The Metrics of Assessment Relate to Standards

Several things about the form in which assessments are reported follow from the centrality of outcomes to assessment. Does the result of an assessment act relate directly to an outcome and standard? Does a grade signify what a student can or can't do? Setting a pass mark is not setting a standard. Unless a pass means something real in terms of what is a particular learning outcome, it is irrelevant whether it is 40 %, 50 % or 60 %. Indeed marks that do not address an outcome are rendered meaningless in this conception of assessment.

The averaging of marks is particularly problematic. On what basis can marks related to different learning outcomes be averaged? This may be justifiable for different measures of the same outcome, but it is meaningless across outcomes. Such an act would assume that all outcomes are equivalent and that performance in one is the same as performance in another. When marks are averaged, information is lost, and the result no longer communicates information about what a student can do. Grade point averages are residues of a norm-referenced system in which students were compared with each other. These do not translate into a standards-based environment. This is not an argument that marks and grades should not be used at all. But they can only be used when they can be directly related to a desired learning outcome. New ways of thinking about progression are needed that are embedded in the language of standards and outcomes, not an inappropriate normalized metric.

Assessment Reporting Is Only as Fine Grain as the Judgment of Outcomes Will Allow

Consideration of marks raises another important issue. Marks and grades cannot be meaningfully reported to a greater level of accuracy than the judgment can stand. This is not a new observation; we have known for 80 years or so that essays, for example, cannot be marked to percentage level accuracy (Hartog, & Rhodes, 1935, 1936), but it has particular significance in an environment of transparency. The question to be faced is: to what level of accuracy can attainment of a particular outcome be judged? If only three or four categories of difference can be ascertained, then this is all that can be legitimately reported. Any finer grain of reporting is spurious and cannot be justified. Just because a test has 100 items, it does not mean that the results can be meaningfully reported to percentage accuracy. It is the attainment of the learning outcomes represented in the test that must be reported, and the level of accuracy for each of these is likely to be considerably less. The traditional Australian university classification of Pass, Credit, Distinction and Higher Distinction is probably the maximum level of granularity that judgment of most tasks against learning outcomes can stand.

As Sadler (2009) has argued, it also means that setting a generic set of standards for a programme or course unit is not enough: they need to be localized in terms of the particular outcomes of a programme or unit. The use of terms such as good, superior, and excellent does not indicate a standard or communicate a level. They need to be replaced by specific descriptions of what a grade means with respect to a particular standard. Grade descriptors are required to be outcome-specific: what in particular is required for a particular grade?

What Does a Standards-Based Approach Not Determine?

There are other matters that a standards-based approach does not restrict. These include that standards be unilaterally applied. That is, students have no say in the standards used. The threshold standards may be non-negotiable, but this says nothing about others. Students should be involved in assessment, through identifying and applying appropriate standards to their work. Such an approach does not imply that all learning can be predetermined. Outcomes may be set which can be addressed in diverse ways through many different products. What counts is whether the outcomes are met, not the particular form in which they are met. There is scope for creativity and meeting varied needs and interests through assessment tasks. Neither does this imply that all learning can be easily judged or is worthwhile. A limitation of conventional assessment methods is that they are often used to test that which it is easy to test, rather than that which is most important to test. A predominance of tests of memorization is an indicator of this. A standards-based framework draws attention to the need for an extended range of approaches that can be used for more outcomes that cannot be encapsulated in simple questions.

What Does a Standards-Based Framework Facilitate?

Many desirable educational features are inhibited by norm-referenced systems but enabled in a standards context. Students can more readily track their performance towards meeting major outcomes that are distributed over many course units. When assessment is recorded against each outcome, improvement can be plotted and thus ipsative assessment permitted (Hughes, 2011). Feedback processes are more likely to influence learning when it is clear that information provided on one task related to one outcome can be utilized in subsequent assessments for the same outcome (Boud & Molloy, 2013). Involvement of students in judging of their own work is more straightforward when they know what outcomes their work seeks to meet. They may need practice in identifying and utilizing criteria, but there are no additional manipulations of marks and grades that distinguish and inhibit knowing what grades stand for.

In summary, a standards-based approach to assessment would necessarily involve:

1. Reporting against standards and learning outcomes
2. The generation of as many reports as there are programme learning outcomes, thus tracking achievement by each outcome
3. Focus on each student meeting every threshold
4. Avoidance of averaging across different learning outcomes
5. Students appreciating what a grade signifies in terms of what they are seeking to be able to do rather than simply what they should know

It also permits the issuing of degree transcripts that relate to meaningful units of activity such as programme learning outcomes, rather than an arbitrary division by subject or course unit. The question: 'what can this graduate do?' can be reflected directly in the information provided in transcripts.

It may also encompass other features beyond the minimum demanded. For example, higher grades may be awarded for achievements beyond the simple meeting of learning outcomes; particular recognition for outstanding achievement on particular major tasks (like Honours projects) or for meeting additional outcomes (e.g. through the use of digital badges) may be given (Oliver, 2016). Higher thresholds may also be set for admission into advanced classes or for selection into higher degrees. All of these are readily encompassed in a standards-based framework.

Most importantly, a standards-based approach enables students to create multiple validated portrayals of achievements for different purposes. These permit students themselves to scale up beyond the standard portrayals that all students get on graduation so they can present themselves in different ways to different audiences (e.g. for employers, for higher degrees). A necessary addition to programmes to allow this is the use of programme-wide portfolios as repositories for all work and assessments or for those that are essential to their future professional work. These need to be designed to facilitate both summative purposes of assessment and formative ones and be able to be curated by students for different purposes (Clarke & Boud, submitted for publication).

Like any approach to educational programmes, a standards-based approach can be misused to centralize decision-making, limit the exercise of professional judgment and inappropriately control legitimate diversity of approach. The major trap in any outcome-oriented approach is that outcomes are viewed behaviouristically and in an overly operational form. To go to these extremes is to miss what a standards approach is seeking to achieve. It is not to produce students who are clones of each other after completing the same programme, but to represent holistically what they can do at an appropriate level of aggregation and to ensure that minimal outcomes are assured in all areas of importance. It is interesting to observe the application of such an approach in some Australian universities over many cycles. The first iteration tends to be compliance driven, and outcomes are written in stereotypical forms that are not aligned with assessments. This is a pretence of a standards-based

approach. The second phase begins to take the process seriously, but if not carefully monitored it can lead to excessive specification of outcomes, many checklists and a proliferation of summative assessments. It is probably a necessary stage in the process as academics come to terms with what a standards-based approach can offer. The third phase, some years later, involves a manageable number of programme learning outcomes and learning outcomes for each unit and the beginnings of good alignment of assessment at both the level of the programme and the course unit and the effective portrayal by students of what the programme has enabled them to demonstrate. As in any major reorientation, unless an institution is prepared to persist through several iterations, then the value of the approach will not be realized.

Progress to Date

Although there are many entire higher education systems that have moved in total or in part to a standards-based or outcome-oriented framework (Australia, Hong Kong, UK), and thus to a more transparent assessment regime, not all of the implications discussed above are yet manifest in all programmes in these jurisdictions, and some have yet to be taken up in major ways. Many of our assessment conventions were created for a different era. The specification of learning outcomes for course units is now commonplace, and the identification of programme level learning outcomes is in progress. When these outcomes are written for the first time, they rarely fully represent what is desired as they may be prompted by the need for compliance rather than an embracing of the framework. However, in the second and third iterations, more robust outcomes can result.

What is much less common is the working through of the implications for marks and grades and the aggregation of results. It is still normal for a single mark to be calculated for a course unit from a weighted average of marks in that unit and for these to be averaged over units in ways that are educationally unjustifiable. Tradition bears particularly heavily on the marking process, and it is often easier to implement change in curricula and learning outcomes than in the assessment process (Deneen & Boud, 2014).

Progress has also been made on the use of learning portfolios, although their use for whole programmes is much less common that it is within course units. It is only as the electronic portfolio becomes commonplace that the digital affordances it provides can enable their use for the purposes described here (Clarke & Boud, submitted for publication). Hard-copy portfolios are too unwieldy to allow for the multiple transformations for different purposes needed for a full scaling up effect on assessment.

Conclusion

As these changes in policy frameworks, assessment policies and curriculum design are adopted, transparency becomes an increasing feature of assessment. Standards and learning outcomes are explicit. The ways in which they are judged can be seen to address them, and the results of assessment acts are clearly related to what students are expected to do. The connection between what is claimed for a programme and what a successful graduate of the programme has achieved is readily available for inspection. Assessment results then can be used as part of quality enhancement and assurance processes in ways that are often not available at present.

References

Australian Learning and Teaching Council (ALTC). (2011). *Learning and teaching academic standards project final report*. Sydney, Australia: Australian Learning and Teaching Council.

Baik, C., & James, R. (2014). Seven strategic areas for advancing the assessment of learning in higher education. In H. Coates (Ed.), *Higher education learning outcomes assessment: International perspectives*. Frankfurt, Germany: Peter Lang.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364.

Boud, D., Lawson, R., & Thompson, D. (2013). Does student engagement in self-assessment calibrate their judgment over time? *Assessment & Evaluation in Higher Education, 38*(8), 941–956.

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

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

Clarke, J., & Boud, D. (submitted for publication). *Refocusing portfolio assessment: Curating for feedback and portrayal*.

Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education, 39*(5), 577–591.

Hartog, P., & Rhodes, E. C. (1935). *An examination of examinations*. London: Macmillan.

Hartog, P., & Rhodes, E. C. (1936). *The marks of examiners*. London: Macmillan.

Higher Education Standards Framework. (2015). *Federal register of legislative instruments F2015 L01639*. Canberra, Australia: Commonwealth of Australia.

Hughes, G. (2011). Towards a personal best: A case for introducing ipsative assessment in higher education. *Studies in Higher Education, 36*(3), 353–367.

Oliver, B. (2016). *21C credentials: Curate, credential and carry forward digital learning evidence*. Sydney, Australia: Office for Learning and Teaching.

Organisation for Economic Co-operation and Development (OECD). (2016). *International assessment of higher education learning outcomes (AHELO) feasibility study*. Retrieved from www.oecd.org/edu/ahelo

Quality Assurance Agency for Higher Education (QAA). (2016). *Quality assurance agency subject benchmark statements*. Retrieved from http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education, 34*(7), 807–826.

Sadler, D. R. (2015). Three in-course assessment reforms to improve higher education learning outcomes. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1064858.

Chapter 3
Surfacing and Sharing Advances in Assessment: A Communities-of-Practice Approach

Dai Hounsell and Tracy X. P. Zou

Abstract The focus of this chapter is on how to encourage the take-up of advances in assessment and feedback practices across and beyond one university, in ways that can bridge subject and organisational boundaries while avoiding top-down prescription and maintaining respect for scholarly autonomy. Against the wider backdrop of key issues that institutions need to grapple with in scaling up assessment renewal constructively, the chapter discusses a communities-of-practice initiative at the University of Hong Kong. It sought to develop an approach to surfacing and sharing understandings about effective innovations in assessment that could help in the pursuit of institutional strategic goals. The approach adopted in this initiative is discussed in terms of the main assessment and feedback themes it addressed, the intended audiences for the project's work, the various strategies deployed in surfacing innovative practices internationally as well as internally and the means by which guiding principles and contextualised instances of developments in practice were more widely shared.

Introduction

With respect to assessment and feedback, universities across the world share a formidable challenge: how can they continue to ensure that their students thrive in authentic and engaging learning environments attuned to twenty-first-century needs? Addressing this challenge is far from straightforward, since three interlocking dilemmas need to be resolved. First, there is the need to strike a balance between local exigencies and emerging directions of travel globally. The aim, in other words, is to ensure on the one hand that developments in assessment within a given university go with the grain of disciplinary and institutional requirements and strategic goals, while on the other there is an openness to what might be learned from advances in practices in other universities, subject areas and national systems of higher education. Second is laying sound foundations for informed

D. Hounsell (✉) • T.X.P. Zou
University of Hong Kong, Hong Kong, China
e-mail: daihounsell@btinternet.com

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_3

choice: how to help colleagues navigate their way through a bewildering maze of possibilities, in ways that avoid prescription by – in Bruner's famous phrase – *showing* in context rather than *telling* out of context (Bruner, 1966, p. 151). Third is identifying a means of brokering efforts to rethink and reshape assessment and feedback practices – in other words, developing resources and strategies that can facilitate the communication of ideas across organisational boundary lines, whether these lie between programmes, levels of study or departments and faculties.

This chapter explores how one higher education institution of global standing, the University of Hong Kong (HKU), tackled this trio of dilemmas through a communities-of-practice initiative that sought to bring to the surface international as well as local advances in assessment and feedback and to share these more widely across and beyond the university. The approach adopted has implications, it is argued, for strategies to increase the scale of efforts to rethink and reconfigure assessment practices.

Dilemmas in Advancing Assessment and Feedback

Global and Local Dimensions

Assessment processes in universities can helpfully be viewed in terms of an interplay between the global and the local. The assumption that assessment has a considerable *global* dimension is well established. It is widely taken for granted, for example, that forms and modes of assessment that have evolved in one given setting nonetheless possess a substantial measure of generality, i.e. that they will be applicable across a wide range of disciplines, institutions, levels of study and even – in a more globalised world – university systems. Yet as Peter Knight (2006) has cogently argued, there is an equally significant *local* dimension to assessment, since the making of judgements about the quality of students' performance or achievement, for instance, is inescapably context bound, i.e. firmly rooted in the contingencies of a particular course, level of study and institutional setting.

This interplay between the local and the global is especially pronounced in the assessment practices of the Western model of universities, in part because of an individualistic culture (Vinther & Slethaug, 2013) in which university teachers in many institutions and subject areas have traditionally enjoyed a degree of discretion with regard to the design of assessments in the courses for which they are responsible. That element of discretion forms part of a larger and long-established ethos of academic autonomy that extends to curriculum design and choice of approaches to teaching and learning. It does also make good sense from the contemporary perspective of constructive alignment (Biggs & Tang, 2011), since congruence between curriculum goals, teaching and assessment is more likely to lead to high-quality learning. For the individual teacher or course team, however, local (i.e. curriculum specific) considerations about assessment have to be weighed against more global ones. The latter include departmental or faculty guidelines and prevailing practices; the 'signature pedagogy' of the discipline concerned

(Shulman, 2005), with respect to the distinctive forms of assessment customarily adopted in the wider subject or professional community; and whatever policy and regulatory frameworks may be in place at institutional level that govern assessment practices in that particular university.

A similar interplay between local and global is evident at the institutional level, where the local dimension may comprise a distinctive assessment 'culture' or 'regime' (Trowler, 2008) which has evolved in that particular university – whether it be, for example, of the main or major determinant of degree outcomes arising from closed-book final examinations, a capstone dissertation or a grade point average. This too may be mediated by global factors such as a national system of quality assurance that includes guiding principles on assessment (see, e.g., QAA, 2014) and the requirements of professional accrediting bodies (national or international) that apply to whatever programmes of study are subject to external validation within the university concerned.

It follows that any initiative that seeks to bring about a rethinking and reshaping of assessment practices – whether within a faculty, across an institution or system wide – has to take account of these complex interrelationships between the global and the local. These interrelationships are shaped not only by multiple regulatory frameworks and layers of accountability, however. They are also moulded by overlapping sets of cultural norms and conventions (disciplinary, institutional, professional, national) that may be largely tacit rather than explicit, yet they can powerfully influence what methods or modes of assessment are held to be valid and appropriate by the wider subject or professional community.

In such circumstances, strategies to foster review and reshaping of assessment which seem more likely to succeed are those which are descriptive rather than prescriptive (i.e. they focus on what can or might be attempted, rather than what must be done) and which present a range of possibilities rather than offer a single pathway forward (and thus enable exploration of those options which seem most attuned to disciplinary cultures and curriculum goals).

Exemplification and Mapping of Changing Practices

A distinctive feature of the higher education assessment literature is the pervasiveness of a form of literature which has been characterised as an 'account of practice'. As defined here, an account of practice is a record of efforts made in a particularised, real-life context (typically a specified course unit or module, a programme of study or a subject area, in a given institution or group of institutions) to introduce a significant change in some aspect of the assessment of students' progress, performance or achievement. Accounts of practice have been a notable if under-noticed feature of the higher education assessment literature, making up two-fifths of the publications identified in a review of the literature on innovative assessment (Hounsell et al., 2007). In some instances, the practice examined may be the focus of thoroughgoing empirical investigation, and thus of interest from a research perspective, but many such accounts are directed chiefly towards a

practitioner audience, within and beyond the subject area concerned, reporting and reflecting on experiences in designing, implementing and evaluating a change in assessment practice. And it is because they are grounded in authentic day-to-day experiences, and documented by established subject insiders, that accounts of practice seem to have a high degree of street credibility for fellow practitioners (Hounsell et al., p. 71). By dint of exemplification, they provide practitioners with a bridge between the global and the local.

When assembled into larger groupings, accounts of practice have also played a vital role over the last two decades in the documentation and dissemination of what has been learnt from developments in assessment and feedback in everyday course settings, as is shown in Table 3.1. It lists 12 booklets, handbooks and websites which together record around 670 accounts of practice. Some were compiled for use chiefly within one or a cluster of institutions, while also being made more widely available; others focus on developments across a national system of universities or in one instance within a subject area. Some adopt a wide-angle lens, while examples have been chosen to illustrate an evidence-informed model or set of guidelines. What is common to all, however, is that accounts of practice, albeit in a variety of forms and depths of detail, have been organised around a thematic or conceptual framework. In other words, they guide practitioners in navigating their way through the resources on evolving practices by combining *exemplification* with *mapping.* The HKU initiative has tried to learn from this twofold strategy for facilitating informed choice and decision-making by practitioners.

Communities of Practice and Brokering

A third element in shaping our approach has been the concept of communities of practice, first introduced by Lave and Wenger (1991) one-quarter of a century ago. Communities of practice were initially conceptualised as groups in which newcomers learn through engaging in 'legitimate peripheral participation', gradually moving to a more central place through constantly observing, interacting with and practising with people having more expertise in the community. In Wenger's later publications (e.g. Wenger, McDermott & Snyder, 2002; Wenger-Trayner, Fenton-O'Creevy, Hutchinson, Kubiak & Wenger-Trayner, 2015), communities of practice are more explicitly defined as groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly. They have also argued that practitioners tend to be associated with multiple communities rather than a single entity:

> Professional occupations [...] are constituted by a complex landscape of different communities of practice – involved not only in practising the occupation, but also in research, teaching, management, regulation, associations, and many other relevant dimensions. All these practices have their own histories, domains, and regimes of competence. (Wenger-Trayner & Wenger-Trayner, 2015, p. 15)

Table 3.1 Compilations of evolving assessment practices in higher education

Thematic focus	Publication details	Country of publication	No. of case examples
Student-centred assessment	Gibbs, G. (1995). *Assessing student centred courses.* Oxford: Oxford Centre for Staff Development	UK	19
Changing assessment	Hounsell, D., McCulloch, M., & Scott. M. (Eds.). (1996). *The ASSHE Inventory: Changing assessment practices in Scottish higher education.* Edinburgh: Association for the Advancement of Sustainability in Higher Education. Retrieved from https://www.researchgate.net/publication/234572401	UK	124
Assessing learning	Nightingale, P., Wiata, I. T., Ryan, G., Hughes, C., & Magin, D. (1996). *Assessing learning in universities*. Sydney: University of New South Wales Press	Australia	62
Formative feedback	Juwah, C., Macfarlane-Dick, D., Matthew, B., Nicol, D., Ross, D., & Smith, B. (2004). *Enhancing student learning through effective formative feedback.* York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/sites/default/files/resources/id353_senlef_guide.pdf	UK	8
Learning-oriented assessment	Carless, D., Joughin, G., & Liu, N. -F. (2006). *How assessment supports learning: Learning-oriented assessment in action.* Hong Kong: Hong Kong University Press	Hong Kong	39
Assessment in the biosciences	Krause, K. -L., & Harris, K. -L. (2007). *Enhancing assessment in the biological sciences.* Melbourne University/Sydney University. Retrieved from http://bioassess.edu.au	Australia	79
Re-engineering assessment	Nicol, D. (2007). *Re-engineering assessment practices.* University of Strathclyde. Retrieved from http://www.reap.ac.uk/reap/assessment/index.html	UK	19
Assessing digitally	JISC (2009). *Effective assessment in a digital age.* Joint Information Systems Committee. Retrieved from http://repository.jisc.ac.uk/6004/1/effectivepracticedigitalage.pdf	UK	10
Enhancing feedback	Hounsell, D., Robinson, N., Crook, A., & Stannard, R. (2010). *Enhancing feedback.* University of Edinburgh. Retrieved from http://www.enhancingfeedback.ed.ac.uk	UK	200+

(continued)

Table 3.1 (continued)

Thematic focus	Publication details	Country of publication	No. of case examples
Assessment futures	Boud, D. (2010). *Assessment futures*. University of Technology, Sydney. Retrieved from http://www.uts.edu.au/research-and-teaching/teaching-and-learning/assessment-futures/overview	Australia	30+
Assessing learning outcomes	NILOA. (2012). *Learning outcomes assessment.* National Institute for Learning Outcomes Assessment, USA. Retrieved from http://www.learningoutcomesassessment.org	USA	9
Capstone projects and dissertations	Healey, M., Lannin, L., Stibble, A., & Derounian, J. (2013). *Developing and enhancing undergraduate final-year projects and dissertations.* York: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/node/8079	UK	70+

According to Wenger, McDermott and Snyder (2002), there are three important features essential to a community of practice. *Domain* is the subject area in which members are interested and passionate about. *Community* denotes the joint activities and ongoing interactions that members are engaged with in order to pursue their shared interest. *Practice* means that each member is a practitioner, and collectively they develop a shared repertoire of resources, including stories, tools and ways of addressing problems. For Wenger-Trayner et al. (2015), it is out of these three elements in combination that a community of practice is constituted and from which it develops.

In higher education, communities of practice have been conceptualised as both naturally occurring phenomena and as ones which could be intentionally cultivated. Concerning the former, Brew (2012) has argued that a discipline, a university as a whole or a network of professionals can each be seen as a community of practice by dint of their modus operandi and of the presence of the three essential elements of domain, community and practice. Indeed, from the standpoint of research and scholarship, it seems relatively straightforward to interpret a discipline in the higher education environment as a community of practice, since academics typically have strong disciplinary allegiances and participate in a range of activities (conference attendance, journal publication, seminars) in order to maintain and advance their expertise through interactions with others sharing similar interests.

From the standpoint of teaching and learning, however, extrapolation may not be so clear-cut. A university teacher exploring the possibility of a specific change in their approach to, for example, assessing students or designing groupwork may well not be able to readily identify relevant instances of such an approach amongst their departmental or faculty peers, nor even amongst their disciplinary peers in other

universities. Indeed, in seeking to capitalise upon the experiences of others who have already experimented with the same or a similar approach, the more promising avenues to be explored may well lie transversely, in subject areas other than one's own. Communication across disciplines is however likely to pose greater challenges to initiatives seeking to cultivate vibrant communities of practices. The shared *domain* is not (as it would be for a research-focused community of practice) the discipline, but rather those elements of the mutual endeavour of advancing students learning that are found to be in common, regardless of disciplinary differences. Further, the 'shared repertoire of resources, stories, tools and ways of addressing problems' that typifies a *practice* may not be as directly accessible as it would be for a research community of disciplinary peers. It may well need to be more consciously teased out and assembled over time.

Similar difficulties arise in transposing the notion of *community* from research to teaching. There are well-established channels in every subject area for communicating conceptual and empirical advances, and all but perhaps the most junior members of the disciplinary community will have acquired considerable familiarity with the communicative norms and conventions associated with authorship of research through reading papers and listening to presentations. Such expertise cannot be so readily taken for granted in the communication of teaching and assessment experiences and insights and especially so where interchange is taking place across subject boundaries.

The point to be made here is that efforts to actively cultivate communities of practice around pedagogical issues need to be alert to such potential challenges and, where appropriate, seek ways of addressing them. And indeed, there is growing evidence of successful attempts to foster communities of practice on teaching and learning topics in universities around the world. Pharo et al. (2014) report that the establishment of such communities in four Australian universities has improved interdisciplinary teaching through facilitating both staff development and institutional learning, while at HKU a community of practice in the Faculty of Dentistry consisting of both faculty members and students was developed through an international peer review and critique process involving five other universities (Gardner, Bridges & Walmsley, 2012). In the USA, a case study of an institution's mechanisms for conducting and enhancing assessment identified the development of communities of practice as a fruitful means of exchanging ideas and perspectives across colleges and disciplines (Guetterman & Mitchell, 2016). In the UK, one study found that developing a community of practice could effectively promote transnational teaching and raise the quality of learning by strengthening partnership (Keay, May, & O' Mahony 2014), while another examined the key role of designated 'associates' in forging links between their discipline-based communities of practice and a centrally coordinated initiative to promote assessment-for-learning approaches across the university (Reimann & Wilson, 2012). The associates thus had a *brokering* role (Wenger, 1998), enabling them to import and export elsewhere new meanings and practices (Reimann & Wilson, 2012, p. 81–82).

As Trowler, Saunders and Bamber (2009) have observed, brokering can play a crucial role in efforts to foster enhancement of curricula, modes of learning and teaching and assessment in higher education that cross boundaries between institutions and departments:

> If we depict educational organisations and the workplace as different activity systems, characterised by different communities of practice, then moving from one to another involves a form of brokerage in which a variety of tools might aid and develop learning – what we might call 'bridging' tools. When groups of teachers are asked to adopt one set of practices rather than another, brokerage and bridging tools are called for. (Trowler et al., 2009, p. 14)

Wenger-Trayner et al. (2015) have similarly commented on the importance of using a range of strategies to enable productive cross-boundary interactions:

> Facilitating boundary crossing, for example, involving certain people in brokering information across different stakeholder groups; creating or improving boundary objects, such as documents, that speak to people in different sectors; organising visits to the practice of potential partners; devising projects that require people from different backgrounds to negotiate a common aim. (Wenger-Trayner et al., 2015, p. 108)

As illustrated in the following section, brokerage across departmental and faculty boundaries, together with the creation of a range of outputs that could serve as 'bridging tools', has been a key feature of the communities-of-practice initiative at HKU.

The Communities-of-Practice Initiative at HKU

The initiative at HKU is a 2-year project aiming to foster the development of communities of practice on themes of strategic significance in sustaining and enhancing the quality of teaching and learning at the university. The present chapter focuses on the first of the project's two themes, *wise assessment*, where the aim has been to surface and share effective assessment practices from within and beyond HKU. The University's Centre for the Enhancement of Teaching and Learning has largely served the major brokering role for the community of practice in an attempt to celebrate, promote and exchange advances in assessment across disciplines and faculties. The brokering role has gradually become more widely shared by faculty members and teachers who, having initially communicated their own attempts to innovate to departmental and faculty colleagues, have since become engaged in university-wide interactions.

The approach that has been adopted to broker advances in assessment at HKU comprises four interrelated components: thematic strands, processes, outcomes and audiences (see Fig. 3.1).

Fig. 3.1 Thematic strands, processes, audiences and outputs in the wise assessment strand of the HKU project

Thematic Strands

Like many universities of global standing, HKU is faced with the challenge of reviewing and enhancing the provision of assessment and feedback in order to ensure that its students continue to thrive in an engaging learning environment attuned to twenty-first-century needs. Its key principles include assessment for learning, alignment with student learning outcomes, diversity of assessment types, equitable assessment and timely and professional feedback (University of Hong Kong, 2015). Assessment and feedback, however, are often the aspects of university study on which students report comparatively lower levels of satisfaction (Carless, 2015). In an early study across higher education institutions in Hong Kong, students had perceived the lack of useful feedback as a problem in the assessment process that inhibited their learning, and a disparity was also evident between tutors' beliefs and students' perceptions of feedback (Carless, 2006). Subsequent annual surveys at HKU have also identified the quality of feedback, together with student uncertainty about goals and standards, as recurring areas of concern in the institution, and these were therefore chosen as two key strands upon which the work of the project would focus.

The other two key strands arose directly from major curriculum initiatives that have called for fresh thinking about assessment. One of these is experiential learn-

ing, a distinctive feature of the undergraduate curriculum at HKU. All undergraduate students are required to tackle real-life issues and problems by making use of the theoretical knowledge they have acquired in the formal curriculum (Ho & Ko, 2013). Needless to say, assessing students' progress and performance in experiential learning presents all sorts of challenges, for example, articulating learning outcomes in authentic ways that can be constructively aligned to assessment and being alert to ethical issues which may have assessment and feedback implications (Zou & Cheng, 2015). The second major initiative is the launch in the academic year 2011–2012 of a Common Core Curriculum (CCC) as a formal requirement for all HKU undergraduate students, following the system-wide reform that increased the length of a first degree from 3 to 4 years (Education Commission, 2003). At HKU, the new CCC seeks to foster a broad perspective and a critical understanding of the complex interconnections of daily life. Each undergraduate student needs to take six six-credit courses from four CCC areas of inquiry (Scientific and Technological Literacy, Humanities, Global Issues and China: Culture, State and Society) to fulfil the graduation requirement. From the outset, there was an expectation that assessment practices in CCC would be more diversified and innovative than had been the case elsewhere in more traditional degree programmes, and many of the emerging CCC course teams have grasped the opportunity to actively experiment with new methods in the design of assessments.

Processes

As Fig. 3.1 indicates, a variety of processes were pursued within the project, including surveying emerging practices, interviewing samples of practitioners, searching the relevant literature and drafting resource materials. These activities serve important brokering functions as they facilitate navigation across disciplinary and departmental boundaries and will result in a set of bridging tools. The activities undertaken for the strand on the CCC offer a means of illustrating how and why these processes were deployed.

Survey A priority in the CCC strand was to establish an accurate and up-to-date picture of the assessment practices adopted. The existence of an informative CCC course handbook, updated annually and following a standard format for each course description, made it possible to derive this overall picture, since the assessment methods used and their relative weightings in arriving at an overall grade were shown for each course. As Fig. 3.2 shows, analysis of the data yielded distributions of assessment methods overall and in relation to the four CCC areas of inquiry (Hounsell & Cheung, 2014). Across the 152 common core courses, a total of 582 assessment tasks could be found, and these were grouped into 11 broad types. While the most commonly adopted methods included traditional forms such as essays and reports (68 %) and exams and quizzes (61 %), there were also extensive instances of less conventional methods that seemed particularly well attuned to distinctive features of the new curriculum such as greater diversity in learning and assessment

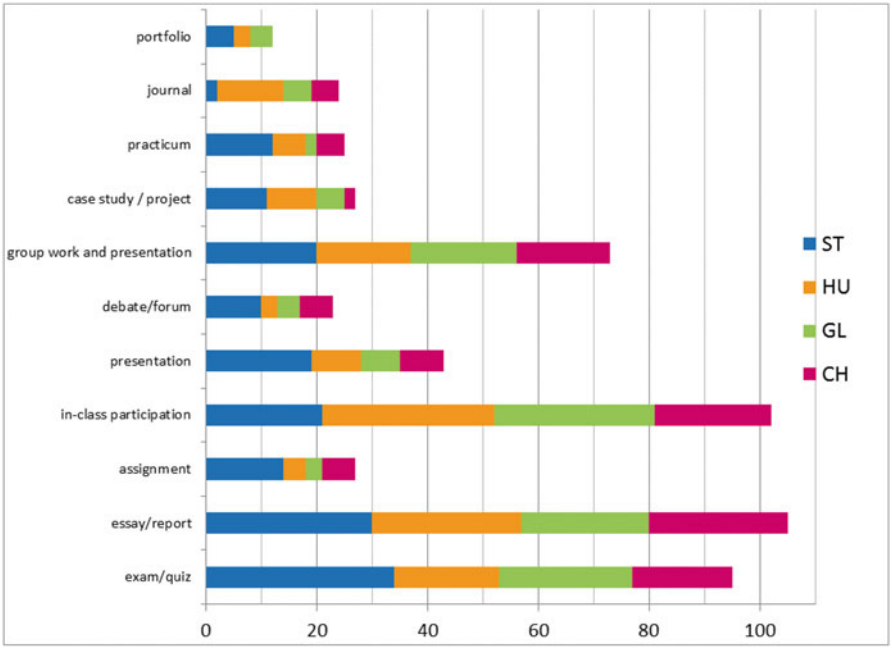

Fig. 3.2 Number of assessment tasks ($n = 592$) by method of assessment across 152 common core courses (Hounsell & Cheung, 2014). (Key: *ST* science and technology literacy, *HU* humanities, *GL* global issues, *CH* China: culture, state & society)

and opportunities for students to work interactively. Three of these departures from convention, all well represented in the CCC, were chosen for more in-depth scrutiny: oral and mixed media presentations by students, groupwork of a diversity of kinds and in-class participation in its various contemporary manifestations.

Interviews The next step was to focus on 'wise practices' in the CCC in relation to these three methods of assessment, as seen from the perspective of the practitioners themselves. Eighteen face-to-face interviews were conducted with a cross section of course coordinators and their teaching assistants. All the interviews had both a global dimension and a local dimension, aiming to identify on the one hand what, in the experiences of the practitioners concerned, seemed to make for effective practice in the use of the chosen assessment method while, on the other, being alert in each instance to those distinctive features of the course setting which had helped to shape how the method had been implemented. After each interview, the project team produced case examples summarising the assessment design and its implementations, which were checked and verified by the interviewee.

Literature Searches A concomitant step entailed a search of the higher education assessment literature for salient material on this and the other three thematic strands of the project that could underpin our efforts in relation both to *mapping* and

exemplification. This meant that, with respect to exemplification, the aim was to track down documented case examples from outside of HKU that could contribute to a pool of contextualised instances of 'wise practice' from across an appropriately wide spectrum of disciplines and professional fields. As regards mapping, relevant conceptual and empirical studies were examined, in a quest for insights and guiding principles that would help to map key considerations in developing effective practices.

Audiences

A key target constituency for the project comprises HKU teachers and other academic staff whose day-to-day responsibilities entail the practice of assessment in its various guises. This group includes not only those who are early adopters of innovative assessment practices but also those who are already engaged in some form of assessment for learning, as well as those for whom the concept of assessment for learning is still a relatively unfamiliar one. Reaching out to potential community members with differing levels of assessment expertise is an important component of brokering. The primary audience also includes a comparatively small but nonetheless influential group of HKU teachers who are not only academics themselves but also have formal organisational responsibilities for promoting teaching and learning, for example, as faculty deans or associate deans. Including this group as a target audience opens up additional opportunities for system-level support and recognition, which was not a principal focus in the initial conceptualisation of communities of practice, but gradually came to be seen as an important factor in 'building a case for action' (Wenger, McDermott & Snyder, 2002, p. 77) across an institution.

Besides HKU teachers, the audiences for the project have also included a broader network of scholars and practitioners, regionally and internationally, with recognised expertise in assessment, and some of them have had a fuller involvement as advisors to the project or as contributors to its resource materials. Interaction with this wider group has been a further means of addressing the global and local dimensions of assessment practices.

HKU students comprise another important group of stakeholders. They have contributed to the project's resource materials and participated in events. Some teachers at HKU have also begun to share the resource materials with their students as a pathway towards greater assessment literacy (Price, Handley, Millar & O'Donovan, 2010).

Outputs

From the outset, the project was committed to generating a range of bridging tools that would offer community participants a measure of choice not only of modes but also of levels of engagement, recognising that colleagues would each be juggling

with competing teaching, research and service commitments, but would also differ in their strength of interest in a given thematic strand – depending on whether, for example, their main concern was to keep in general touch with evolving approaches to assessment or that they had identified a promising assessment opportunity that they were keen to learn more about. Outputs were therefore in a variety of modes: written resources, face-to-face events and video materials.

The written resources chiefly took the form of 12 'briefings', varying in length from two to eight sides of A4 and in printed form as well as downloadable via the project's website, the *Wise Assessment Forum* (www.cetl.hku.hk/wise-assessment-forum/). Seven of these were compiled by the project team, distilling key elements of effective practice by drawing on the project's survey and interview material as well as literature searches, with ample illustrations of effective practices in the form of case examples from a range of subject areas and sourced from within HKU as well as globally. A further five were specially commissioned from recognised experts, whether by adapting and reprinting with permission high-quality published material that would engage and inform the target audiences or by contracting accessible introductions to more substantial material available from published sources.

The face-to-face interactions have mainly comprised a series of 'join-the-conversation' events, publicised university wide to encourage boundary crossing, but sometimes oriented towards a particular interest group such as CCC course teams, and involved the originators of case examples and student representatives as panellists. In these events, as their name suggests, the accent is on dialogue and interchange, with the minimum of formality, and an opportunity to debate issues raised in the printed briefings.

The third cluster of outputs involves digital videos of two kinds. The first of these is a 'vox pop' on feedback as seen from the perspective of HKU students, 20 of whom shared on video their experiences of getting feedback on the assignments and assessments. The resulting vox pop is intended as a stimulus to discussion, whether by teaching colleagues in a subject group or department, or between, for instance, a teacher and his or her class of students. Second is a set of videos, dubbed 'talking heads', each of which is a short, edited recording of an interview with a prominent scholar about their insights on assessment for learning. The goal is to communicate these insights succinctly in an accessible way, linked to fuller published work which can be followed up where appropriate.

Concluding Comments

The approach which the HKU project has followed in scaling up the renewal of assessment and feedback has four main characteristics:

- An a priori acknowledgement of the extent to which assessment practices in their everyday course settings have features which are shaped by local circumstances (the subject area, curriculum goals, a departmental culture, available resources)

as well as by more global concerns such as institutional strategic goals as well as understandings about, for example, the validity and robustness of a given form of assessment
- To exemplify advances in assessment practices across a range of subjects and contexts and also to distil principles that can guide adoption and adaptation on a wider scale
- The value of a communities-of-practice perspective, particularly the need for brokerage to encourage and facilitate interchange, especially across organisational and disciplinary boundaries
- The deployment of a variety of modes of communication, with the aim of accommodating differing forms and levels of engagement from target audiences

This approach has limitations as well as strengths which need to be borne in mind when considering how it might inform scaling-up initiatives in other universities. First, tracking down, mapping and exemplifying information about relevant developments in assessment and feedback practices are demanding both in time and in expertise, whether the focus is locating and documenting internal instances or searching and retrieving already-published accounts from databases and websites. Collaborative efforts by more than one institution, or perhaps a crowd-sourcing approach across a subject area, may therefore offer greater cost-effectiveness. Second, the resources of the project have thus far been mostly deployed in creating bridging and brokering tools, and a great deal of follow-up work – no less demanding of time and expertise – needs to be done over the coming months and years if productive communities of practice are to flourish.

Finally, it should be emphasised that the HKU project does not represent a universal blueprint for scaling-up assessment but rather (in keeping with its central ethos) a set of practices from which others may be able to learn. As Bamber, Trowler, Saunders and Knight (2009) have argued:

> Effective change is embedded in its context and comes when those involved make it their own through use and adaptation to local histories and contexts. (p. 2)

Acknowledgements The HKU Community of Practice Project was made possible by generous funding from the University Grants Committee (UGC) of the Hong Kong Special Administrative Region. We are also grateful to Ms. Sonia Cheung for data collection and analysis and to Professors Grahame Bilbow and Ian Holliday for their encouragement and support.

References

Bamber, V., Trowler, P., Saunders, M., & Knight, P. (Eds.). (2009). *Enhancing learning and teaching in higher education: Theory, cases, practices*. Maidenhead, UK: Open University Press/Society for Research into Higher Education.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Maidenhead, UK: Open University Press.

Brew, A. (2012). Teaching and research: New relationships and their implications for inquiry-based teaching and learning in higher education. *Higher Education Research & Development, 31*(1), 101–114.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.
Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.
Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.
Education Commission (Hong Kong). (2003). Review of the academic structure of senior secondary education. Retrieved from http://www.e-c.edu.hk/eng/reform/index_e.html
Gardner, K., Bridges, S., & Walmsley, D. (2012). International peer review in undergraduate dentistry: Enhancing reflective practice in an online community of practice. *European Journal of Dental Education, 16*(4), 208–212.
Guetterman, T. C., & Mitchell, N. (2016). The role of leadership and culture in creating meaningful assessment: A mixed methods case study. *Innovative Higher Education, 41*(1), 43–57.
Ho, M., & Ko, A. (2013). *Integrating experiential learning into formal curriculum: The HKU experience*. Presentation made at the 4th Asia-Pacific Regional conference on service learning, Hong Kong. Retrieved from http://commons.ln.edu.hk/cgi/viewcontent.cgi?article=1063&context=osl_conference
Hounsell, D., Blair, S., Falchikov, N., Hounsell, J., Huxham, M., Klampfleitner, M., & Thomson, K. (2007). *Innovative assessment across the disciplines: An analytical review of the literature.* York, UK: Higher Education Academy (HEA). Retrieved from http://www.heacademy.ac.uk/ourwork/research/litreviews
Hounsell, D., & Cheung, S. (2014). *Patterns of assessment in the common core curriculum at the University of Hong Kong*. Centre for Enhancing Teaching and Learning, University of Hong Kong. Retrieved from http://www.cetl.hku.hk/wise-assessment-forum/assessment-in-the-common-core-curriculum/
Keay, J., May, H., & O' Mahony, J. (2014). Improving learning and teaching in transnational education: Can communities of practice help? *Journal of Education for Teaching, 40*(3), 251–266.
Knight, P. (2006). The local practices of assessment. *Assessment & Evaluation in Higher Education, 31*(4), 435–452.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.
Pharo, E., Davison, A., McGregor, H., Warr, K., & Brown, P. (2014). Using communities of practice to enhance interdisciplinary teaching: Lessons from four Australian institutions. *Higher Education Research & Development, 33*(2), 341–354.
Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education, 35*(3), 277–289.
Quality Assurance Agency for Higher Education (QAA). (2014). *UK Quality Code for Higher Education – Chapter B6: Assessment of students and the recognition of prior learning*. Gloucester, UK: Quality Assurance Agency for Higher Education. Retrieved from http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-codeg
Reimann, N., & Wilson, A. (2012). Academic development in 'assessment for learning': The value of a concept and communities of assessment practice. *International Journal for Academic Development, 17*(1), 71–83.
Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus, 134*(3), 52–59.
Trowler, P. (2008). *Cultures and change in higher education: Theories and practice*. Basingstoke, UK: Palgrave Macmillan.
Trowler, P., Saunders, M., & Bamber, V. (2009). Enhancement theories. In V. Bamber, M. Saunders, P. Trowler, & P. Knight (Eds.), *Enhancing learning, teaching, assessment and curriculum in higher education* (pp. 7–15). Maidenhead, UK: Open University Press/Society for Research into Higher Education.
University of Hong Kong. (2015). *Quality manual.* Retrieved from http://tl.hku.hk/quality-manual/
Vinther, J., & Slethaug, G. (2013). The influence of internationalisation and national identity on teaching and assessments in higher education. *Teaching in Higher Education, 18*(7), 797–808.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, MA: Cambridge University Press.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Boston: Harvard Business School.

Wenger-Trayner, E., Fenton-O'Creevy, M., Hutchinson, S., Kubiak, C., & Wenger-Trayner, B. (Eds.). (2015). *Learning in landscapes of practice: Boundaries, identity, and knowledgeability in practice-based learning*. London: Routledge.

Wenger-Trayner, E., & Wenger-Trayner, B. (2015). Learning in a landscape of practice: A framework. In E. Wenger-Trayner et al. (Eds.), *Learning in landscapes of practice: Boundaries, identity, and knowledgeability* (pp. 13–29). London: Routledge.

Zou, T. X. P., & Cheng, S. (2015). *Assessing experiential learning at HKU (Wise Assessment Briefings, No. 5)*. Hong Kong, China: University of Hong Kong, Centre for the Enhancement of Teaching and Learning. Retrieved from www.cetl.hku.hk/wise-assessment-forum/

Chapter 4
Inspiring Transformation Through TESTA's Programme Approach

Tansy Jessop

Abstract This chapter explores evidence from 75 undergraduate degree programmes at 14 UK universities about students' experience of assessment and feedback. The data was collected through the *Transforming the Experience of Students Through Assessment* (TESTA) project. TESTA illuminates students' whole programme experience of assessment and feedback within the context of modular curriculum design. The methodology consists of an audit to ascertain dimensions of the assessment environment and focus groups with students. Analysis explores the relationship between assessment design and students' lived experience. Findings show the prevalence of high summative and low formative assessment diets and disconnected feedback which students find difficult to use. High summative diets reinforce students' instrumental approach to learning. A lack of formative assessment impacts on engagement in learning, diminishing opportunities for risk-taking, creativity and wider reading. Findings from TESTA have prompted educationally principled strategies: rebalancing the number of summative and formative tasks with greater connections between them, devising formative assessment tasks valued by both students and staff and designing feedback to feedforward. Wider implications include establishing institutional mechanisms to ensure principled, evidence-based and programme-focused assessment and feedback design within existing curriculum processes.

Introduction

This chapter is based on triangulated data about programme assessment from the *Transforming the Experience of Students Through Assessment* (TESTA) project (TESTA, 2015). TESTA started as a 3-year funded UK Higher Education Academy (HEA) National Teaching Fellowship Project (2009–2012) to investigate programme-level assessment on seven programmes in four similar small universities. The purpose of the research was to explore the impact of modular

T. Jessop (✉)
Southampton Solent University, Southampton, UK
e-mail: tansy.jessop@solent.ac.uk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_4

systems on assessment and feedback design and consequently on how students learn. By 2016, 4 years after funding ceased, more than 50 universities in the UK and universities in Australia, India, Canada, the USA and South Africa have drawn on TESTA's approach. This is testimony to the sector's appetite for understanding whole programme assessment patterns, but also to the value of TESTA's research methodology and its participatory change process.

Modular assessment has come under criticism for fragmenting the curriculum, fostering a lack of connection and coherence and blurring chronological progression. The introduction of semesters and modular curricula has compressed learning into short, contained units to the extent that 'slow learning' and formative assessment are squeezed out (Harland, McLean, Wass, Miller, & Sim, 2014; Knight, 2001; Knight & Yorke, 2003; Rust, 2000). TESTA provides an evidence base which gives insight into the impact of modular assessment environments on student learning. It enables programme teams to redesign assessment and feedback holistically, with articulation between the evidence, assessment principles and quality assurance frameworks (Gibbs & Dunbar-Goddet, 2007, 2009; Jessop, El Hakim, & Gibbs, 2011, 2014; Jessop, McNab, & Gubby, 2012). TESTA's approach has brought credible evidence to sector-wide discussions about developing a strategic focus on programme assessment and feedback (Bloxham & Boyd, 2007; Knight, 2000; PASS, 2009–2012).

The following sections outline TESTA's research methodology and explore key findings, with examples of best practice arising from undertaking the change process. It concludes with a strategy for scaling up assessment transformation institutionally.

Research Methods

Previous studies have used various combinations of the three TESTA methods to provide a critical perspective of degree programme environments. A previously published TESTA study triangulated the audit, the Assessment Experience Questionnaire (AEQ) and focus group data to provide an analysis of programme assessment environments (Jessop, El Hakim, & Gibbs, 2013). A follow-up study triangulated audit and AEQ data to illuminate disciplinary assessment practices (Jessop & Maleckar, 2014). Since the publication of these studies, based on smaller data sets (23 and 18 programmes, respectively), many more programmes have undertaken TESTA across the UK. This chapter analyses data through the lens of two of the three TESTA methods, namely, the audit and focus groups. The rationale is to add new knowledge based on a larger sample than previous studies which relied heavily on statistical methods (Jessop et al., 2013, 2014). This chapter gives a distinctive qualitative perspective on programme assessment environments using large-scale data, by triangulating the hard count data of the TESTA audit with student voice data from focus groups.

The TESTA audit distinguishes elements of the assessment environment (Gibbs & Dunbar-Goddet, 2009). It consists of discussion with the programme leader over course documents to map the key features of assessment and feedback over 3 years of an undergraduate programme (Jessop, 2010a). These features are:

- Number of summative assessments
- Number of formative assessment tasks
- Number of different varieties of assessment types
- Proportion of the assessment diet by examinations
- Time it takes to return marks and feedback
- Amount of written feedback over 3 years
- Amount of oral feedback over 3 years

The programme audit represents the 'planned curriculum' (Stenhouse, 1975), providing hard count data about key features of students' experience over 3 years of an undergraduate degree. In most UK universities, the definitive documents of a programme normally undergo revision and scrutiny every 5 or 6 years, with modules, assessments, learning outcomes and the content of a programme being revised in the light of institutional, staffing and subject developments. This process, known as periodic review, provides programmes with external and internal scrutiny to assure universities and students of the quality of provision. While this documentation is publically available, the TESTA audit brings to light the less visible aspects of a programme which may not be contained in the documents, for example, formative assessment tasks. The audit is a mixed methods approach drawing on documentary and interview evidence.

Focus groups provide the second vein of data for this study. The hallmark of focus groups is their 'explicit use of group interaction to produce data and insights that would be less accessible without the interaction found in the group' (Morgan, 1997, p. 2). TESTA focus groups prompt discussion on four themes: the assessment diet, feedback, the influence of assessment on study habits and students' understanding of goals and standards (Jessop, 2010b). These themes are hospitable and allow students to discuss wider issues about assessment and feedback on their programme. Typically a focus group consists of an hour-long discussion between a group of between five and eight final year students, facilitated by a researcher.

Sampling of programmes varies institutionally. At Winchester, TESTA began as a voluntary exercise which programmes signed up for out of interest or the desire to enhance their programme's assessment. Since 2013, TESTA has been embedded in cyclical quality processes, so that all undergraduate programmes undergoing six yearly periodic reviews are required to participate in TESTA. This evidence gathering and planning informs curriculum design. Other institutions have nominated programmes to participate; alternatively whole departments, faculties and programmes have signed up (or been signed up) to participate in TESTA.

Data analysis of the audit is a relatively simple process of transposing the data, recorded on a flipchart, into a brief report which summarises occurrences of assessment, and volumes of feedback, for example. The rigour of the audit is ensured through investigator triangulation (Cohen, Manion, & Morrison, 2007) when more

than one researcher is eliciting the data. More often, the accuracy of interpretation is ensured through 'member-checking' (Lincoln & Guba, 1985) which involves the draft audit document being sent to the programme leader to check the accuracy of information and interpretation.

Focus group data are recorded and transcribed. The transcripts are uploaded into qualitative data analysis software (Atlas.ti), for the purpose of thematic analysis. The researcher codes units of meaning, using either thematic coding (drawing on existing constructs about how students learn from feedback) or generative coding (open to new constructs from the data). An example of a thematic code might be 'confusing criteria', with criteria being an explicit area of investigation. In contrast, a generative code might be 'marker variation', which students raise as an implicit barrier to understanding goals and standards. The levels of codes also vary between concrete descriptions to more abstract notions such as strategic or instrumental approaches to learning. There is a guide to coding focus group data on the TESTA website (Jessop, 2011).

Normally, the TESTA process consists of representing audit, Assessment Experience Questionnaire and focus group data in a case study which is discussed with programme teams. The case study is the focal point for developing strategies to enhance curriculum design and pedagogic practice. In this chapter, the key findings draw only on audit and focus group data. The audit data provide a summary of the numbers/proportions of certain assessment and feedback activities, in ranges and medians, across the whole sample. As in any qualitative research with large volumes of textual data, coding segments of data underpins the process of developing themes. The key findings in this chapter represent recurrent themes, derived from a systematic qualitative analysis of transcripts from participating programmes.

Key Findings

Three main themes in TESTA data are now explored. The first is the variation in assessment patterns and implications for student learning. The second demonstrates the prevalence of high summative assessment diets in contrast to low occurrences of formative assessment. Thirdly, the data sheds light on episodic and haphazard feedback which does not connect to the next task or across modules. These themes all impact on the student learning experience, contributing either to surface or deep learning (Marton & Saljo, 1976), or strategic behaviour (Miller & Parlett, 1974). Student alienation, characterised by 'playing the game' in a performative way, arises partly from flaws in the design of assessment and feedback (Boud & Molloy, 2013; Mann, 2001; Miller & Parlett, 1974).

Before discussing the themes, it is worth clarifying how the terms formative and summative assessment are used in this chapter. TESTA defines summative assessment as tasks which are graded and count towards the degree either as pass/fail or as grades; in contrast, formative tasks do not count towards the degree, are required to be done by all students and elicit feedback. Formative assessment has been described as a 'fuzzy concept', and its elusiveness has stimulated much debate

and contestation (Taras, 2008; Torrance, 2012; Yorke, 2003). While recognising the problematic nature of defining formative and summative assessment, TESTA adheres to Shepard's distinction between summative and formative: 'summative assessment measures students' achievement by a grade, while formative gives qualitative insights about students' understandings and misconceptions to improve learning' (Shepard, 2005). This distinction is founded on the belief that formative generally plays a different role to summative, privileging reflection and action on feedback, and orienting students towards future performance. In contrast, summative assessment often occurs at the end of modules and orients students towards grades rather than future performance, which may occur in an unrelated area of study, albeit the same discipline.

The key findings section expands on each theme using data from TESTA. Following each theme, there are examples of best practice drawn from programmes and institutions which have engaged with the change process in TESTA. Strategies to implement best practice emerge from a rich discussion of the data in the light of assessment principles.

Theme 1: Variations in Assessment Environments

Undergraduate degree programmes demonstrate extreme variations in their assessment environments. The disciplines represented among the 75 within the sample include:

- Pure sciences (e.g. mathematics, chemistry)
- Applied sciences (e.g. engineering, pharmacy)
- Humanities and social sciences (e.g. history, sociology)
- Applied 'soft' disciplines (e.g. education, social work)
- Creative subjects (e.g. drama, dance, creative writing)

The following table shows the range and medians of TESTA audit data across a 3-year programme of study within the sample of 75 programmes (Table 4.1).

Table 4.1 Ranges and medians for TESTA audit data (n=75 programmes)

Characteristic	Range	Median
Number of summative assessments	12–227	43
Number of formative assessments	0–116	5
Varieties of assessment	5–21	11
Proportion of tasks by exam	0–87 %	20 %
Time to return marks and feedback	10–42 days	21 days
Volume of oral feedback	37–1800 min	282 min
Volume of written feedback	936–22,000	6165 words

Variations may occur because of different 'ways of thinking and practicing' in the disciplines (Hounsell & Anderson, 2009). Research on 'signature pedagogies' (Shulman, 2005) illustrates the ways in which different professions and disciplines enact their knowledge, shown through 'idiosyncratic organisation, set of artefacts, assumptions and practices peculiar to learning and teaching in the discipline' (Donald, 2009, p.40). In the sciences, for example, assessment patterns often contain small and frequent formative tasks, partly to ensure that students master concepts incrementally, and to avoid gaps in understanding linked to the next concept (Jessop & Maleckar, 2014).

The curriculum design process also contributes to variations. Aside from more tightly regulated professional programmes, most lecturers exert a significant degree of autonomy in the design of assessment tasks on modules (Bridges, 2000). Programmes are commonly assembled module-by-module according to the content being covered, without an eye on the whole programme's assessment design. Module leaders often design assessment tasks in isolation and without connection to the wider programme (Jessop et al., 2012).

TESTA has demonstrated that variations in assessment environments exist on a scale that requires attention to ensure comparability in the student learning experience. The most striking variations occur in the number of summative and formative assessments, the proportion of exams and the amount of feedback students typically receive. Variations are likely to influence study behaviours. For example, a high proportion of summative assessment with low formative has been shown to lead to narrowly focused effort, strategic behaviour and a lack of deep learning, evidenced in focus group comments and by low scores on the Quantity of Effort and Deep Learning Scales on the AEQ (Jessop et al., 2013). Many small and frequent summative assessment tasks foster instrumental and grade-conscious approaches to learning (Harland et al., 2014). Bite-sized small tasks may also lead to superficial learning as students are not challenged to undertake large, independent tasks which integrate and connect learning from across the programme (Ashford-Rowe, Herrington, & Brown, 2013; Harland, McLean, Wass, Miller, & Sim, 2014, Jessop et al., 2013). Students also describe incidences of marker variation, which are evidenced in low scores on the Clear Goals and Standards Scale on the AEQ (see Jessop et al., 2013).

TESTA has helped institutions and programmes to address variations in practice within individual universities. The following case studies illustrate actions taken to reduce variations, in the interests of parity of the student experience:

Case Study 1: Addressing variations at the institutional level
The Problem: Students experience widely differing assessment and feedback practice, particularly with some programmes having extremely high summative demands alongside invisible, uneven or non-existent formative tasks.

(continued)

There is a huge variety of randomly sequenced types of assessment across the degree with confusing demands for students.

Strategy: All programmes go through TESTA at periodic review. TESTA demonstrates how many summative assessment tasks a student will experience, a figure which is not self-evident from the modular nature of the documents. The concept of a reasonable programme assessment load is discussed against the backdrop of sector-wide data and assessment principles, with the aim of rebalancing the ratio of summative and formative. Appropriate and well-designed formative assessments are written into the documentation and planned to articulate with summative tasks. Varieties of assessment are sequenced through the degree, in many cases streamlining these so as to enable a coherent journey of learning through the degree.

Case Study 2: Variations within a programme

The Problem: Students describe wide variations between markers on a programme, characterised by some lecturers marking harshly, others more generously. Student focus groups evidence perceptions of bias, varying styles of marking, different approaches and standards.

Strategy: Calibration of standards

Facilitate a programme calibration exercise involving all team members. Begin by discussing criteria using flipchart and discussion; compare with existing written criteria and discuss tacit criteria and standards. Mark two or three anonymous written pieces. Collect marker's individually assigned marks before a round-table discussion. This negates programme team power relations and hierarchies, ensuring that markers commit to their marks before discussing why they have assigned them. Display the range of marks. Discuss marks in relation to the agreed criteria and come to consensus. The discussion should begin to articulate more of a common standard. Repeat with a different piece. Repeat calibration annually.

Note: Calibration is different from moderation. It is a much more all round, relaxed and open discussion of team standards. It is not linked to time-bound marking processes and formal institutional quality assurance processes.

Theme 2: High Summative and Low Formative Assessment

TESTA audit data shows that students typically encounter 43 summative tasks over 3 years, while experiencing only five formative tasks. This is a ratio of 8:1 of summative to formative tasks. In contexts of high summative assessment, students frequently encounter many, bite-sized summative tasks, which may distribute effort

but not challenge them; these students often become strategic performers in an 'assessment arms race' (Harland et al., 2014), depleting their capacity for deep learning. On the surface, they will have worked harder than other students, but their learning is likely to have been trivialised by small, frequent and narrowly focused tasks. In focus groups, students describe how summative assessment dominates their study behaviour and narrows their focus. Below are quotations which exemplify these issues:

> A: A lot of people don't do wider reading. You just focus on your essay question.
>
> B: I always find myself going to the library and going 'These are the books related to this essay' and that's it (Archaeology).
>
> If someone said what did you learn on your degree, I'd basically sum it up as saying I learnt what I needed to learn for assignments; I learnt what I learnt because I needed to complete an assignment, rather than I learnt because I was really interested in the whole thing (English Language Studies).
>
> In Weeks 9 to 12 there is hardly anyone in our lectures because we're too stressed. I'd rather use those two hours of lectures to get the assignment done (Theology and Religion).

While summative assessment tasks may be learning oriented (Carless, 2007, 2015), TESTA data shows that this is difficult to achieve across compartmentalised modules designed in an atomised way. Students describe the timing and volume of assessment as interfering with learning-oriented assessment, especially when too much assessment crowds out reflection and fosters a grade-driven approach (Knight & Yorke, 2003; Lizzio, Wilson, & Simons, 2002). In the context of high summative diets, an instrumental culture flourishes, because students focus on achieving grades in the shadow of ever-present deadlines (Harland et al., 2014; Jessop et al., 2013). Modular design has multiplied the number of assessment tasks and disconnected them from one another across modules. Students indicate how stressful and demotivating a succession of summative tasks can be, replacing assessment *of* learning with assessment *as* learning and bypassing assessment *for* learning (Torrance, 2007):

> The quantity of assessed work is very tiring. We'd rather genuinely study the subject (Education).
>
> It's been non-stop assignments, and I'm now free of assignments until the exams – I've had to rush every piece of work I've done (History).
>
> There was a full two weeks of madness, because there was the poster submission, thesis, then the poster presentation, then the exam. It was a very stressful period at the time. Motivation was hard to come by (Pharmacy).

Focus group and audit data imply an unspoken agreement between academics and students that summative assessment is the main way to drive student effort. Without the incentive of a grade, some students say that they are disinclined to undertake academic work. As one student wryly observed, 'The lecturers have the problem of actually getting the students to go away and learn. If it is not being assessed, if there're no actual consequences of not doing it, most students are going to sit in the bar' (Computing Student). The audit data bears out a tentative embrace of formative tasks in programme design, with lecturers underplaying it to the extent that one

in five programmes contains zero formative assessment, and students typically experience less than two formative tasks each year.

Extremely low formative assessment occurs in spite of overwhelming evidence in the literature of its effectiveness in helping students to learn from assessment. Black and Wiliam's large-scale analysis of factors influencing learning concluded that 'innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains' (Black & Wiliam, 1998, p. 40). The reasons for formative assessment's effectiveness include its capacity for 'short-circuiting the randomness and inefficiency of trial-and-error learning' (Sadler, 1989, p. 120), and its capacity to help students fine-tune their work by coming to a deeper understanding of goals and standards (Boud, 2000; Nicol & Macfarlane-Dick, 2006).

Students prioritise assessment tasks which count towards their degrees. In the context of high summative assessment demands, it is unsurprising that formative tasks are undervalued when they compete for time and effort with tasks which count. The combination of high summative demands on concurrent modules, and short semesters, make it almost inevitable that formative tasks are squeezed out. Comments from students in focus groups evidence these competing priorities:

> It didn't count for anything, so if you didn't do it, it didn't matter (Mathematics).
>
> It's a little bit pointless for me because I'd rather put all my energy and efforts into marked ones and get really high grades on them and not bother with the others (Philosophy).
>
> What is the point of putting that much effort in when I have this much time to do an assessment that counts towards my degree? I find it really frustrating when people ask for ten page reports and presentations which don't count and I am thinking why am I doing this?! It's brilliant practice but... (Business and Management).

The low value of formative is compounded by issues with the distribution of assessment in compact semesters. Most programmes have two assessment points per module, which cluster at the mid- and end points of modules. The timing of formative tasks is a complex and important aspect of design, influencing students' capacity to engage with formative tasks.

The following case studies demonstrate actions which programme teams have taken to address the challenges of high summative and low formative assessment diets:

Case Study 1: Rebalancing summative and formative

Problem: TESTA uncovers that several programmes in a business school have a typical assessment load of 48 summative tasks. Students are working for each assessment *as* learning, ignoring wider reading and set tasks.

Strategy: Departmental decision is made by the head of school to revalidate all programmes with mandated limits of one summative assessment on each module and three formative tasks leading up to the summative. Timing and sequencing of all assessments are agreed to prevent competing clashes. All

(continued)

programmes move to a summative assessment load of 24 tasks and a ratio of 3:1 of formative to summative.

Design complexities: There is variation in the success of the formative design because some formative 'teaches to the test', with the result that students fine-tune similar work to the summative to achieve better grades. Strategies are put in place to support lecturers in the design of formative which synchronises with summative but is challenging and stand-alone, yet conceptually linked to the summative.

Feedback complexities: Formative work is often peer reviewed in class, so that lecturers do not increase marking loads by having four marking occurrences (3 × formative plus 1 × summative) instead of two as in the past (2 × summative). Best practice ensures discussion and dialogue about marking and student use or co-creation of criteria.

Case Study 2: Students as producers of formative work

Problem: TESTA shows ineffective formative tasks which are only done by keen students or 'dashed off' by students who do not value formative work, particularly when it competes with summative demands on other modules.

Strategy: Whole programme strategy is adopted to reduce summative assessment load on all modules. Well-designed formative tasks command student attention and interest, replacing certain summative tasks. Principles of good formative assessment practised as a result of undertaking TESTA include as follows:

(a) Public facing so as motivate students to perform and contribute to knowledge generation.
(b) Links to the summative task in a challenging and enriching way.
(c) Builds in elements of collaboration and accountability.
(d) Involves challenging research, theory or project design.
(e) Links to the discipline and has a genuine purpose.
(f) Encourages students to be creative and take risks.
(g) Engenders feedback from peers or the tutor and encourages reflection and inner dialogue.
(h) Students are required to undertake it as a gateway to completing the summative task.

Idea 1: Blogging as formative

Students blog fortnightly in class on academic readings. In alternate weeks, students read several fellow students' blogs and spend the in-class hour commenting on posts. The summative assessment is constructed around

(continued)

conceptual understandings developed through the blogging or may be a synthesis of different arguments and positions. 'Think aloud' data on blogging shows that it prompts engagement with academic texts, reflection, distribution of effort and deep learning (Jessop, 2015).

Idea 2: Project development and design

Assessment consists of large, challenging, collaborative (or individual) projects which involve designing publically available artefacts, such as films or posters or a publically disseminated research project, to address an issue or problem in the discipline. In one case, the team of lecturers split a 12-week module into 6 × weeks of lectures and 6 × weeks of collaborative project work, guided in class. Students produced media artefacts which were showcased in the final weeks.

Theme 3: Disconnected Feedback

Written feedback volumes are calculated by sampling in-text and summary feedback from each cohort, from a variety of different markers. Typically, students receive around 7400 words of written feedback over a 3-year undergraduate degree. This represents a significant amount of time, thought and effort in crafting feedback, yet in the National Student Survey (NSS), students rate the effectiveness of feedback lower than any other aspect of teaching. Question 9 on the NSS 'Feedback on my work has helped me clarify things I did not understand' is consistently the lowest scored of the assessment and feedback questions. In England, only 66 % (2014) and 67 % (2015) of full-time undergraduates on taught courses affirmed that feedback helped them to clarify things they did not understand (HEFCE, 2015). At least one third of feedback is not working.

Evidence from TESTA gives some explanation for the broken state of feedback. It sheds light on the waste of resources when feedback ends in a cul-de-sac. Feedback which is end loaded, occurring after the module has finished, has little chance of improving student performance. Audit and focus group data show a lack of articulation of feedback across modules, with few opportunities for feeding forward designed into the process. One-off, episodic and piecemeal feedback which 'dangles the data' (Sadler, 1989) has questionable value when it is not designed as a process which helps students to reflect and act on it with a future orientation (Boud & Molloy, 2012, 2013). In a modular system, students are more inclined to compartmentalise their learning and ignore feedback which does not have much chance of feeding forward, as these comments demonstrate:

> It's difficult because your assignments are so detached from the next one you do for that subject. They don't relate to each other (Media Studies).
>
> It is so dependent on whether it is the first assignment for the module or not . . . say you have a fifteen credit module where you have two essays and it's the first essay back,

then you would probably take quite a lot on board but then for the essay after that I was quite pleased with my marks so I just was like fine! As bad as it sounds, the module is over (Business).

Because it's at the end of the module, it doesn't feed into our future work (History).

The feedback is generally focused on the module (Primary Education).

I read it and think 'Well, that's fine but I've already handed it in now and got the mark. It's too late' (Creative Writing).

Through doing TESTA, programme teams have developed strategies for making feedback connect across modules, exemplified below:

Case Study 1: Building reflection into feedback in a way which leads to action

Problem: Students do not make use of end-of-module feedback unless there is a problem with their mark.

Strategy: Programmes give students feedback on ways to improve their work, which students are required to reflect on in writing in their next task. The lecturer will not mark the next task unless students have shown how they have addressed previous feedback.

Case Study 2: Building dialogue into feedback in a way which leads to action

Problem: Students feel that they are the passive victims of feedback which is a one-way transaction from expert to novice.

Strategy 1: Students open the conversation by indicating what they would like feedback on, for example, targeted towards what they feel they understood or argued well or where they feel conceptually fuzzy. Markers respond to the conversation with feedback. This is dialogue!

Strategy 2: Markers release comments only. Students need to write a brief response to feedback and attend a tutorial to receive their mark. Students have reflected on and engaged in dialogue about their feedback without being distracted by the mark.

Strategy 3: Markers give generic feedback about strengths, weaknesses and issues based on a sample of marking very soon after the hand-in date. This helps students to think through their own practice from memory, prompting inner dialogue.

Case Study 3: Integrated assessment across modules

Problem: Students make little use of end-of-module feedback because it is too late, or modules are viewed as 'finished business'.

(continued)

Strategy: Curriculum design makes more explicit links and connections across modules through a capstone assessment which threads across several modules. This may be a reflective portfolio, a group project or a research paper.

Scaling Up Transformation: An Embedded Institutional Approach

TESTA provides an example of a systematic approach to transforming the student experience of assessment and feedback. It is based on reconceptualising assessment and feedback design as integrated, connected and sequential within curriculum design processes *at the programme level*. There are four key dimensions to bringing about institutional transformation in assessment and feedback to improve student learning, developed through 6 years of implementing TESTA's research and change process. These are:

- Taking a whole programme view of assessment and feedback
- Using an evidence-led approach with a strong element of student voice
- Putting the evidence and principles in the hands of teams to make changes
- Adopting a systemic approach through quality assurance processes

A whole programme approach to assessment and feedback prompts two shifts in perception: firstly, lecturers shift from thinking about curriculum and pedagogy from the silo perspective of 'my' module to a connected view of 'our' programme. Secondly, lecturers start to understand what learning looks like from the programme-wide experience of a student juggling modules and deadlines, in contrast to the funnelled view of seeing students as participants on 'my' module. The programme view enables a shift from a teacher-centred to a student-centred paradigm of learning from assessment and feedback.

The second dimension is TESTA's robust evidence-base, triangulating audit, survey and focus group data in rich and textured case studies. Programme teams describe the data as plausible and compelling. It validates hunches and intuitions about the assessment environment by bringing particular evidence into focus. The evidence may be challenging and discomforting, but teams are able to engage with it because of its externality, its robustness and its genuine enhancement focus. The way the case studies are constructed ensures that the student voice data is prominent, rather than wrapped up in authorial interpretation (Jessop, 2013; Richardson, 1990).

The third important element in scaling up institutional change through TESTA has been placing the evidence in the hands of the team in order for them to develop holistic curriculum and pedagogic strategies. This respects the autonomy, agency and disciplinary knowledge of programme teams. Meeting over evidence takes the

shape of a consultancy briefing, with rich discussion, contestation, new ideas and 'what if' questions being raised. It is fertile ground on which to base curriculum design and pedagogic decisions.

The final dimension of assessment transformation using TESTA has been to scale it up institutionally, by embedding it in quality assurance processes. Systematic embedding of TESTA in cyclical periodic review contributes to evidence-led curriculum design on all revalidating degrees. This model drives institution-wide and programme-specific enhancement effects at the Universities of Winchester and Dundee, with further universities exploring its use. As more programmes engage at Winchester, deans of faculty describe a 'TESTA effect': a new appreciation of formative assessment, coupled with principled and evidence-based assessment design. The chair of the UK's assessment in HE conference has described TESTA as 'the only thing I can find that seems to be making systematic headway beyond the individual module' (S. Bloxham, personal communication, 17 March, 2016). Isolating reasons for TESTA's systematic headway is complex, but two aspects stand out: the focus on the programme and the marriage between an enhancement project and a quality assurance process. These signify TESTA's potential to impact the institutional assessment culture, in particular, its capacity to create a powerful learning experience for students through systematic transformation.

References

Ashford-Rowe, K., Herrington, J., & Brown, C. (2013). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education, 39*(2), 205–222.

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

Bloxham, S., & Boyd, P. (2007). Planning a programme assessment strategy. In *Developing effective assessment in higher education* (pp. 157–175). Maidenhead, UK: Open University Press.

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

Boud, D., & Molloy, E. (Eds.). (2012). *Feedback in higher and professional education: Understanding it and doing it well*. New York: Routledge.

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

Bridges, D. (2000). Back to the future: The higher education curriculum in the 21st century. *Cambridge Journal of Education, 30*(1), 37–55.

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

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.

Cohen, P. L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York: Routledge.

Donald, J. G. (2009). The Commons: Disciplinary and interdisciplinary encounters. In C. Kreber (Ed.), *The university and its disciplines: Teaching and learning within and beyond disciplinary boundaries* (pp. 35–49). New York: Routledge.

Gibbs, G., & Dunbar-Goddet, H. (2007). *The effects of programme assessment environments on student learning*. Higher Education Academy Report. Retrieved from https://www.heacademy.ac.uk/sites/default/files/gibbs_0506.pdf

Gibbs, G., & Dunbar-Goddet, H. (2009). Characterising programme-level assessment environments that support learning. *Assessment & Evaluation in Higher Education, 34*(4), 481–489.

Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2014). An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education, 40*(4), 528–541.

Higher Education Funding Council for England (HEFCE). (2015). *2015 – Higher education funding council for England*. Retrieved from http://www.hefce.ac.uk/lt/nss/results/2015/

Hounsell, D., & Anderson, C. (2009). Ways of thinking and practising in biology and history: Disciplinary aspects of teaching and learning environments. In C. Kreber (Ed.), *The university and its disciplines: Teaching and learning within and beyond disciplinary boundaries*. New York: Routledge.

Jessop, T. (2010a). *Ten steps to auditing a programme*. Retrieved from http://www.testa.ac.uk/index.php/resources/research-tool-kits/category/11-researchtoolkits

Jessop, T. (2010b). *Focus group questions*. Retrieved from http://www.testa.ac.uk/index.php/resources/research-tool-kits/category/11-researchtoolkits

Jessop, T. (2011). *Coding focus group data*. Retrieved from http://www.testa.ac.uk/index.php/resources/research-tool-kits/category/11-researchtoolkits

Jessop, T. (2013). *TESTA case studies*. Retrieved from http://www.testa.ac.uk/index.php/resources/case-studies/category/8-case-studies

Jessop, T. (2015, December). *Flipping the technology-pedagogy equation: Principles to improve assessment and feedback*. Seminar presentation at the University of Sussex, Brighton, UK.

Jessop, T., El Hakim, Y., & Gibbs, G. (2011). TESTA: Research inspiring change. *Educational Developments, 12*(4), 12–16.

Jessop, T., El Hakim, Y., & Gibbs, G. (2013). The whole is greater than the sum of its parts: A large-scale study of students' learning in response to different programme assessment patterns. *Assessment & Evaluation in Higher Education, 39*(1), 73–88.

Jessop, T., El Hakim, Y., & Gibbs, G. (2014). TESTA 2014: A way of thinking about assessment and feedback. *Educational Developments, 15*(2), 21–23.

Jessop, T., & Maleckar, B. (2014). The influence of disciplinary assessment patterns on student learning: A comparative study. *Studies in Higher Education, 41*(4), 696–711.

Jessop, T., McNab, N., & Gubby, L. (2012). Mind the gap: An analysis of how quality assurance procedures influence programme assessment patterns. *Active Learning in Higher Education, 13*(3), 143–154.

Knight, P. T. (2000). The value of a programme-wide approach to assessment. *Assessment & Evaluation in Higher Education, 25*(3), 237–251.

Knight, P. T. (2001). Complexity and curriculum: A process approach to curriculum-making. *Teaching in Higher Education, 6*(3), 369–381.

Knight, P. T., & Yorke, M. (2003). *Assessment learning and employability*. Maidenhead, UK: Society for Research into Higher Education & Open University Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*(1), 27–52.

Mann, S. J. (2001). Alternative perspectives on the student experience: Alienation and engagement. *Studies in Higher Education, 26*(1), 7–19.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: Outcome and process. *British Journal of Educational Psychology, 46*(1), 4–11.

Miller, C. M. L., & Parlett, M. R. (1974). *Up to the mark: A study of the examination game*. London: Society for Research into Higher Education.

Morgan, D. L. (1997). *Focus groups as qualitative research*. London: Sage Publications.

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.

Programme Assessment Strategies Project (PASS). (2009–2012). *University of Bradford*. Retrieved from http://pass.brad.ac.uk

Richardson, L. W. (1990). *Writing strategies: Reaching diverse audiences*. Newbury Park, CA: Sage Publications.

Rust, C. (2000). A possible student-centred assessment solution to some of the current problems of modular degree programmes. *Active Learning in Higher Education, 1*, 126–131.

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

Shepard, L. A. (2005, October). *Formative assessment: Caveat emptor*. Paper presented at 'The Future of Assessment: Shaping Teaching and Learning' Conference, New York, NY.

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus, 134*(3), 52–59.

Stenhouse, L. (1975). *An introduction to curriculum research and development*. London: Heinemann Educational Books.

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education, 9*(2), 172–192.

Transforming the Experience of Students through Assessment (TESTA). (2015). The University of Winchester. Retrieved from http://www.testa.ac.uk

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

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*(3), 323–342.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*, 477–501.

Part II
Assessment for Learning Strategies and Implementation

Chapter 5
Making Assessment for Learning Happen Through Assessment Task Design in the Law Curriculum

Rick Glofcheski

Abstract Across the disciplines in higher education, too little attention is paid by those who design and deliver courses to the role of assessment as a driver of learning. This is certainly the case in legal education. A lecture-based, teacher-centred approach predominates, which produces a largely passive learning, an approach that is reflected in the assessment. The emphasis is on doctrinal instruction, issue coverage, accreditation and ranking. Thus, there is plenty of scope for scaling up. In this chapter, the author describes the principal method of learning and assessment in law schools and the modest learning outcomes it can produce. The author proposes some simple strategic moves in assessment design that can expand the range of achievable learning outcomes in legal education and facilitate the development of skills necessary for professional life. These moves involve the adoption of authentic materials for use in learning and assessment and the introduction of task-based assessments in which students take the lead role in the construction and management of their learning artefacts. They are simple and economical, can be applied in large classes and have the potential for adaptation across the disciplines.

Introduction

The teaching and learning of law presents a case that is particularly ripe for the scaling up of assessment for learning. That is because legal education is steeped in traditions that have proved resistant to change, and there is as yet only limited understanding of assessment for learning in law schools. This has happened in part because of a long-held belief in the distinctiveness of law as a subject of study and a commitment to teaching traditions that law teachers themselves were schooled in (Sturm & Guinier, 2007). In designing assessment, teachers take their cue from their own teachers and from their perceptions of professional expectations.

R. Glofcheski (✉)
Faculty of Law, University of Hong Kong, Hong Kong, Hong Kong
e-mail: rickg@hku.hk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_5

The psychology and theory of learning have had little influence on legal educators and the design of law curricula (Schwartz, 2001). As a consequence, not much has changed in a century (Rankin, 2011).

A major tension within legal education has to do with its very function. Should educators concentrate on preparing students for the profession, or should they provide a liberal university education? Is it their mission to offer training or education? Although legal academics generally take the latter position (LeBrun & Johnstone, 1994; Bradney, 2003; Rochette, 2011), this schism has produced curricula that may not do either very well.

Another tension has to do with the role of assessment. This arises in large part from the influence of the profession. When assessment enters the conversation in legal education, it is more in the context of its gatekeeper function and the age-old (albeit understandable) concern with standards and competencies. The public should be protected from graduates who cannot reach expected standards (Sullivan, Colby, Wegner, Bond & Shulman, 2007).

Among the questions and concerns addressed in this chapter are the following: Do current assessment practices serve learning goals very well? Does assessment bear much relation to what law graduates do in their careers? Can assessment be used more strategically to achieve more meaningful and sustainable learning? Can it be done in large law classes within existing resources? What is the role of assessment for learning in the design of teaching and learning in law?

The author will argue that much can be achieved through the scaling up of assessment for learning in law, even by the individual instructor. There is scope for scaling up assessment for learning in many respects – among them, diversity of assessments, feedback processes and assessment literacy. Not all can be considered here. Of the four main assessments for learning strategies identified by Carless in Chap. 1 of this volume, the focus here is on the first – productive assessment task design. The author focuses on two aspects of assessment for learning for scaling up: authenticity in assessment and aligned project or activity-based assessment.

Legal Education Today

Legal education relies heavily on doctrinal instruction. Doctrinal instruction relies on the "case method", introduced at Harvard University in the late nineteenth century (Stuckey et al., 2007). In this model, students read case reports, usually written by appellate courts. In lectures, the case reports are subjected to a critical reading by the teacher, to uncover their meaning and their contradictions internally and across the body of case law (Schwartz, 2001). Student classroom participation is limited. In the US classroom, teachers employing the so-called Socratic method will isolate an individual student for interrogation, while other students observe. In law schools in the rest of the common law world, a less aggressive approach is favoured, and any student participation is more likely to take the form of voluntary, albeit infrequent, Q&A. In all such classes, the main activity of students is note-taking.

Waye and Faulkner (2012), describing two different but influential systems of legal education, confirm that not much has changed:

> Despite the long entrenched rhetoric of student-centred learning, in reality most legal education in Australia, as in the United States, is homogenised and monologist, rather than dialogic and transformational. Students are generally expected to absorb material delivered by lecturers or online in text form, then apply it to hypothetical problems in situations far removed from the professional practice they are likely to encounter upon graduation. (p. 563)

This educational model, as explained by the author elsewhere (Glofcheski, 2015), is adopted by teachers from their own teachers, perpetuating the same method over the generations. This is a pedagogy premised on the somewhat crude notion of knowledge transfer: the lecturer transmits knowledge that students are required to master and apply on final examinations (Biggs & Tang, 2011). It is sometimes explained as a modelling function. Students, observing their teachers, will learn to "think like a lawyer" (Sturm & Guinier, 2007). This notion of thinking like a lawyer has permeated legal education for more than a century, although its meaning has never been entirely clear.

Recent reviews of legal education decry the emphasis placed on doctrinal instruction (Legal Education and Training Review (LETR), 2013; Stuckey et al., 2007; Sullivan et al., 2007), while other critics question the very notion of what it means to think like a lawyer (Hess, 2002, as cited in Stuckey et al., 2007), and whether legal education as currently configured can deliver that outcome (Sullivan et al., 2007). Two of the five key observations of the 2007 US-based Carnegie Institute's report on legal education are a propos:

- Compared to other professional fields, which often employ multiple forms of teaching through a more prolonged socialisation process, legal pedagogy is remarkably uniform across variations in schools and student bodies. (Summary, p. 5)
- Assessment of student learning remains underdeveloped. Summative assessments are useful devices to ensure basic levels of competence. But there is another form of assessment, formative assessment, which focuses on supporting students in learning rather than ranking. (Summary, p. 7)

Fortunately, this does not comprise the entire picture, and at many law schools, there are now other means of teaching and learning law, although confined mostly to upper-year elective courses with small cohorts. Such classes are often conducted in seminar format, providing greater scope for discussion and participation. In some of those courses, students conduct research and write essays, independently or under limited supervision, while attending classes. Many law schools have introduced experiential learning in the form of internships and clinical legal education, although, due to economies, these options are available only to a small number of students. Some more adventurous teachers have introduced learning portfolios, reflective diaries and the like (Waye & Faulkner, 2012). These initiatives are all to be commended as more likely to produce sustainable learning. But the mainstream of teachers continues the convention of teacher-centred lecturing, particularly so in the foundational compulsory courses in the first 2 years of study. Teachers are comfortable with the conventional methods, having been schooled in them.

Attempts at serious curricular reforms are rare, and inevitably resisted, at many levels (Glofcheski, 2015), by teachers, by the institution (Johnstone & Vignaendra, 2003) and even by students.

Assessment in Law

Assessment in law is not terribly imaginative. Boud's description is a propos: assessment is seen by most "almost exclusively as an act of measurement that occurs after learning has been completed, not as a fundamental part of teaching and learning itself" (Boud, 2006, p. xviii). The end-of-course, in-hall final examination is the main instrument (Rochette, 2011). Examination papers typically consist of questions designed for coverage and drawn from the teacher's imagination. Final examinations typically carry 70–100 % of the course assessment weighting. This weighting is likely to vary across jurisdictions. In a subject-specific survey conducted by the author in 2010, all 22 respondents from five jurisdictions confirmed that their final examination comprised on average 70 % of the course assessment, and for four of them, it was 100 % (Glofcheski, 2010a). Rochette (2011) in a survey of Canadian law teachers reports 17 % of respondents employing a 100 % weighted final examination. Where it carries less than 100 %, the balance is typically made up of short essay assignments, class participation and possibly a learning portfolio, though the latter is rare. Such heavily weighted final examinations are problematic, from a learning perspective, as discussed below. It is even doubtful whether they perform their assumed primary function – the evaluation of doctrinal knowledge – very well (Motley, 1985–1986). Among the many learning-related problems, such heavily weighted assessment tends to concentrate student work at the end of the course, leaves little scope for meaningful feedback and the possibility of "feedforward" and, as a single instrument for assessment, fails to accommodate students whose learning can be better demonstrated through other means. One positive development is the use by some teachers in upper-year courses of take-home final examinations, to be written over the course of a fixed but longer period of time than an in-hall examination. This has advantages in that there is greater scope for research and reflection, and it more closely resembles a real work setting. However, it poses problems of its own, including risks of plagiarism and collaboration (assuming collaboration is not allowed), and at any rate is not widely used (Rochette, 2011).

In recent reviews and studies of legal education, attributes and outcomes have been identified that more closely approximate what is expected of a sensitive, socially aware legal professional. The 2007 Carnegie study reports that:

> . . . a more adequate and properly formative legal education requires a better balance among the cognitive, practical, and ethical-social apprenticeships. To achieve this balance, legal educators will have to do more than shuffle the existing pieces. It demands their careful rethinking of both the existing curriculum and the pedagogies law schools employ to produce a more coherent and integrated initiation into a life in law. (Sullivan et al., 2007)

The recent studies call for law schools to reduce the emphasis currently placed on doctrinal instruction and to integrate the teaching of knowledge, skills and values, instead of treating them as separate subjects addressed in separate courses (Stuckey et al., 2007; Sullivan et al., 2007).

These studies represent a broad indictment of legal education as conventionally practised. The deficiencies identified are easy enough to comprehend, but can much be realistically expected from individual teachers, or even programme directors, given the absence of incentives and constraints such as those outlined by Carless in Chap. 1 of this volume. A major factor inhibiting the introduction of more meaningful assessment methods is class size. In large classes lecturing and assessment by final examination are easy to execute. How then might an assessment for learning approach contribute to the enhancement of assessment and learning?

Assessment for Learning

Assessment for learning places learning at the centre of any assessment activity. Assessment has a particular capacity to make a contribution to learning because we know students will work hard for success at it. Students are strategic (Entwistle & Ramsden, 1983) and will learn what they think they will be tested on (Biggs & Tang, 2011). From the students' perspective, assessment defines the academic agenda (Ramsden, 2003; Snyder, 1970). Assessment offers a learning moment, a powerful learning opportunity that should not be missed. Assessment can and should be designed such that the learning of students who participate in it will be advanced.

Given that students will do what is necessary for academic success, arguably any assessment would produce some learning. However, whether it does so effectively or whether it produces the right kind of learning is another matter and depends on careful assessment design, including alignment. To ensure that student effort contributes to the right kind of learning, the learning programme must be designed to achieve the learning outcomes recognized as important and valued in that discipline, but, importantly, it must be aligned with the assessment. This approach recognizes the importance of assessment and its capacity to advance learning.

An assessment for learning approach can have a positive influence on the teaching and learning of law, given the widespread use of the in-hall final examination and the hypothetical narrative. When preparing for such assessments, typical student practice is to become familiar with all the cases, and then practice writing answers to the kinds of hypothetical narratives expected from that teacher. An obvious problem with hypothetical narratives is that they are recognizable as hypothetical. They are often exaggerated and in the sequences presented bear limited if any resemblance to real-world occurrences. They are teacher-created and carry the expectation of a preconceived solution that the students must uncover. Students may focus on what the teacher was thinking rather than the problem itself. Moreover, as fictional creations these narratives are not, as a rule, conducive to the discussion of serious

social policy issues, a dimension of legal learning that critics of legal education see as important but lacking in current legal education (Watson, 2001).

Authenticity

Authenticity of assessment is a core feature of assessment for learning (Sambell, McDowell, & Montgomery, 2013) and can have a very positive influence on students' longer-term learning capacities. The role of higher education in equipping students to become lifelong learners is widely acknowledged.

> The raison d'etre of higher education is that it provides a foundation on which a lifetime of learning in work and other social settings can be built. Whatever else it achieves, it must equip students to learn beyond the academy, once the infrastructure of teachers, courses and formal assessment is no longer available. (Boud & Falchikov, 2007, p. 399)

Authentic learning and assessment focus on real-world, complex problems and their solutions (Lombardi, 2007). In authentic assessment, the cognitive demands are similar to those that starting professionals might be confronted with in their working life (Savery & Duffy, 1995). Moreover, authenticity of learning and assessment can increase student motivation. Students, aware that what they are learning is what they will do post-graduation, are likely to take their learning and assessment tasks more seriously.

The relevance and value of authentic learning, and its capacity to produce lifelong learners, have not been lost on legal educators. In the past half century or so, clinical legal education has taken hold in law schools around the common law world. Clinical legal education takes many forms, but most commonly involves students taking real cases and advising real clients, under the supervision of lawyers or law professors. The work is credit-bearing, taken as an elective in place of classroom-based courses, and assessment is often conducted on a pass-fail basis. One cannot imagine a more authentic and rich learning environment, and the law academy is to be credited with this initiative. However, such courses are expensive, requiring close supervision by teachers and lawyers. They are cost-inefficient when compared to classroom-based courses. Thus, clinical legal education is available only to a small number of students and is no panacea.

As already observed, the end-of-course, in-hall final examination is the most predominant form of assessment (Clegg, 2005; Glofcheski, 2015; Rochette, 2011) and is not likely to be dispensed with soon. Moreover, they are thought by some to have a higher degree of reliability than other more subjective forms of assessment (Race, 2006), although this was long ago questioned in the law context (Motley, 1985–1986; Stuckey et al., 2007). The chief characteristic of these law examinations is the teacher-invented hypothetical narrative. Naturally, students work hard to master this format, which unfortunately bears little resemblance to real-world problems. In this environment, the author, who taught a foundational compulsory law class of more than 250 students and employed the conventional final

examination format from the beginning of his teaching career, was at an impasse. A funded study, essentially a survey conducted by the author and colleagues into the learning habits of law students (Tai, Lee, & Glofcheski, 2006), revealed that students deliberately avoided a deep approach to learning because it was found unnecessary for success in hypothetical problems. Moreover, participants in the study were emphatic that they would not devote time and effort to non-assessed tasks, or tasks that bore little or no resemblance to the final examination, even when the tasks were designed to be interesting and interactive. Finally, they acknowledged that the surface approach that they admitted to taking did not produce sustainable learning, most of them confessing to have forgotten most of the subject matter.

News Reports

In the face of this damning indictment, an immediate solution did not present itself, in particular given the constraints of resources and class size. However, it was apparent that assessment could play a role in improving the situation. Moreover, to get students to think deeply, and to take their learning seriously, greater authenticity of learning and assessment would be necessary.

Having discussed with students in the 2006 survey focus group about their approach to the hypothetical narrative, the author determined to try to make the assessment content more authentic. This could help students make connections between their learning and real-world problems. But how could greater authenticity be achieved? By definition, any problem created by the teacher would suffer from the same deficiencies. The solution came from a somewhat obvious source. The author in his daily reading of the newspaper regularly encountered reports on events that were relevant to the subject under consideration – in this case tort law. Although not immediately apparent to the casual reader, on a close reading, many news reports concern issues of wider legal and social significance. A bit of scratching at the surface can often reveal a host of legal issues. This is so in respect of most areas of law. Indeed, many areas of study across the disciplines are regularly reported on in the media, implicitly if not explicitly. Using news reports in place of hypothetical narratives for assessment purposes could achieve an important objective. Asking students to respond to a question taken from the real world would have the potential to help sensitize students to legal and social issues and help them make connections between law as studied and as experienced in the community and practised in the profession.

The adoption of news reports as assessment questions was introduced gradually, beginning in 2008, first in the end-of-semester test (one question, carrying 20 % of the course weighting) and then in the final examination in 2008 (carrying 60 %), where two of seven questions (students select three) were news reports. The news reports were lightly edited by the author for clarity and compactness and to ensure coverage of legal issues. In 2010, the author took the decision to use only news reports in assessments and, importantly, news reports in their unedited, original

form. This was a difficult step because it greatly reduced teacher control over the form and content of the assessment. But this was a productive move because the teacher was replaced by a neutral third party as the one having control over the question content. The purpose of this move was explained to students that it was their role to sort out what was legally relevant and what was not – no different than the teacher in this respect.

The case in support of the use of news reports as assessment and learning materials is strong. By definition the material is realistic, in that the event, albeit reported by a journalist for news-reporting purposes, did happen. Moreover, invariably it is of some social consequence given that it was determined to be a newsworthy event. An important feature is that the material as reported by journalists in newspapers is generally complex, requiring multiple perspectives in analysis, including social policy. Social policy is something hard for a law student to take seriously in the conventional, teacher-invented, improbable hypothetical exam question scenario.

Another feature of authenticity is that the material in a news report is often factually incomplete. This is not unlike what a lawyer in practice should expect to encounter when interviewing a client, who inevitably fails to provide all the legally relevant facts while including irrelevancies. This greater degree of authenticity has the potential to foster a more serious approach to analysis and can help develop the habit of identifying issues in unflagged situations. The use of news reports fosters good reading habits and an awareness of the community and social policy issues. It requires students to make connections between their learning of legal doctrine and the kinds of problems that are occurring in the world around them. The use of news reports requires original thinking on the part of students. It helps them better understand the indeterminate nature of the law. The teacher did not create the narrative in the news report and has no claim to a single answer or even the correct answer. A student interviewee responded as follows: "It's more interesting to read a real than a fabricated case. When it is fabricated, you have the pressure in mind that the teacher actually has the answer before he sets the question, but if it is a real life case, it's up to you to make a logical argument" (Carless, 2015, p. 98).

Moreover, reading and analysing news reports will foster the habit of learning. Aware that assessment questions will be taken from the world of real events, students will get started early and learn to independently identify legally relevant events being reported in the media and to subject those news reports to independent legal analysis. These are skills and habits that cannot be learned in classroom doctrinal instruction.

The adoption of news reports proved to be labour-friendly and cost-efficient. A ready supply of authentic assessment material is available, and the form of assessment did not require a radical overhaul. It was probably an example of what Elton and Johnston (2002) meant by "doing things better" (rather than "doing better things"):

> when one thinks of the vast number of timed examinations that go on in universities, then a small improvement in timed examinations might have a far bigger effect . . . than the replacement of the timed examination by a superior method of assessing. (p. 7)

In order to ensure that an assessment activity produces the right kind of learning, it is important that it be aligned to the learning activities. Learning activities that are not so aligned run the danger of being ignored, or at least of not being taken seriously. Thus, if news reports are to comprise the raw material for assessments, they should also be used in learning activities and assignments, whether assessed or not. For this reason the author took the decision to replace all weekly tutorial problems which, as with other courses, were hypotheticals, with news reports. The learning gains described in the description of the examination assessment (above) apply equally here. Moreover, students tend to spread their efforts more evenly across the semester, knowing the connection between the learning activities and the summative assessment.

Of course, this assessment and learning format cannot claim to be fully authentic. For full authenticity, at minimum, advice to a real client would be required, perhaps by contacting the persons identified in the news report. For various reasons, including ethical considerations, that is not practical. However, as Boud (2009) contends, authentic contexts "need not necessarily involve students being placed in external work settings, but involve the greater use of features of authentic contexts to frame assessment tasks". The proposed model achieves a high degree of authenticity, is economical and adaptable and hence can be scaled up.

Students have responded well to this form of learning and assessment. Students are more engaged in tutorial discussions. Not surprisingly, they are more likely to take seriously and to participate in a discussion about important current happenings that have been reported on. A survey commissioned by the author that was conducted after the 2010 midterm test produced an approval rating of over 90 % (Glofcheski, 2010b), with students reporting a variety of insightful reasons for valuing this form of assessment and learning.

There are other possible permutations of this method that can be suited for a particular course and particular circumstances and learning outcomes. It may be possible, in some disciplines, to find other sources of authentic material that can be used for assessment purposes. For instance, in a course on patent and copyright law, it could be objects taken from department store shelves, the Internet or other sites of human activity. The key point is that whatever assessment mechanism is chosen, relevant and sustainable learning should be a key consideration, so that students are able to make connections between their academic learning and the world in which they live and in which they are about to work.

Activity-Based Authentic Assessment and Learning

An important and universally recognized learning outcome in law is the ability to identify legal issues in a narrative, independently from any guidance from a supervisor or teacher, and to provide legal analysis of those issues. To some degree, the teacher-invented hypothetical narrative used by most teachers in tutorials and assessments could be said to provide some opportunity for the achievement of

this outcome. However, as discussed, the hypothetical nature of such narratives has disadvantages, indeed risks, and misses the opportunity of authenticity and all the advantages it has to offer. Moreover, given that such hypothetical narratives are designed or selected by the teacher, the student is tipped off that the narrative will concern legal issues from the subject syllabus being examined. To that extent, teacher-selected work assignments and assessment questions can never really advance the learning outcome of independent identification of unflagged legal issues. To resolve that, it is necessary to require students to engage in the hunt for legal issues in their natural settings, i.e. in the events that happen in the community.

Project- or activity-based learning is related to authentic learning. Authentic tasks generally involve learning by doing, are ill-defined and are completed over a period of time (Herrington & Herrington 2006). Students undertaking such tasks have the opportunity to engage with authentic material and make connections between their classroom learning and the real world. Students learn best by doing rather than listening (Race, 2010). Listening produces very low levels of learning (Bligh, 2000). Engagement with the material is essential to achieving higher-order learning (Bloom, Engelhart, Furst, Hill, & Krathwohl 1956). For these reasons activity-based learning is preferred over the passive sort of learning that takes place in lectures (Gibbs, 1981). "Learning is not a spectator sport . . . [Students] must talk about what they are learning, write about it, relate it to past experiences, apply it to their daily lives. They must make what they learn part of themselves" (Chickering & Gamson, 1987, p. 140).

Activity-based learning holds out the possibility of a multiplicity of learning outcomes that capture the sorts of attributes that law graduates are expected to possess, including the possibility of independent learning and the skill of self-evaluation and self-monitoring of learning. The law curriculum, indeed any university curriculum, must be designed so that, in addition to reading and listening, students are required to do things. Deep learning can take place only if students engage with the material. Engagement requires doing something. And, of course, students must be assessed on these activities, in order to ensure that they are motivated and learn to do them well. This requires the introduction of more project- or activity-based assessments.

Reflective Media Diary

If law students are to acquire the skill of issue identification in unflagged contexts, they must be given the opportunity to look for issues in those undisclosed settings. Indeed, they must be required to do so repeatedly, over time. To be asked to do so only once cannot possibly produce the learning outcome in any meaningful or substantial way. The task must be designed so that students work independently, over time, to develop and acquire that skill. Learning portfolios can offer such an opportunity. They are self-directed and are completed over a period of time.

With this in mind, the author considered that a new learning tool was required in his course, bearing the characteristics described. It would have to carry assessment

weight, to ensure student engagement, while avoiding an increase of supervision or tutor marking load. The author designed and proposed to his students a variation of the learning portfolio, one that would align with the news-based authentic assessment and learning activities already described. The author consulted a focus group of graduated tort law students on the design of the assessment, and then proposed it to the class, at this introductory stage on an optional basis.

In the model proposed, students are required to independently identify tort law-related events as reported in the media. They keep a diary for the first 5 months of the course, of selected events reported in the news media that they identify as relevant to the course subject matter. For each item diarized, students record their legal reflections and make some attempt at legal analysis, bearing in mind that their formal study of the issues may be in the rudimentary stages or may not even have begun. The diary is web-based to ensure authenticity, timing and integrity of work. There is no need for teacher monitoring or intervention, an important consideration given the competing demands on teachers' time. The student monitors events as further reported in the media. At the conclusion of the diary period, the student selects ten diarized items for submission and assessment. Rubrics and exemplars are available from the beginning to ensure students' understanding of the learning outcomes and what is quality work. Within 6 weeks of the submission of the diary, the student selects two–three events for a detailed legal analysis, which is submitted as Part B, the capstone portion of the project. Students are instructed to select for Part B analysis news items that they perceive to be particularly legally problematic and that can showcase the learning of a range of legal issues and the development of advanced analytical skills.

As in the discussion of examination assessment, the authenticity and relevance of this task-based assessment have the potential to sensitize the student to social and legal issues relevant to the community and thus foster a more serious approach to legal analysis and the development of the habit of identifying issues in unflagged situations. Again, the material that students work with mimics reality. It is likely to be legally incomplete, in the same way that a client can be expected to present what he thinks is relevant, leaving the lawyer to excavate for more.

In the first year that the project was offered, it was optional, to be taken in place of a more conventional piece of legal analysis. Only 12 of 260 students took it up. In the following year, also optional, 62 students took it up. In the third year, almost all took it up, recognizing how well it coordinated with the primary means of assessment on the final examination. The project is now a compulsory component of the course, carrying 30 % of the course weighting.

Students have responded very well to this assessment (Glofcheski, 2011). Their specific comments suggest new insights in their understanding of the law and its interface with the community and a newly discovered ability to reflect on and monitor their own learning. As expressed by one student in separate interviews recorded by Carless (2015), "it makes you really think because most of the facts in the news are not the study of law; they are just common situations. There is no guided answer, so many issues are unclear, and we have to produce our own analysis" (p. 93). The learning is self-managed, knowledge is constructed by the

student and, given its habit-forming nature, it holds out the possibility of lifelong learning – the real possibility for learning beyond the assessment. It potentially stimulates deep approaches to learning in that students are encouraged to look for patterns, and it develops student metacognition as they sharpen their focus during the process of developing their diaries (Carless, 2015). It is now a compulsory component of both courses taught by the author (tort law and labour law). There is great scope for adaptations in other courses and disciplines and, hence, great scope for scaling up.

Conclusion

The assessment for learning initiatives discussed in this chapter has proven very effective in achieving high levels of student engagement and in advancing learning in the courses in which these learning activities have been introduced. In the examinations and graded assessments, they have been found by the teacher and tutors to help students acquire a better understanding of the social dimensions of law and bridge the gap between classroom learning and the world outside. Teachers of law, or of any discipline in higher education, might adapt them in ways appropriate to the discipline. They are mere examples of what can be achieved in the design of assessment for learning. An important point is that they were devised and introduced independently by the author, albeit after student consultation and agreement, demonstrating the author's belief that scaling up of assessment for learning starts with, and can be achieved by, the individual teacher, in partnership with learners.

References

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Maidenhead, UK: Open University Press.

Bligh, D. (2000). *What's the use of lectures?* San Francisco: Jossey-Bass.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives*. New York: David McKay Company.

Boud, D. (2006). Foreword. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. xvii–xxix). London: Routledge.

Boud, D. (2009). How can practise reshape assessment? In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 29–43). Dordrecht, The Netherlands: Springer.

Boud, D., & Falchikov, N. (2007). Aligning assessment with long term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413.

Bradney, A. (2003). *Conversations, choices and chances: The liberal law school in the twenty-first century*. Portland, OR: Hart Publishing.

Carless, D. (2015). *Excellence in university assessment*. London: Routledge.

Chickering, A. W., & Gamson, E. F. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education, 39*(7), 3–7.

Clegg, K. (2005). *Playing safe: Learning and teaching in undergraduate law*. UK Centre for Legal Education. Retrieved from http://www.ukcle.ac.uk/projects/past-projects/ncle/

Elton, L., & Johnston, B. (2002). *Assessment in universities: A critical review of research*. York, UK: Learning and Teaching Support Network Generic Centre.

Entwistle, N., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

Gibbs, G. (1981). *Twenty terrible reasons for lecturing*, *SCED Occasional paper No. 8, Birmingham*. London: Financial Conduct Authority.

Glofcheski, R. (2010a). *Tort law assessment survey* Retrieved from http://www.assessmentproject.com.hk/resources.html

Glofcheski, R. (2010b). *Students' tort test experience*. Retrieved from http://www.assessmentproject.com.hk/resources.html

Glofcheski, R. (2011). *RMD survey*. Retrieved from http://www.assessmentproject.com.hk/resources.html

Glofcheski, R. (2015). Rethinking teaching, learning and assessment in the twenty-first century law curriculum. In C. Gane & R. H. Huang (Eds.), *Legal education in the global context*. Farnham, UK: Ashgate.

Herrington, A., & Herrington, J. (Eds.). (2006). *Authentic learning environments in higher education*. Hershey, PA: Information Science Publishing.

Hess, G. (2002). Heads and hearts: The teaching and learning environment in law school. *Journal of Legal Education, 52*(1–2), 75–111.

Johnstone, R., & Vignaendra, S. (2003). *Learning outcomes and curriculum development in law (chapter 13)*. Report of Australian Universities Teaching Committee. Retrieved from http://www.cald.asn.au/docs/AUTC_2003_Johnstone-Vignaendra.pdf

LeBrun, M., & Johnstone, R. (1994). *The quiet revolution: Improving student learning in law*. Sydney, NSW: The Law Book Company.

Legal Education and Training Review (LETR). (2013). *Legal education and training review*. Retrieved from http://letr.org.uk/

Lombardi, M. (2007). Authentic learning for the 21st century. *ELI Paper 1*. Retrieved from http://net.educause.edu/ir/library/pdf/ELI3009.pdf

Motley, J. (1985–1986). A foolish consistency: The law school exam. 10. *Nova Law Journal, 723*.

Race, P. (2006). *The lecturer's toolkit* (3rd ed.). London: Routledge.

Race, P. (2010). *Making learning happen: A guide for post-compulsory education* (2nd ed.). London: Sage Publications.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). New York: Routledge.

Rankin, S. (2011). Tired of talking: A call for clear strategies for legal education reform: Moving beyond the discussion of good ideas to the real transformation of law schools. *Seattle Journal for Social Justice, 10*(1), 6–11.

Rochette, A. (2011). *Teaching and learning in Canadian legal education: An empirical exploration.* Doctoral dissertation, McGill University Canada. Retrieved from http://digitool.library.mcgill.ca/R/?func=dbin-jump-full&object_id=103488&local_base=GEN01-MCG02

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.

Savery, J., & Duffy, T. (1995). Problem based learning: An instructional model and its constructivist framework. In G. W. Brent (Ed.), *Constructivist learning environments*. Englewood Cliffs, NJ: Educational Technology Publications.

Schwartz, M. (2001). Teaching law by design: How learning theory and instructional design can inform and reform law teaching. *San Diego Law Review, 28*, 347.

Snyder, B. R. (1970). *The hidden curriculum*. New York: Alfred A. Knopf.

Stuckey, R., et al. (2007). *Best practices for legal education: A vision and a road map*. Clinical Legal Education Association. US: University Publications, University of South Carolina. Executive summary retrieved from http://www.cleaweb.org/Resources/Documents/best_practices-cover.pdf

Sturm, S., & Guinier, L. (2007). The law school matrix: Reforming legal education in a culture of competition and conformity. *Vanderbilt Law Review, 60*(2), 515.

Sullivan, W., Colby, A., Wegner, J., Bond, L., & Shulman, L. (2007). *Educating lawyers: Preparation for the profession of law*. San Francisco: Jossey-Bass/Wiley.

Tai, B., Lee, A., & Glofcheski, R. (2006). *How law students learn*. Retrieved from http://www.assessmentproject.com.hk/resources.html

Watson, A. (2001). Legal education reform: Modest suggestions. *Journal of Legal Education, 51*(1), 91–97.

Waye, V., & Faulkner, M. (2012). Embedding e-portfolios in a law program: Lessons from an Australian law school. *Journal of Legal Education, 61*(4), 560–584.

Chapter 6
Clinical Assessment Judgements and 'Connoisseurship': Surfacing Curriculum-Wide Standards Through Transdisciplinary Dialogue

Susan M. Bridges, Claire M. Wyatt-Smith, and Michael G. Botelho

Abstract Outcomes-based models in higher education recognize the key role of standards-based assessment in fulfilling the goal of curriculum alignment. However, writing clear yet nuanced descriptors or specifications is a continuing challenge. We examine an illustrative case of an assessment expertise project in dentistry at the University of Hong Kong to examine the role of standards descriptors within an overarching, programme-level assessment strategy. The project centred on examiner judgement during in situ clinical assessment tasks. Key to validity and reliability is making such judgements defensible, visible and accessible to students and examiners. Articulation of latent expertise and 'connoisseur' use of clinical performance criteria addressed the notion of accessibility. Connoisseurship extends beyond knowing the stated or explicitly defined criteria, to 'know-how' in using explicit, latent and meta-criteria. Central to the approach taken was how standards-based assessment can be enacted in clinical education by making latent, expert judgement practices explicit and adopting more flexible approaches than the traditional 'rubric'. Building a programme-level standards-based assessment culture draws upon dialogues across disciplines – both clinical and educational. Further research needs to illuminate professional judgement as process as well as the evaluative knowledge that is the source of assessment criteria and standards.

S.M. Bridges (✉)
Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong
e-mail: sbridges@hku.hk

C.M. Wyatt-Smith
Australian Catholic University, Brisbane, Australia

M.G. Botelho
University of Hong Kong, Hong Kong, China

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_6

Introduction

Outcomes-based models in higher education recognize the key role of standards-based assessment (SBA) in fulfilling the goal of curriculum alignment (Biggs & Tang, 2011). In clinical dental education, 'learning outcomes must be set to prepare all potential registrants for safe and independent practice' (General Dental Council, 2015, p. 5). The clinical assessment project presented in this chapter had its genesis in the observation that, in health professions education, 'faculty members who are well-trained clinicians or scientists often have little formal training in education' (O'Donnell, Oakley, Haney, O'Neill, & Taylor, 2011, p. 1163). The project, therefore, focused on supporting outcomes-based assessment in clinical higher education through a staff development initiative. A specific deliverable was to support the development of a coherent and robust set of assessment documents, specifically standards descriptors (also known as rubrics) for assessment tasks under the local nomenclature of clinical 'key skills'.

Conceptually and pragmatically, a performance-based clinical assessment task such as competency skills testing (see, e.g. Huth et al., 2015) differs from the non-clinical hypothetical tasks such as 'situational judgement tests' (SJTs) which can link key attributes to competencies (Patterson, Zibarras, & Ashworth, 2016). The clinical competency skills test or 'key skills' assessment, as applied to the undergraduate dental education programme discussed here, focuses on students' clinical performance, in real clinics with real patients. As a performance-based assessment in clinical education, authentic tasks are highly complex as:

> such performance is a composite of (clinical) cognitive, psychomotor and affective abilities (attitudes) of the individuals alongside their non-clinical skills like team working, situational awareness, etc. (Khan & Ramachandran, 2012, p. 921)

As such, the clinical key skills assessments comprise a series of complex, high-stakes evaluation events that are recognized by students and faculty assessors as authentic to clinical standards testing. However, there are challenges of standardizing the assessments not only individually but across the whole suite of assessment tasks.

From a faculty professional development perspective, impetus for the project was threefold. First was the desire to move beyond the discourse of objective and subjective assessments to focus instead on professional judgement and connoisseurship. Second was the recognition that standards and judgement lie at the heart of intelligent accountability (Sahlberg, 2008). Third was the need to develop a transdisciplinary approach to translational innovation in assessment. For definitional clarity, we understand transdisciplinary research to be where investigators from different disciplines work jointly to create new conceptual, theoretical, methodological and translational innovations that integrate and move beyond discipline-specific approaches to address a common problem (Aboelela et al., 2007).

The dual focus on translational innovation and transdisciplinary understandings was foundational to the project's design intent of scaling up assessment capability from an individual 'key skills' assessment to the whole suite of 13 key skills used

across the duration of the programme. This meant bringing together academics and clinicians across fields in exploratory dialogue with a focus on common, curriculum-wide task design and their use of insights and standards in an outcomes-based curriculum.

Underpinning Understandings

Connoisseurship or 'guild knowledge' (Sadler, 1987) is foundational to how experts engage in practice and in how they recognize quality. This is the case even though such knowledge is not always easily surfaced and articulated (see also Hounsell and Zou, this volume). In the clinical assessment project, guild knowledge contributed both to expert performance and to the assessment observation of students' clinical skills. The prism of clinical competence (Miller, 1990) supported this focus, providing a theoretical framing that located performance integrated into practice at the top of the clinical competence pyramid. Our focus, therefore, was on how students extended through Miller's levels of 'knows, knows how and shows how' to the top level of action, i.e. what the student does in key skills assessment tasks.

The approach chosen was to work with faculty staff, including academics and clinicians, to tap into their own expertise, prior evaluative experience in assessing clinical skills, discipline expertise and professional practice history. This had not been previously attempted. How staff appraised students' achievement of clinical skills through to grading had not been a focus for curriculum-wide professional development and research into teaching and learning. By working through dialogue, including individual discussions, small focus group meetings and collaborative working parties, the venture involved at least a measure of risk taking on their part.

At the commencement of the project, assessment statements of criteria had been previously available in course outlines; however, the academic participants had not focused on their own judgement practices. These included the essential features of the assessment process that helped them arrive at final judgements of quality of students' work. In effect, the stated assessment criteria had remained unproblematized and taken as received with knowledge inherited from previous clinicians or assessment writers. The prevailing assumption was that the criteria were written in terms that were self-evident and widely understood and that staff would be able to apply them in assessment. Also of note was that the criteria were stand-alone features judged on a pass/fail binary. Prior to the project, there was no information in the form of standards descriptors to support differentiation of performance quality. The project described here provided an important juncture point in the programme's clinical skills-based assessment to explore collaboratively the clinical assessor's latent, insider knowledge with regard to quality and their expectations of student performance in clinical key skills assessments.

In what follows, we establish the theoretical framing and main premises that informed the authors' approach to standards descriptors within an overarching, course- or programme-level assessment strategy. We share various approaches and models in developing not only the descriptor proforma but also how these can build an assessment culture. To demonstrate this, an illustrative case of an assessment expertise project in an integrated undergraduate dental curriculum (Botelho, Lo, Bridges, McGrath, & Yiu, 2013, Bridges, Yiu, & Botelho, 2016; Yiu et al., 2011, 2012) will explore these tensions and pose possible ways forward in building a SBA culture. The goal is to consider how, by making latent features of judgements explicit and by adopting more flexible approaches to the traditional 'rubric', SBA is enacted as part of a renewed programme-level assessment culture.

In the illustrative case, we explore how connoisseurship is different from simply knowing the stated or explicitly defined criteria. It involves 'know-how' in assessing: applying three categories of criteria – explicit (stated), latent (unstated) and meta-criteria (knowing how to combine and deploy or manage compensations and trade-offs) (Wyatt-Smith & Klenowski, 2013). Working from this position, 'rubric' formats can usefully denote standards and task-specific criteria for the novice assessor. They can, however, be potentially confining as to what assessors actually consider, or even should consider, in appraising student performance. This may reflect the notion that fair assessment necessarily involves a rigid adherence to predetermined criteria specifications. Fair assessment could also mean, however, being open to *the X-factor* – the surprise performance where valid features relevant to the task have not been previously specified. There is no doubt that thorny issues about fairness in assessment are at play in discussions about the utility of stated criteria, as well as the critical issue of whether it is reasonable to anticipate all relevant criteria can be wholly prespecified before assessment takes place.

In response, we argue for the centrality of the notion of professional judgement across all assessments, but especially when considering in-the-moment, performance-based clinical assessment tasks. Central to the approach taken to validity and reliability in the project is the aim of making judgements of the quality of students' completed tasks defensible, visible and accessible to students, examiners and regulatory bodies. Such reliability is necessary as there is a notion of failing to fail in the healthcare setting when underperforming students will manage to pass (Bush, Schreiber, & Oliver, 2013) based on a lack of calibration of examiners or an assessment tool that lacks sensitivity. There is also a strong trend of moving away from measuring competency by establishing levels of clinical activity with a focus on a standards- and quality-based approach with triangulation of other activities to determine professional competence (Dawson, Mason, Bissell, & Youngson, 2016). The use of defined and explicit grade descriptors assessing clinical performance, and an emphasis of professional judgement (see also van der Vleuten et al., 2012), would appear to be critical to this triangulation process.

Theoretical Framing

Standards-Based Assessment of Performance-Based Tasks

The starting premise for the project is that good assessment practice enhances both the quality and accountability of medical and dental education. Drawing on van der Vleuten's (1996) utility equation, the research team accepted that the construct validity of assessments had already been assured through curriculum certification processes. The core priority was the reliability of staff assessments of student achievements and the practicability of approaches to enhance reliability and comparability. Related priorities were the practicability and cost-effectiveness of quality assurance processes including new approaches to staff use of standards and moderation of results.

We also framed the project within SBA taken as key to an outcomes-based approach to teaching and learning. The core premises of SBA are grounded in Sadler's (1987) theorizing of formative assessment and his more recent writing on achievement and grading. In this light, the issues of validity, fidelity and standards, understood as 'fixed external anchors' for informing grading decisions, are viewed as central to SBA (Sadler, 2010). In this line of argument, validity and fidelity are both concerned with assessment evidence and are understood as connected. A distinguishing feature of the project is how the curriculum development team sought to connect these issues of *evidence* with issues of *judgement*, *with special interest in prior evaluative experience and connoisseurship as primary sources of evaluative criteria.*

Sadler (2010) proposed that validity indicates the scope and soundness of the assessment evidence to support interpretations as to student achievement, thus indicating the central concern with measurement. He also aligned fidelity to the legitimacy of the assessment evidence in that legitimacy comes from determining that the evidence generated from assessments is suitable or eligible for inclusion in the appraisal process, setting 'sharp boundaries for the object that is to be appraised' (Sadler, p. 742). In an outcomes- and competency-based model in clinical education, the logic is that the selected assessment item or task must assess the outcomes and competencies that it purports to assess. Recent work in rubric development in dental education has indicated their utility for clinical dentistry in both clarifying and defining performance expectations (O'Donnell et al., 2011).

In this clinical assessment project, curriculum developers undertook foundational work to enable task design with high validity and fidelity. For validity, a curriculum mapping project ensured alignment of the outcomes to the key skills assessment events was completed (Bridges, Yiu, & Botelho, 2016). For fidelity, the assessment evidence was reviewed for both 'scope' and 'soundness' in order to enable the strong inferences that could be drawn about underlying achievement. Additionally, each cross-curriculum key skill acts as a gatekeeping hurdle mechanism to the students' continued enactment of the clinical procedure in patient care.

An aim of the project was to explicate how judgements were arrived at and in talking through the processes that senior staff relied to 'see' or discern quality, tap connoisseurship and generate standards descriptors to be shared amongst staff and, in turn, students. As outlined below, this involved a process of talking with experienced staff in a range of settings (interviews, workshops and focus group meetings) to capture expert knowledge of the features or characteristics of fine performance that they looked for in assessing student achievement in the clinical setting. A related intention, and a secondary stage of the project, was to make these expectations explicit and available to staff, and in turn to students, in effect de-privatizing how judging of students' achievement of outcomes and competencies in clinical performance occurred.

Judgement, Connoisseurship and Evaluative Experience

A central premise of the project was that standards descriptors and exemplars of student and expert work could play a mediating role in the building of the 'guild knowledge' and the move from novice to expert. Further, judgement made against fixed external anchor points permits grading that is not influenced by factors unrelated to the performance being assessed, including 'how other students in the class perform, or each student's individual history of achievement' (Sadler, 1987). The project worked from the position that standards can play a key role in supporting efforts to deliver fair and reliable assessment, as well as enabling scrutiny back on to the nature of the assessment evidence to be collected in an iterative process of inquiry (see Fig. 6.1).

The primary function of standards is to present a formalized statement of expected quality at either an absolute level or levels. Irrespective of the form that standards take, typically they incorporate features of quality that we refer to as criteria. As noted earlier, there are three main types of criteria: explicit or stated criteria; latent or unstated criteria that may emerge in the course of an appraisal; and meta-criteria relating to how criteria can be combined or traded off, say higher performing features compensating for those of lower or inferior quality (Sadler, 2005; Wyatt-Smith & Klenowski, 2013). There is a developing body of writing arguing that while standards have a utility as fixed, external anchor points, this utility is enhanced when standards are accompanied by illustrative exemplars that show how the requirements of standards can be met (Hendry, Armstrong, & Bromberger, 2012). Standards tend to be abstract, while exemplars can provide concrete references for what the standards look like in practice. Exemplars in dentistry can take various forms, including static, often simulated work samples (Perry, Bridges, & Burrow, 2015) or recordings of professional practices or performances in action (see, e.g. Botelho, 2016). Dialogue was key to articulating expert notions of quality and in judging student performance of professional clinical skills, knowledge and dispositions.

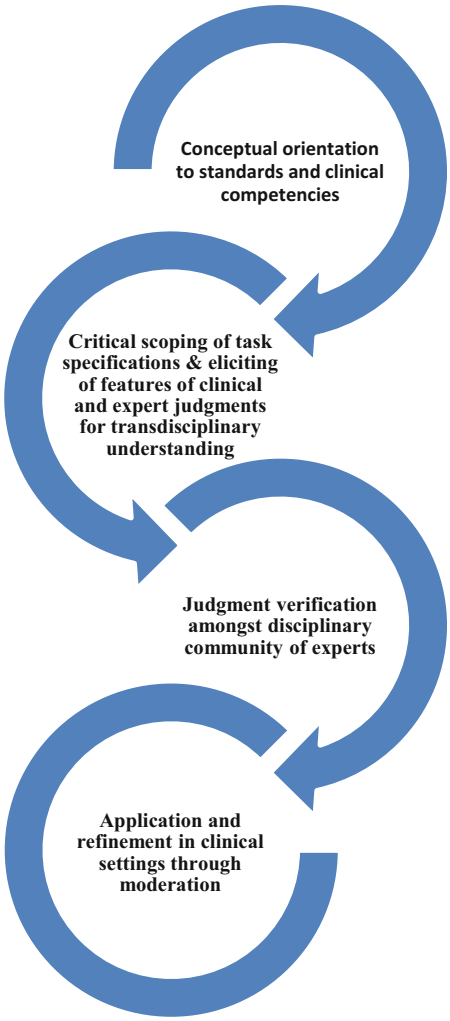

Fig. 6.1 Professional development cycles

In the faculty professional development innovation led by the authors, we designed a fit-for-purpose four-stage dialogic process of inquiry detailed below. The goal of the dialogues across disciplines and years of the curriculum was the principled generation of an agreed framework for determining standards and explicating descriptors, which would clarify distinctions between grade boundaries in a high-stakes clinical assessment task. In this work, we applied a notion of 'connoisseurship' to explore and make transparent all three categories of criteria, namely, explicit, latent and meta-criteria. From this vantage point, a related interest was the outcome students achieved and how professional judgements were informed by, but not limited to, the standards descriptors and the planned curriculum.

Project Overview

The process of developing refined criteria was already in place, but these were diverse in terms of their format and expected function. A key task for this project was to achieve consistency to support the reliability and validity of the tools. A core goal was to design a standardized intra-programme, cross year-level suite of assessment tasks with accompanying features of quality as recommended by international professional bodies (Plasschaert et al., 2007). The external assessment expert's role was to train and facilitate *principled dialogue* to create the standardized, robust approach necessary for quality assurance and accountability around grading and the implementation of SBA. The term 'principled dialogue' is taken to refer to trans-disciplinary dialogue around the task involving staff with expert knowledge in the discipline clinical skills, as well as others with expertise in educational assessment and standards. Explication of 'connoisseurship' became key to the process of four interconnected inquiry cycles outlined in Fig. 6.1 and discussed below.

The Process of Explicating Latent Judgement: Developing Clinical Assessor 'Connoisseurship'

In all cycles of the project, key participants were drawn from four distinct disciplinary fields and professional roles:

1. Clinical curriculum co-ordinator
2. Educational assessment expert
3. Higher education expert
4. Clinical disciplinary representatives

The first three constituted the 'expert panel', which led and facilitated dialogues at whole group and individual levels. The clinical disciplinary representatives were from the areas of family dentistry, oral rehabilitation (prosthodontics), oral and maxillofacial surgery, oral radiology, paediatric dentistry, oral diagnosis and treatment planning. These clinical faculty members reported that this assessment was a critical aspect of their work and that the sustained and iterative engagement in the design process enabled them to improve their thinking and practices in new ways, both individually and as an interdisciplinary team, to develop new, transdisciplinary understandings at both conceptual and pragmatic levels (Aboelela et al., 2007).

Cycle 1: Conceptual Orientation to Standards and Clinical Competencies

After initial briefings and observations of clinical learning in dentistry, members of the expert panel led a hands-on workshop attended by curriculum leaders and

the clinical disciplinary co-ordinators responsible for monitoring the key skills assessment items which were distributed across the 5 years of the integrated, problem-based curriculum model (Bridges, Yiu, & Botelho, 2016). The introductory workshop presented the conceptual framing for the project in addressing the nature of SBA and the planned process of formulating descriptive criteria for the evaluation of clinical competency skills.

One of the initial hands-on workshop tasks stimulated a critical review of existing key skills criteria against the graduate competency statements. This then drew the group into reflective discussions as to how their tasks articulate the expectations of the curriculum and required competencies. This was consistent with the project's focus on validity and fidelity. Participants sought to align the suite of 13 key skills clinical assessments that occur across the 5 years of the clinical curriculum to the graduate competency statements. This focus on assessment fidelity mentioned above focused sharply on evidence of learning. Specifically, the discussions focused on the breadth and depth of the demands of the key skills tasks and their cumulative increase in difficulty across the years of the curriculum. In these conversations, clinicians considered aspects of performance that extended beyond task performance to patient preparation and student self-evaluation.

They also reviewed again the staging and sequencing of the key skill across the years of the curriculum. This was seen as critical to the overarching assessment map which was aligned to the integrated curriculum design reflecting the notion of advancing complexity in terms of professional knowledge, clinical skills and professional dispositions, including ethics.

Cycle 2: Critical Scoping of Task Specifications and Eliciting of Features of Clinical and Expert Judgements for Transdisciplinary Understanding

Building on Cycle 1, the expert panel had generated a draft template for task-specific standards descriptors related to clinical performance (see sample in Tables 6.1 and 6.2). Participants then worked individually and collaboratively with laptops in the same room to revise the proposed template for purpose-built design, given that each task had its own specific clinical demands. The final product was an initial draft of a new key skills-specific standards descriptor (also referred to in the literature as a 'rubric') for which their discipline was responsible. The programme- and year-level learning outcomes and graduating competencies were on hand to guide the development of task-specific criteria to support 'constructive alignment' (Biggs & Tang, 2011).

Disciplines took turns in presenting the initial standards descriptors (rubrics) on an interactive whiteboard (IWB) to the group with some feedback and sharing. This was an opportunity to put the demands of the task and the descriptions of quality under the microscope with regard to levels of detail and associated criteria

for possible assessment. Dialogue focused on the type of evidence that the task was expected to generate, including the scope of the evidence and the range of quality. The type of language used was discussed to help allow differentiation between standard levels as well as improving clarity on what was actually important and being assessed.

Table 6.1 Sample 'hurdle' in clinical key skills: a pre-assessment requirement

	Graduate competencies	**Grade**	
Case preparation	**Domains and competencies**	**Fail/do not proceed**	**Pass/proceed**
	Domain 6: treatment planning	**Case assessment**	**Case assessment**
	6.1 Comprehensive treatment plan established to promote oral health and to prevent and manage oral diseases and oral conditions	Oral health is not satisfactory for receiving prosthodontic treatment. Preventive, periodontal and/or restorative treatment is not complete or is inappropriate. Denture design has not been approved, and/or casts have not been surveyed and correctly mounted. Trial tooth preparations have not been performed on a duplicate cast. Aesthetic wax-up (if necessary) has not been done	Oral health is satisfactory for receiving prosthodontic treatment. Preventive, periodontal and/or restorative treatment is complete and appropriate. The denture design is approved, and casts have been correctly mounted and surveyed. Trial tooth preparations have been performed on a duplicate cast. Aesthetic wax-up/try-in (if indicated) has been performed
	6.2 Obtaining consent for all treatment planned		
Patient care		**Insufficient**	**Sufficient**
	Domain 1: professionalism	Inadequate infection control measures of the clinical working environment, patient, dental instruments or associated materials for safe clinical care	Appropriate infection control measures of the clinical working environment, patient, dental instruments or associated materials (impression) for safe clinical care
	1.7 Working effectively with the team to maintain the highest possible standards of infection control and in the control of all hazards associated with oral healthcare delivery, including the appropriate clinical use of dental biomaterials		

(continued)

Table 6.1 (continued)

	Domain 2: ethics and jurisprudence	Inadequate information and/or informed consent provided	Appropriate information and/or informed consent provided
	Major competence: ethical practice		
	2.3 Sufficient information provided in order to gain informed consent for decisions regarding oral healthcare delivery		
	Domain 3: communication	Inadequate care and empathy displayed for patient care with notable/avoidable damage to abutment(s) or other tissues	Appropriate care and empathy displayed for patient care with no/minimal damage to abutment(s) or other tissues
	Major competence: professional communication		
	3.1 Empathy shown with the patient		

This shared dialogue across the clinical disciplines provided opportunity to pool shared expectations as to the salient features of both task and clinical execution. Such a process of de-privatizing amongst the various disciplinary experts themselves opened up conversations about quality, thereby expanding the existing assessment culture which had, to some extent, been previously 'siloed' into disciplines. At the end of the workshop, clinical disciplinary representatives were asked to complete their key skills standards descriptors and to resubmit these for feedback, consolidation and alignment after discussion with colleagues in their area of the aims, objectives and details of the assessment rubrics. In addition, they were asked to map the year-level and programme-level outcomes to the key skills.

Further, in this cycle, the dialogue focused sharply on defining the breadth and detail of each key skills performance task to be assessed. A key juncture point of debate related to those tasks involving patient care as part of a holistic treatment plan. For these more complex tasks, the expert panel proposed to employ a 'hurdle' as a compliance requirement (see Tables 6.1 and 6.2). This was conceived in order to address key issues raised by clinicians across multiple disciplines concerning 'patient preparation' and treatment readiness. This proposed 'hurdle' requirement was defined in terms of four 'graduate competencies' and their task-aligned, specified requirements which were seen as common requirements before proceeding to perform the clinical procedure on patients, namely:

1. Professionalism

- 'Working effectively with the team to maintain the highest possible standards of infection control'

Table 6.2 Sample clinical key skills standards and criteria (oral rehabilitation)

Performance	Domain 7: delivery of oral healthcare			
	7.14 Replacing missing teeth and supporting tissue where appropriate, by means of removable prostheses			
	Criteria	**Concerns**	**Safe**	**Proficient**
	Supporting knowledge	Insufficient knowledge of the key elements of principles, practice and materials relevant to the key skills performance	Sufficient knowledge of the key elements of principles, practice and materials relevant to the key skills performance	Accurate, detailed and elaborated knowledge of principles, practice and materials relevant to the key skills performance
	Case presentation	Inadequate or inaccurate justification for the proposed partial denture design and/or tooth preparation plan	Accurate justification for the critical features of the proposed partial denture design and tooth preparation plan	Accurate and detailed justification for the proposed partial denture design and tooth preparation plan, as well as recognize and propose alternate satisfactory designs
	– Justification and rationale for denture design			
	– Appraisal of trial tooth preparations			
	Tooth preparation procedure	Rest seat, guide planes	Rest seat, guide planes	Rest seat, guide planes
		Tooth preparations are over- or underprepared and/or with incorrect features	Tooth preparations are acceptably prepared	Tooth preparations are neat, smooth and well-finished as planned
	Outcome evaluation	Inaccurate or inadequate self-evaluation of key features of the tooth preparations	Accurate self-evaluation of key features of the tooth preparations	Detailed and accurate self-evaluation of the features of the tooth preparations
	– Appraisal of tooth preparations			

2. Ethics and jurisprudence
   - *Ethical practice* in terms of 'sufficient information provided in order to gain informed consent for decisions regarding oral healthcare delivery'
3. Communication
   - 'Empathy shown with the patient'
4. Treatment planning
   - 'Comprehensive treatment plan established to promote oral health and to prevent and manage oral diseases and oral conditions'
   - 'Obtaining consent for all treatment planned' (see Table 6.1)

Cycle 3: Judgement Verification Amongst Disciplinary Community of Experts

In this cycle, the clinical disciplinary representative presented the draft standards descriptor for discussion with the other three expert panel members for critical review.

Throughout the entire project, learnings for all participants concerned the language choices we made to capture and communicate the evidence staff wished to collect and how to talk about that evidence in terms of quality. Collectively, we were exploring how to see quality and how to talk about quality. The complexity of the criteria and accompanying standards descriptors proved challenging for clinical assessors. Early drafts of the standards descriptors were often too detailed. They tended, at times, to be led by a more behaviourist, checklist approach in a desire to display objectivity and transparency. The frequently revisited notion was one of salience, i.e. *from the pool of possible features to attend to*, *what are the most salient for the purpose of assessing quality?* The learning for the group was that the attempt for an exhaustive listing of all possible criteria was potentially self-limiting, i.e. the attempt to fully prespecify or anticipate all relevant criteria and rigidly adhere to these precluded the opportunity to *see* in students' work previously unanticipated features of performance.

The successive drafts were moderated within the transdisciplinary panel to focus on the most salient features of the performance (criteria) and their related quality (standards descriptors). The distinctiveness of the standards, that is, the features that distinguished one level of performance from another was at times limited by language, especially the tendency at first to employ 'good, better, best' continuum of the stem descriptor. The probing nature of the discussion encouraged critical reflection to discern task requirements and to surface expert connoisseur notions of quality. The tension arose between overspecification of detail in the pursuit of comprehensiveness and 'objectivity' and the need to distil central features of particular salience and then unpack those in terms of the standards.

Cycle 4: Application and Refinement in Clinical Settings Through Moderation

Cycle 4 is the next phase of development and will be reported in subsequent publications. The goal of this cycle, however, is for clinical areas to modify the created standards-based rubrics for assessing students' competency and to collate clinical performance-based skills outcomes that can be captured to use as exemplars for the different standards. These exemplars can be used before the assessment procedure so that staff can collectively moderate and calibrate their judgements

and scope of assessment before conducting the summative assessment procedures. Additionally, exemplars with rubrics provide a powerful tool for student learning, and there is ongoing potential for this to be explored.

Discussion

The clinical key skills were identified as a critical gatekeeping task in the faculty's clinical curriculum. They may be considered a precursor to what medical educators (ten Cate et al., 2015) have recently proposed as entrustable professional activities (EPAs). These assessment tasks are 'units of professional practice (e.g. anaesthetic care of an uncomplicated patient)... executable within a given time, observable, measurable, confined to qualified personnel and suitable for focused entrustment decisions' (ten Cate et al., 2015, p. 985). This indicates a shift from describing competency or experience in terms of numbers of skills performed to focusing on assessing graduate readiness through performance integrated into practice.

The transdisciplinary approach taken in the dental education project described above supported the assessment of a critical constellation of clinical skills and dispositions conducted in an authentic setting and in delivery of care to actual patients. The dialogic inquiry allowed unstated (latent) criteria to be identified and become explicit. These latent criteria emerged in the discussions by reference to deep structures and prior evaluative experience (Wyatt-Smith & Klenowski, 2013). The iterative process of dialogue between clinicians across disciplines and facilitated by experts in higher education and SBA enabled a context for sharing transdisciplinary understandings and guided clinical faculty to construct appropriate standards descriptors for the clinical key skills assessments.

Outcome 1: Notions of Quality as Central to Standards, Judgements and Fitness for Purpose

Capturing the continuum of quality and discerning the grade boundaries became key points of discussion. This was, however, highly productive conceptually given recent calls in medical education to consider 'competence as a point on the spectrum of performance' (Khan & Ramachandran, 2012, p. 920). As clinicians, the disciplinary experts held a responsibility for quality assurance in that passing a clinical key skill was, in effect, the permission to treat a patient independently. This is 'high-stakes assessment' in many aspects. It is high stakes for the patient (now and for future patients). It is also high stakes for the maintenance of professional standards and the professional standing of the provider.

What became evident in the entire initiative was the issue of *fitness for purpose*. The five-standard continuum (*concerns*, *reservations*, *safe*, *competent*, *proficient*) employed in the formative assessments during clinical sessions reflected a focus on feedback for improvement purposes. The various key skills staged at critical junctures across the curriculum constituted a sequence of summative, 'hurdle' assessments which were necessary for advancement from 1 year to the next. They also were critical milestones in attaining and demonstrating the achievement of professional clinical competencies. The employment of a three-standard continuum (*concerns*, *safe and proficient*) was seen, at this point in clinical preparation as more salient to the purpose of professional gatekeeping. A continuum of quality that could be captured across three standards was seen as more clearly denoting and explicating quality and expectations of performance.

Outcome 2: Criteria and Standards as an Informing Framework for Judgements

The criteria and standards descriptor can be understood as an 'anticipatory set' for informing decision-making. While they may appear fixed at a point in time, over time they are best understood as remaining open to review and refinement. This openness addresses key developments in terms of new expectations based on performance outcomes, including exceptional performances that may go beyond stated expectations.

Now that the criteria and standards descriptors have been formulated, the next step is to address how they are used by individual staff and teams and also in moderation practice for accountability. Quality assurance claims rest not only on the provision of the descriptors, but also on how they are used. Research shows that for optimum effect, standards descriptors or standards are accompanied by exemplars that illustrate the features of quality captured in the standards. Building on the curriculum alignment project presented in this chapter is the opportunity for developing digital exemplars along with student work samples and standards descriptors, with staff sharing recorded commentaries about professional judgement. Preliminary discussions have been undertaken to establish the dual purposes of this work, namely, to inform staff training for calibrating standards-referenced judgement in clinics and also to inform student learning. In effect, the exemplars together with the standards descriptors would serve as valuable concrete examples of high-quality work and develop students' knowledge of assessment expectations of quality. In turn, these can be used to support students' self-assessment skills as intensive training has been shown to align more accurately to tutor assessments when judging quality (Huth et al., 2015).

Conclusions

From the commencement of the project, the research team was interested on how clinicians assessed student performance in clinical key skills. The search was not for the 'right' assessment technique but rather the actual resources that the clinicians relied on in arriving at judgement. The focus, therefore, extended beyond the curricular resources to judgement approaches and, specifically, the mix of explicit or stated criteria they relied on, as well as the latent and meta-criteria that came into judgement processes. Some staff reflected on their own clinical skills development and how they had learnt through 'observing', 'doing' and 'reflecting' on how well they had performed with the setting of quality at the centre of all.

In the case of assessing clinical key skills, the sources of evaluative criteria appear to be:

(a) Each clinician's expert disciplinary knowledge
(b) Their prior clinical professional practice
(c) Their cumulative evaluative experiences
(d) The institutional context framing their current practice and learning opportunities

The exploration of the source of evaluative criteria holds great potential as a new area of assessment research in higher education in general and clinical education in particular. From the outset, we understood that the clinicians' prior clinical experience typically remained private and, therefore, unarticulated.

In the connoisseurship model, however, it is the potent mix of expert knowledge and evaluative experience that can inform the critical moments whereby the clinician articulates what was previously the inside-the-head expert judgement. When this judgement is articulated, it is opened to scrutiny by clinical colleagues, educationalists and novice clinicians alike in order to build shared understandings of quality. It is the raising of these 'deep structures' to the surface that enables shared understanding. The ongoing explication of the elements of these latent criteria was the greatest challenge to the project. Charting the 'deep structures' of expert judgement is the beginning of the work that needs to be undertaken in clinical skills assessment and much more needs to be done in the field.

The conceptual and applied outcomes of this project have direct relevance to clinical contexts. This is particularly the case in discipline areas where the application of standards for learning improvement and accountability involve educators and students reviewing current pedagogical practices and adopting new assessment and accountability practices. Core to the discussions across the cycles was the process of 'finding the right words' to explicate insights as to 'how to see quality'. In part, this involved connecting the language used in talking about clinical skills with the language of assessment in order to convey what constituted quality when making judgements on observed clinical performance. More work is required in the field of clinical judgement. The professional development model which is an outcome of this project is one critical step forward. A second outcome concerns staff learning

how to engage in talk of assessment evidence and expectations of quality, connecting connoisseurship, clinical expertise, standards and judgement with potential benefit to students' own knowledge and skills in self-monitoring.

References

Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., et al. (2007). Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research, 42*, 329–346.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. Maidenhead, UK: McGraw-Hill and Open University Press.

Botelho, M. (2016). The communal consultation: Video-recorded student consultations. *Medical Education, 50*, 573–574.

Botelho, M. G., Lo, E. C. M., Bridges, S. M., McGrath, C. P. J., & Yiu, C. K. Y. (2013). Journal based learning, a new learning experience building on PBL at HKU. *European Journal of Dental Education, 17*, e120–e125.

Bridges, S. M., Yiu, C. K. Y., & Botelho, M. G. (2016). Design considerations for an integrated, problem-based curriculum. *Medical Science Educator*. doi:10.1007/s40670-016-0255-6.

Bush, H. M., Schreiber, R. S., & Oliver, S. J. (2013). Failing to fail: Clinicians' experience of assessing underperforming dental students. *European Journal of Dental Education, 17*, 198–207.

Dawson, L. J., Mason, B. G., Bissell, V., & Youngson, C. (2016). Calling for a re-evaluation of the data required to credibly demonstrate a dental student is safe and ready to practice. *European Journal of Dental Education*. doi:10.1111/eje.12191.

General Dental Council. (2015). *Preparing for practice*. London: General Dental Council.

Hendry, G., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment and Evaluation in Higher Education, 37*(2), 149–161.

Huth, K. C., Baumann, M., Kollmuss, M., Hickel, R., Fischer, M. R., & Paschos, E. (2015). Assessment of practical tasks in the Phantom course of conservative dentistry by pre-defined criteria: A comparison between self-assessment by students and assessment by instructors. European Journal of Dental Education doi:10.1111/eje.12176.

Khan, K., & Ramachandran, S. (2012). Conceptual framework for performance assessment: Competency, competence and performance in the context of assessments in healthcare – Deciphering the terminology. *Medical Teacher, 34*(11), 920–928.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), 63–67.

O'Donnell, J. A., Oakley, M., Haney, S., O'Neill, P. N., & Taylor, D. (2011). Rubrics 101: A primer for rubric development in dental education. *Journal of Dental Education, 75*(9), 1163–1175.

Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teacher, 38*(1), 3–17.

Plasschaert, A. J. M., Manogue, M., Lindh, C., McLoughlin, J., Murtomaa, H., Nattestad, A., & Sanz, M. (2007). Curriculum content, structure and ECTS for European dental schools. Part II: Methods of learning and teaching, assessment procedures and performance criteria. *European Journal of Dental Education, 11*(3), 125–136.

Perry, S., Bridges, S. M., & Burrow, M. F. (2015). A review of the use of simulation in dental education. *Simulation in Healthcare, 10*(1), 31–37.

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*, 191–209.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education, 30*(2), 175–194.

Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education, 35*(6), 727–743.

Sahlberg, P. (2008). Rethinking accountability in a knowledge society. *Journal of Educational Change, 11*(1), 45–61.

ten Cate, O., Chen, H. C., Hoff, R. G., Peters, H., Bok, H., & van der Schaaf, M. (2015). Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Medical Teacher, 37*(11), 983–100.

van der Vleuten, C. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Science Education, 1*, 41–67.

van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher, 34*(3), 205–214.

Wyatt-Smith, C., & Klenowski, V. (2013). Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy and Practice, 20*(1), 35–52.

Yiu, C., McGrath, C., Bridges, S., Corbet, E., Botelho, M., Dyson, J., & Chan, L. K. (2011). Graduates' perceived preparedness for dental practice from PBL and traditional curriculum. *Journal of Dental Education, 75*(9), 1270–1279.

Yiu, C., McGrath, C., Bridges, S., Corbet, E., Botelho, M., Dyson, J., & Chan, L. K. (2012). Self-perceived preparedness for dental practice amongst graduates of The University of Hong Kong's integrated PBL dental curriculum. *European Journal of Dental Education, 16*(1), e96–e105.

Chapter 7
The Use and Design of Rubrics to Support Assessment for Learning

Anders Jönsson and Ernesto Panadero

Abstract Rubrics are assessment instruments designed to assist in identifying and evaluating qualitative differences in student performance. Research into scoring rubrics has shown that they can serve two purposes: (1) aid assessors in achieving higher levels of consistency when scoring performance tasks and (2) promote learning and/or improve instruction by making assessment expectations explicit and aiding the feedback process. In this chapter we summarise research on the formative use of rubrics, in order to identify how the use and design of rubrics may be optimised for the purpose of supporting student learning in an environment that often stresses independence and management of own learning. The presentation is organised around two different pathways through which rubrics may support student learning. These are through (a) facilitating the understanding and use of feedback and through (b) supporting students' self-regulated learning. We also analyse what is known about the implementation of rubrics in higher education, with a particular focus on more sustained and widespread implementations. The implications of these findings for both practice and future research on rubrics are discussed.

Introduction

Research on the use of scoring rubrics has shown that these assessment instruments can aid assessors in achieving acceptable levels of consistency when scoring performance tasks. Research has also documented positive educational consequences of rubric usage, such as supporting students' development towards independent learners and improved student performance. These effects seem to come from the fact that rubrics make expectations and criteria explicit, which in turn facilitate other processes, such as interpreting and using feedback. As a consequence, rubrics have

A. Jönsson (✉)
Kristianstad University, Kristianstad, Sweden
e-mail: anders.jonsson@hkr.se

E. Panadero
Autonomous University of Madrid, Madrid, Spain

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_7

been used as a tool for communicating expectations, as well as a support for other assessment for learning (AfL) practices (Jönsson & Svingby, 2007; Panadero & Jönsson, 2013).

Although most studies explicitly linking the use of rubrics to AfL are performed in school settings, there are a number of studies from higher education contexts as well. A majority of these studies, however, are small scale and use short-term interventions (Brookhart & Chen, 2014; Panadero & Jönsson, 2013), which makes it difficult to get an overview of how to successfully use and design rubrics for formative purposes. The first aim of this chapter is therefore to summarise research on rubrics, in order to identify how the use and design of rubrics may be optimised for the purpose of supporting student learning in an environment that often stresses independence and management of students' own learning. The second aim is to analyse what is known about the implementation of rubrics in higher education, with a particular focus on more sustained and widespread implementation.

What Is a Rubric?

Assessing complex tasks, such as dental students communicating with patients or science students formulating arguments in socio-scientific issues, is difficult, and a common recommendation for increasing the quality of teachers' assessment has therefore been to use detailed scoring protocols, preferably together with sampled responses which exemplify the points on the scoring scale (e.g. Linn, Baker, & Dunbar, 1991). An instrument that effectively matches these recommendations is the rubric. Rubrics are designed for assisting assessors in judging the quality of student performance, and in order to accomplish this, all rubrics have three fundamental features in common. First, in order to assist in identifying the qualities to be assessed, the rubric includes information about which aspects or criteria to look for in student performance. Second, in order to assist in judging the quality of student performance, the rubric includes descriptions of student performance at different levels of quality. By combining the aspects to be assessed with the descriptions of quality into a two-dimensional matrix, along with a scoring strategy (i.e. the third feature), a rubric comes into existence.

How the Use of Rubrics May Support Student Learning

As established by reviews on research about rubrics, the transparency provided by rubrics may facilitate other AfL related processes, such as interpreting and using feedback or assessing the performance of peers (Jönsson & Svingby, 2007; Panadero & Jönsson, 2013). But there are also other less direct ways in which the use of rubrics has been shown to support student learning, such as reducing anxiety or supporting self-regulated learning (SRL) strategies. Below, these two pathways – i.e.

the more direct pathway through facilitating the understanding and use of feedback, as well as the indirect pathway via SRL strategies – are outlined and illustrated with relevant studies.

Pathway 1: Facilitating the Understanding and Use of Feedback

According to literature reviews and meta-studies, such as the one by Hattie and Timperley (2007), feedback can have a strong influence on student learning. However, a number of students do not use the feedback they receive and therefore do not realise the potential of feedback for learning (Jönsson, 2013). There may be several reasons for not utilising feedback, but two of the main obstacles seem to be that students either do not understand the feedback or do not know how to use it. By using rubrics to support the feedback process, both of these difficulties may be mitigated.

First, by making assessment criteria explicit, the understanding of feedback can be facilitated. As an example, Case (2007) wanted to reconfigure the feedback process through students' active engagement with explicit assessment criteria. To this end, he devised an electronic template for feedback, which incorporated the assessment criteria (in the shape of a rubric) as the main feedback element. Along with the feedback form, there was a bank of electronically stored statements, which served as the basis of more specific feedback. This bank included comments relating to common student weaknesses, mistakes, recommended improvements and strengths. Findings show that the changes had a positive effect on student performance and that these improvements were largely attributable to the fact that the feedback process enhanced students' awareness and understanding of learning outcomes and assessment criteria. It is also noteworthy that there was a reduced marking time due to the use of statement banks and fewer student queries as a result of the explicit engagement with criteria. Similar results have been obtained in studies combining rubrics with exemplars (Jönsson, 2010).

Second, the fact that rubrics – by definition – include descriptions of student performance at different levels of quality means that there is an inbuilt 'feed-forward' feature in all rubrics. This, in turn, may aid teachers when struggling to provide constructive feedback to students. For instance, Schamber and Mahoney (2006) report that most faculty members in their study found a rubric for assessing critical thinking to be very useful in order to provide students with feedback on drafts. Students also found the same rubric useful for clarifying expectations. Furthermore, field supervisors reported that a rubric for clinical supervision in counsellor education facilitated the provision of feedback on a continuous basis as well as giving concrete recommendations for improvements (Hanna & Smith, 1998).

Feedback need not only be provided by teachers but also by peers or the student herself. There are, however, few scientific studies reporting on effects of self- and peer assessment using rubrics (Panadero & Jönsson, 2013). Also, in a number of

such studies, the main focus has not been on learning, but on accuracy (Jönsson & Svingby, 2007). One interesting example of research on self- and peer assessment using rubrics is a study by Reitmeier, Svendsen and Vrchota (2004). In their study, students worked with self- and peer assessment of oral communication skills in a food preparation course, and both the scores of the presentations and the grades were higher as compared to the scores and grades attained in the previous semester. These changes were attributed to the use of rubrics. In another study, dental students used a rubric for communication skills to provide feedback to each other during role play with a simulated patient. The evaluation of this exercise was very positive and most students thought they learned from the task and that the criteria were useful for them (Lucander, Knutsson, Salé, & Jönsson, 2012). Similar results have been reported for students using a rubric to monitor their task performance in statistics and epidemiology, as well as for real estate broker students (Jönsson, 2014).

Pathway 2: Facilitating Students' Self-Regulated Learning

Rubrics have also been demonstrated to support student learning through different aspects of SRL. Some of these studies overlap somewhat with previously mentioned studies on self-assessment, where students were seen to use rubrics to monitor and evaluate their task performance (e.g. Jönsson, 2014). This is not surprising, given that self-assessment is considered an integral part of most models of SRL, since students need to monitor and evaluate their progress in order to regulate their work (Panadero & Alonso-Tapia, 2013). For instance, according to Zimmerman's (2013) cyclical model of self-regulation, which is one of the most cited in the literature, SRL consists of three phases (forethought, performance and self-reflection), and self-assessment is an important element in at least two of these phases (i.e. performance and self-reflection).

The forethought phase, on the other hand, is more often associated with goal setting, planning and self-efficacy, because students do not necessarily self-assess in this part of the cycle. However, the forethought phase does provide important conditions for the self-assessment to come, for example, by establishing the assessment criteria (Panadero & Alonso-Tapia, 2013). There seems to be no research where students have used rubrics as a tool for goal setting, but there are a number of studies investigating students' planning. For instance, in a study by Andrade and Du (2005), students reported using the rubric to plan their approach to an assignment, much like a recipe or a map. Also, in a replication study by Reynolds-Keefer (2010), students' responses indicate that rubrics aided them in both planning and when performing the assignment. Most students claimed to read the syllabus and then start working on the assignment, using the rubric as a reference point. Several students also stated that they worked through the assignment by reading the rubric and working on one portion at a time, merging all the separate parts before submitting.

In Panadero and Jönsson (2013), rubrics were found to mediate improved student performance by, for instance, reducing anxiety. These effects too are contingent on the transparency provided by the rubrics. For instance, when asked about anxiety, pre-service teachers spoke about increased confidence and making it easier to hand in assignments when having access to a rubric (Andrade & Du, 2005).

Closely related to the issue of anxiety is students' perceptions of their own capability, and the impact on student self-efficacy from using rubrics has been investigated in a number of studies. The findings, however, are inconclusive. For instance, Panadero (2011) investigated the relationship between self-efficacy and rubric use in three different studies. In this research, self-efficacy was impacted by the use of rubrics in only one of the cases (with secondary students) and only in interaction with the type of feedback received. It should be noted that these studies did not control for whether the students had received feedback by the teacher, which is a factor suggested to significantly affect students' perceptions of their performance. Still, it is difficult to find studies where higher education students' self-efficacy is affected by the use of rubrics. An interesting exception is a study by Yopp and Rehberger (2009), where low-performing mathematics students were given a rubric and told that they had to master the material (i.e. obtain a score of 100 % based on the rubric). While instruments for measuring both general and subject-specific self-efficacy were administrated to the students, only the mathematics self-efficacy changed during the intervention. According to the authors, a possible explanation for this could be that academic attitudes and beliefs are difficult to change in a single specific course. This explanation is corroborated by Balan (2012), but since the self-efficacy questionnaires in the research by Panadero (2011) were task specific, the difference between general and subject- or task-specific measures does not seem to explain the inconclusive results in relation to self-efficacy and rubrics.

Panadero and his colleagues have also done a number of studies relating to students' learning orientations and SRL. In one of their investigations, they found that the level of SRL strategies, as measured through think-aloud protocols, was higher in a group of secondary education students using rubrics as compared to students in a control group (Panadero, Alonso-Tapia, & Huertas, 2012). In another study, it was found that scores on a performance- and avoidance-oriented SRL scale decreased for pre-service teachers using rubrics (Panadero, Alonso-Tapia, & Reche, 2013). In yet another study, Panadero and Romero (2014) found that a group of pre-service teachers using rubrics scored higher on a learning-oriented SRL questionnaire, as compared to students who were asked to self-assess their work without any instrument to facilitate the self-assessment. Performance- and avoidance-oriented SRL scores also decreased significantly in the rubric group.

The findings above are indications of positive effects on students' SRL, but there are other findings as well. The students using rubrics in the study by Panadero and Romero reported higher levels of stress while performing the task as compared to the control group. Also, the learning-oriented SRL scores decreased for first year psychology students who used rubrics (Panadero, Alonso-Tapia, & Huertas, 2014).

This means that while the use of rubrics may decrease performance- and avoidance-oriented SRL strategies, which are often detrimental for learning, they do not necessarily increase learning-oriented SRL.

In summary, the use of scoring rubrics has been shown to facilitate student learning through two (more or less) distinct pathways. The first pathway involves aiding students in understanding and using feedback from teachers, peers or themselves. The fact that rubrics contain different levels of quality also means that it is easier to provide – and possibly to interpret – constructive feedback. The second pathway involves an impact on SRL and related factors such as self-efficacy, student anxiety and learning-orientation, factors that may in turn affect student performance. Again, an important requirement seems to be the transparency provided, which makes it possible for students to estimate their own capability, as well as to plan, monitor and evaluate their work according to the explicit criteria. This means that students can exert more control of their own learning, which potentially reduces anxiety and negative SRL strategies. At the same time, however, this control and responsibility can create feelings of stress and also make students more performance oriented. An extreme version of this performance orientation is reported by Torrance (2007), where an unfortunate combination of explicit criteria and coaching in post-secondary education resulted in a situation where 'criteria compliance' tended to replace productive learning.

How the Design of Rubrics Can Support Student Learning

As noted by Dawson (2015), the term 'rubric' is sometimes used with different meanings. Rubrics may therefore be 'embraced and resisted based on often imprecise and inconsistent understandings of the term' (p. 1). Consequently, Dawson proposes a framework for the categorisation of rubrics, encompassing 14 design elements that can be used to make distinctions between different rubrics. This framework, with Dawson's nomenclature, is used in this section.

It should be noted that Dawson (2015) does not express support or preference for any decisions regarding any particular design. In his case this has to do with his research approach, but generally it also has to do with the fact that many research articles do not include complete information about the rubrics used and that most of the design elements of rubrics have not been systematically investigated (Panadero & Jönsson, 2013). Still, there are some tentative recommendations that can be made, based on existing research. From their review of 75 studies of rubrics, Jönsson and Svingby (2007) identified a number of important design features of rubrics, such as 'specificity' (i.e. whether the rubric is more task specific or more general), 'scoring strategy' (i.e. whether the scoring is holistic or analytic) and 'quality levels' (i.e. the number and type of levels of quality). Importantly, these recommendations differ depending on the main purpose of using rubrics. For summative purposes, Jönsson and Svingby suggest using task-specific rubrics with few levels in order to increase reliability. Such a design would not, however, be appropriate for formative

purposes since fewer quality levels would make the rubric less useful for providing and understanding constructive feedback. Furthermore, a rubric that can only be used for one particular task is of less use for formative assessment. As it may take a while for students to learn how to use a rubric, the rubric needs to be applicable to at least a group of similar tasks (called 'task-type rubrics' in Dawson's terminology). There are examples of successful use of generic rubrics, for instance, a study by Balan (2012), where a generic rubric for mathematical problem-solving in upper secondary school was used. On the other hand, it took the students several weeks to comprehend and use the rubric for self- and peer assessment purposes. Learning to use this rubric was therefore an investment, which paid off since the students were able to use the rubric during the remaining part of the course, whenever they encountered a mathematical problem-solving situation. Generally, however, it is doubtful whether such generic rubrics provide enough guidance to students in higher education contexts, since the students often work more independently and within shorter courses or modules.

Another important design feature, which is affected by the purpose of the assessment, is the scoring strategy. Students' strengths and weaknesses are the raw material for formative assessments, and this information becomes visible through an analytic scoring strategy. But in contrast to summative assessments, the assessment of separate criteria does not have to be summarised into a total score. A rubric does not even have to include any numerical scores. As noted by Dawson (2015), the quality levels in a rubric may come from taxonomies like the SOLO taxonomy, but they can also be expressed through grade levels or statements of student proficiency. Consequently, there is nothing inherent in the design or use of rubrics that has to do with quantification of knowledge.

The distinction between identifying strengths and weaknesses versus scores is important for several reasons. While scores may also be used to identify strengths and weaknesses, their main merit is that the addition of scores is compensatory. This means that weaknesses in some areas can be compensated for by strengths in other areas. On the other hand, as soon as the aggregated score is calculated, the original pattern of strengths and weaknesses disappears (Sadler, 2005). A summary score therefore has little value for formative assessment purposes, which means that formative assessment should primarily focus on an analytical assessment of strengths and weaknesses, not on aggregated scores or holistic judgements.

A design feature of rubrics that has become more or less common practice is to have the same number of quality levels for all criteria. Even if this might give an impression of coherence and logic, it has been shown that such an arrangement can affect the scoring process, so that assessors to a greater extent give the same score in relation to several criteria (i.e. a kind of halo effect). As a result, validity is affected negatively since the variance of what is actually assessed becomes narrower than in the original construct. This problem may disappear, however, by designing rubrics with a different number of quality levels for different criteria (Humphry & Heldsinger, 2014).

Jönsson (2014) has done an in-depth study on students' use of rubrics in professional education, where he investigated three different assessment situations,

which included the use of rubrics. In one case, public health students in a course on statistics and epidemiology constructed a questionnaire and a database with fictitious data. In the second case, real estate broker students systematically reviewed a house. And in the third case, dental students were assessed by peers when communicating with a simulated patient. As shown by results from questionnaires and interviews with students, students in all cases perceived the criteria as both comprehensible and useful. They also actively used the rubric. For instance, the real estate broker students used the rubric for planning, as well as to monitor and evaluate their task performance. The dental students used the criteria to discuss each other's performances and give each other feedback, as well as to reflect about their performance as professionals beyond the scope of the assignment. From these cases, Jönsson identified two important factors, which seemed to facilitate the communication of expectations to the students. These factors are called 'accessibility' and 'alignment'.

The teachers made the rubrics *accessible* to the students, both in terms of understanding and availability. First, the teachers explained the meaning of the criteria in the rubric, which was done criterion by criterion, holistically or by letting the students use the criteria during an instructional event. Second, the teachers presented the rubrics to the students *before* they carried out their assignments. This means that the students could use the rubrics as guides when planning, monitoring and evaluating their performance. Third, the teachers made the rubrics available to the students by publishing the documents digitally or by handing them out on paper. The students did not, therefore, have to rely on teachers' oral description and their own interpretations and notes. Instead, the criteria could be reviewed and discussed both individually and among peers. The students could also have the rubrics beside them when they performed their assignments or, in the case of the dental students, assessed their peers.

Alignment refers to how the rubrics were aligned with the tasks. In these cases, the rubrics were analytic and of task-type specificity. But they were also 'direct' by focusing on the performance of skills that students were expected to master. The directness of rubrics has the advantage of potentially aiding in the improvement of the skills sought to assess. This is in opposition to indirect assessments, where the connection between the skills sought for and what is actually assessed may not be clear to the students, due to the sometimes complex transformation of scores to interpretations of student performance. In this case, the directness of the criteria was seen to facilitate student engagement with the rubrics, to guide their performance and as tools for self-assessment and reflection.

Taken together, the possibility to provide recommendations on how to design and use rubrics is limited by the lack of a common terminology. There is also still much work to be done in systematically investigating different aspects of design and use (Dawson, 2015; Panadero & Jönsson, 2013). Still, there are some recommendations that can be made for designing and using rubrics to support AfL practices:

- Use an *analytic scoring strategy without summarising into a total score*, so that the aspects to be assessed are explicitly spelled out and – most important – strengths and weaknesses in relation to individual criteria are discernible.
- Use *several quality levels*, so that the quality sought becomes visible to the students and for aiding in producing and understanding constructive feedback. It should be noted that having the same number of levels for all criteria may compromise the validity of scoring.
- Use *task-level specificity*, so that rubrics are neither too closely tied to the particular task nor too generic. Instead, rubrics need to be applicable to several, but similar, tasks assessing the same competency.
- Make the rubrics *accessible* to the students by (a) explaining the criteria and quality levels, (b) making the rubric available, digitally or on paper, and (c) providing the students with the rubric before they perform the task.
- Use *direct criteria*, so that they may guide student performance and facilitate self-assessment and reflection.

Towards a Sustained and Widespread Implementation of Rubrics in Higher Education

As suggested by the research reviewed above, a wider implementation of rubrics in higher education could have great potential, both for students' short-term performance and for their development towards independent learners. To date, however, research on larger-scale implementations of rubrics for formative use is lacking, and most studies are based on small samples and short-term interventions. This is particularly true in the higher education context, whereas in school settings, there are a number of studies with larger samples and longer interventions (Brookhart & Chen, 2014; Panadero & Jönsson, 2013).

Although few, there are some studies in a higher education context, with at least medium-sized samples. For instance, besides the previously mentioned study by Jönsson (2014) in which 166 students participated, there is another study by the same author with a large sample size (Jönsson, 2010). In this study, a rubric was shared with pre-service teachers before performing an online examination of complex teaching skills. The same examination was studied during three consecutive years with a cohort of pre-service teachers in science and mathematics ($n = 170$, 154 and 138). Some changes were implemented after the first year, which were thought to increase the transparency of the assessment. For instance, there were some clarifications in the rubric, and the students could access exemplars illustrating the quality levels in the rubric. These changes in transparency lead to major improvements in students' performance on the examination, especially for those students who claimed to have used the rubric actively during the examination. In both of these studies, however, the rubric intervention is limited. In Jönsson (2010), students were provided the rubric (along with exemplars) before the examination,

and in Jönsson (2014), the students were provided the rubric before they performed an authentic task. In the latter case, the criteria and levels of quality were explained to the students, but students were not offered any training in using the rubric in either of the cases.

In a study by Kocakülah (2010), on the other hand, physics students (n = 153) took active part in constructing a rubric. First, the students were introduced to the concept and design of rubrics and were then asked to design a rubric on their own in groups. The students were free to select the type of rubric they wanted, but afterwards they had to present their rubric to the whole class and justify their choices. From the presentations, the best rubric was chosen, collectively modified and finally used by the students while doing the test. Another example of a more extensive intervention is the study by Reitmeier et al. (2004) mentioned previously. Here rubrics were used for teacher, self- and peer assessments, and all students were required to self- and peer assess repeatedly over the course of a semester. As can be seen, these interventions are not as limited as in the above-mentioned research. Interestingly, however, positive effects are reported in a number of studies in higher education, seemingly regardless of the extent of the intervention. One exception is a study by Green and Bowser (2006), where there were no significant differences between scores from students who had access to a rubric and those who did not. As pointed out by Reddy and Andrade (2010), however, the sample size in this study was very small (n = 16), which – together with the limited intervention – makes it difficult to draw any conclusions from this comparison. But this is not the only example. In Panadero et al. (2013), investigating 69 pre-service teachers, no significant effects for students' performance or self-efficacy were documented. On the other hand, students preferred the use of rubrics as compared to so-called scripts, and there was a positive effect on students' SRL strategies.

Taken together, the general picture is that the use of rubrics is appreciated by the students and that rubrics often contribute to improved performance or students' SRL strategies. As noted by Panadero and Jönsson (2013), this is in contrast to studies performed in school settings, where longer interventions are often needed in order to produce a clear effect. Obviously, many students in higher education are capable of using rubrics productively, for instance, to monitor and evaluate their work, with only minimal guidance.

Even if the findings presented above may seem promising, it should be kept in mind that there are still relatively few studies addressing formative aspects of using rubrics and that most of these are based on small samples and short-term interventions. Furthermore, in most studies, students could be characterised as 'rubric neophytes', which means that not much is known about how rubrics are used by students more accustomed to such instruments or how students are affected when they are exposed to rubrics during longer periods of time. For instance, fears have been voiced about the limiting effects of rubrics (e.g. Sadler, 2009; Wilson, 2006). By providing criteria beforehand, we may guide some students, helping them to focus their efforts on what is considered important and worthwhile, but we may at the same time restrict others.

With reference to the fears of hampering students by providing them with criteria specified beforehand, Sadler (2009) makes a case for 'emerging criteria'. This means that assessors should address criteria that surface in the moment of assessing a particular piece of work – much like the appraisal by connoisseurs of art, wine, etc. The main argument is that when breaking down holistic judgements into more or less discrete components, these components – no matter how many they are and no matter how carefully they are selected – cannot sufficiently represent the full complexity of the multi-criterion qualitative judgement made by the connoisseur. Furthermore, qualities not represented by the criteria might be filtered out and not taken into account by the assessors. Instead of relying on analytic assessment and pre-set criteria as a vehicle for transparency in assessment, Sadler therefore argues that students need to develop a conceptualisation of what constitutes 'quality' by repeatedly evaluating authentic work.

A problem, however, is that novices may not know what to look for in authentic work. This is evident in a number of studies. An illustrative example is provided by Orsmond and Merry (1996), where students were asked to assess each other's work. Even though all criteria were explained to the students, they were unable to recognise some of these criteria in the work by their peers. For instance, a majority of students had actually drawn a 'clear and justified conclusion' (which was a criterion), but did not know it.

In our understanding, the question of using pre-set criteria or developing a conception of quality through evaluating authentic work is therefore not a question of either one or the other. Rather, what seems to be needed is an integration of both. Students need explicit criteria to know what to look for in authentic work, but they also need to experience authentic work in order to know how the criteria may be realised. Rubrics can provide a scaffolding structure for students when learning to identify indicators of quality, but like other scaffolding structures, it can be disregarded if not needed and gradually phased out as the students become more independent.

This is exemplified in the study by Jönsson (2014), where the rubrics provided an important support for low-performing/low-confidence students, whereas some confident students actually refused to use the rubric because they wanted to manage on their own. The findings from this study thus suggest that rubrics do not necessarily limit high-performing/high-confidence students, since they may choose not to use it.

Conclusion

The main idea permeating this chapter is that rubrics can not only be used by teachers to increase the reliability and validity of summative assessments but also shared with the students. Rubrics then become a tool for communicating expectations, which is a fundamental requirement for successful AfL implementation. As a tool for communication, rubrics can be used to support different AfL processes,

such as giving and understanding feedback, as well as self- and peer assessment. Perhaps even more important, the transparency provided by the use of rubrics has been shown to support students' SRL, so that students are able to plan, monitor and evaluate their performance with the aid of a rubric. In order to enhance these effects, rubrics should be designed to promote formative assessment practices and also made accessible to the students.

Taken together, a wider implementation of rubrics in higher education could have great potential, both for students' short-term performance and for their development towards independent learners. However, since research on long-term use of rubrics is lacking, as well as research on the effects of rubrics for students with different performance backgrounds or learning orientations, a general recommendation could be to advance slowly towards a sustained and widespread implementation of rubrics in higher education.

References

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation, 10*(3), 1–11.

Balan, A. (2012). Assessment for learning. A case study in mathematics education (Unpublished doctoral dissertation). Malmö University, Malmö, Sweden.

Brookhart, S. M., & Chen, F. (2014). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343–368.

Case, S. (2007). Reconfiguring and realigning the assessment feedback processes for an undergraduate criminology degree. *Assessment and Evaluation in Higher Education, 32*(3), 285–299.

Dawson, P. (2015). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment and Evaluation in Higher Education*. doi:10.1080/02602938.2015.1111294.

Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration, 45*(1), 185–202.

Hanna, M. A., & Smith, J. (1998). Using rubrics for documentation of clinical work supervision. *Counselor Education and Supervision, 37*(4), 269–278.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Humphry, S. M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher, 43*(5), 253–263.

Jönsson, A. (2010). The use of transparency in the 'interactive examination' for student teachers. *Assessment in Education: Principles, Policy & Practice, 17*(2), 185–199.

Jönsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63–76.

Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment and Evaluation in Higher Education, 39*(7), 840–852.

Jönsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144.

Kocakülah, M. S. (2010). Development and application of a rubric for evaluating students' performance on Newton's laws of motion. *Journal of Science Education and Technology, 19*(2), 146–164.

Linn, R. L., Baker, E., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

Lucander, H., Knutsson, K., Salé, H., & Jönsson, A. (2012). "I'll never forget this": Evaluating a pilot workshop in effective communication. *Journal of Dental Education, 76*(10), 1311–1316.

Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*(3), 239–250.

Panadero, E. (2011). Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics. (Unpublished doctoral dissertation). Universidad Autónoma de Madrid, Madrid, Spain.

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.

Panadero, E., & Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited. *Educational Research Review, 9*, 129–144.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy and Practice, 21*(2), 133–148.

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts: Effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806–813.

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: Effects on first year university students' self-regulation and performance. *Infancia y Aprendizaje: Journal for the Study of Education and Development, 37*(1), 149–183.

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts: Effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation, 39*(3), 125–132.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448.

Reitmeier, C. A., Svendsen, L. K., & Vrchota, D. A. (2004). Improving oral communication skills of students in food science courses. *Journal of Food Science Education, 3*(2), 15–20.

Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment Research and Evaluation, 15*(8), 1–9.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education, 30*(2), 175–194.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education, 34*(2), 159–179.

Schamber, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. *The Journal of General Education, 55*(2), 103–137.

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

Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Portsmouth, NH: Heinemann.

Yopp, B. D., & Rehberger, R. (2009). A curriculum focus intervention's effects on prealgebra achievement. *Journal of Developmental Education, 33*(2), 28–30.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147.

Chapter 8
Students' Experiences of Assessment for Learning

David Carless

Abstract This chapter is based on 90 individual semi-structured interviews asking students about their experiences of assessment for learning in five disciplines: architecture, business, geology, history and law. Four features of assessment valued by students are discussed: assessment mirroring real-life uses of the discipline, flexibility and choice, developing understanding of expectations and productive feedback processes. A striking finding was student cynicism about rubrics or lists of criteria in contrast to their enthusiasm for exposure to exemplars of previous student work. Two challenging modes of assessment are also focal points for analysis: the assessment of participation and group assessment. Assessed participation through verbal and written means was perceived quite positively by student informants. Group assessment attracted mixed student views and might be enhanced by feedback processes involving interim reports of progress to discourage procrastination and free-riding. The chapter concludes with some discussion of how the analysis of exemplars and productive feedback designs could be scaled up and further investigated.

Introduction

Assessment for learning (AfL) aims to change assessment from something done to students to something done with, and for, them (Brown & Knight, 1994). A repercussion of this goal is that it is vital to investigate students' perceptions of assessment. There is quite a lot known about the student experience of assessment, but the topic remains relatively sketchily explored, although for an exception, see Jessop (this volume, Chap. 4). It is a trend in the UK National Student Survey, for example, that students perceive concerns about how assessment and feedback are handled (Williams & Kane, 2009), but such relatively broad-brush treatments do not provide much detail of the student reaction to assessment. The aim of this chapter is to investigate how students perceive key aspects of their assessment experience.

D. Carless (✉)
Faculty of Education, University of Hong Kong, Hong Kong, Hong Kong
e-mail: dcarless@hku.hk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_8

Understanding the student response is a facilitating factor for the scaling up of AfL because the student perspective lies at its heart.

My recent research investigated the assessment practices of five teacher recipients of awards for excellence. In the book treatment of that research, student perceptions were triangulated with teacher interviews and classroom observational data and weaved within narratives of the courses (Carless, 2015a). In this chapter, I want to bring together students' responses in a more focused way to analyse their perceptions of the assessment processes which they experienced in the modules under investigation.

Literature Review: Students' Perspectives on Assessment

Rather than attempting a more wide-ranging review of literature, for the purposes of this chapter, I discuss selected literature which speaks to the issues to be discussed in the Findings. The rationale for this strategy is to provide a coherent treatment of key issues raised by student informants.

The existing knowledge base on students' responses to assessment suggests a number of themes. Students seem to welcome alternative assessment when it seems fair and relates to real-life application of disciplinary knowledge (McDowell & Sambell, 1999). Students are sometimes reported to be unreceptive to innovative assessment but may also relish some variety compared to what they have done before (Carless & Zhou, 2015). Students' response to assessment emanates from the totality of their previous experiences of learning and being assessed (Boud, 1995). Accordingly, it is students' perceptions rather than the actual features of assessment tasks which have most impact on how they respond to assignments (Lizzio & Wilson, 2013).

Students seem to favour assessment designs which involve some choice for them to tailor assignments to their preferences or capabilities (Bevitt, 2015; Lizzio & Wilson, 2013). Personal interest in the task improves performance and enhances persistence in the face of adversity (O'Keefe & Linnenbrink-Garcia, 2014). Choice also seems to encourage students to adopt deep approaches to learning (Craddock & Mathias, 2009), whereas lack of autonomy may increase anxiety and use of surface learning strategies (Coutts, Gilleard, & Baglin, 2011).

Students are more likely to have positive perceptions of assessment tasks when it is clear what is required. Rubrics bring much-needed transparency to the assessment process, and there is potential for positive impacts on student learning (Jonsson and Panadero, this volume). On-display assignments in which student work is openly visible rather than being private are a useful means of indicating how rubrics are operationalized (Hounsell, McCune, Hounsell, & Litjens, 2008). Exemplars of student work are a further means of illustrating how quality is achieved, and there is plenty of evidence that students are positive about being exposed to exemplars (Hendry, Armstrong, & Bromberger, 2012; Lipnevich, McCallen, Miles, & Smith, 2014). A concern discouraging teachers from using exemplars is that some students

may interpret them as model answers to be imitated, and this may reduce student creativity or lead to formulaic unimaginative work (Handley & Williams, 2011). The sensitive handling of the teacher-led discussion phase about the exemplars may be a means of reducing this problem by unpacking the nature of quality work (To & Carless, 2015).

Students' concerns about the value and usefulness of feedback processes in higher education have been well-rehearsed over the last 15 years or so. For many students, end-of-course comments often seem like a perversely belated revelation of things that should have been made clear earlier (Crook, Gross, & Dymott, 2006). A key strategy to tackle this issue of timeliness is to embed integrated cycles of guidance and feedback within the learning activities for a course (Hounsell et al., 2008). A related line of thinking involves repositioning feedback as a fundamental component of curriculum design rather than a marking and grading routine through which information is delivered by teachers to learners (Boud & Molloy, 2013).

Group assessment is a mode of alternative assessment which often provokes negative reactions amongst students (Flint & Johnson, 2011). Students seem to find it difficult to work effectively in groups and are concerned that assessed group work may negatively impact on their overall grades (Pauli, Mohiyeddini, Bray, Michie, & Street, 2008). Group work is often perceived as unfair because of the free-riding phenomenon (e.g. Davies, 2009). Students may not tackle group work together, instead dividing up the tasks and doing them individually (Brown & McIlroy, 2011). This works against a powerful rationale for working in teams in that the more complex learning is, the less likely that it can be accomplished in isolation from others (Boud, 2000). Students with negative experiences of group assignments express a need for more teacher involvement to support group processes (Volet & Mansfield, 2006).

Another potentially controversial mode of assessment is assessing participation. This should not imply grades being awarded for attendance but could be worthwhile if it involves well-defined intellectual contributions to the course. Assessment of student participation may engender various benefits in terms of student preparation prior to class, development of oral communication skills and regular engagement and involvement (Armstrong & Boud, 1983). Students seem to appreciate course material more when participation is demanded but at the same time report being less likely to enrol in such courses because of the anxiety it provokes (Frisby, Weber, & Beckner, 2014). Probably the biggest challenge for the assessment of participation is that students often have limited understanding of how their participation grade is determined due to its subjectivity and somewhat ill-defined nature (Mello, 2010).

Method

The basis for this chapter is 90 individual semi-structured interviews with 54 undergraduate students at the University of Hong Kong, an English-medium research-intensive university. Some students were interviewed more than once,

and they came from the disciplines of architecture, business, geology, history and law. Thirty-nine of the student informants were Hong Kong Chinese, 11 were mainland Chinese, 3 European (all studying business) and 1 South Asian (studying architecture).

Students were interviewed individually for about 30–45 min to ascertain their perceptions of different key issues which related to their experiences in the courses being researched. Aspects investigated were wide-ranging, including students' perceptions of the assessment tasks in the course, their views about rubrics and exemplars, their opinions of feedback processes and other relevant course-specific issues. Not all of these issues were covered within a single interview because the aim was to elicit students' views which most resonated with their immediate assessment experiences.

The analysis involves the interpretation of a large corpus of student interviews. In view of the inevitable difficulty of selecting representative student quotations and to avoid repeating quotations appearing in Carless (2015a), I have decided to adopt a somewhat unusual procedure. In what follows, I summarize rather than quote what the students said. Despite limitations that this entails, my hope is that the student views emerge whilst also saving space for bringing out commentary, inferences and implications.

Findings

Students carried out four of the most common modes of assessment in contemporary higher education: examinations; written assignments, such as essays or reports; oral presentations; and group projects. There were also a variety of other tasks. In architecture, students were assessed on a portfolio of designs which they had developed during the semester. In law, students curated over time an analysis of legal cases reported in the local media through a 'reflective media diary': an assessment task which combined elements of traditional portfolio and e-portfolio. In business and history, students were assessed on participation comprising both written and verbal participation elements. There were also some discipline-specific assessment tasks. In geology, students carried out laboratory reports on features of rocks which also rehearsed some of the skills needed in the subsequent examination. In law, there was an option of a 'photo essay' whereby students photographed a potential tort law issue from daily life and wrote a short legal analysis. In history, students could choose between a fieldwork report on a museum visit and a 'scavenger hunt': an Internet-based simulation involving visiting sites of historical significance and scavenging for clues.

First, I discuss students' perceptions of four key task features which were centrally evident in the data: assessment mirroring real-life uses of the discipline, flexibility and choice, understanding expectations and productive feedback processes. These relate to the AfL strategies of productive task design, developing student understanding of quality and effective feedback processes (Carless, this

volume, Chap. 1,). Second, I focus on two specific challenging modes of assessment: assessing participation and group assessment because from a student perspective, they are at the same time potentially worthwhile, yet often problematic.

Assessment Mirroring Real-Life Uses of the Discipline

Students reacted enthusiastically to assessments which bore a relevance to real-world applications of the discipline. For example, law students made a number of positive comments about the assignments which were related to how law was practised in daily life. These included their reflective media diary about legal cases reported in the media as well as an assessed report of a self-organized visit to a labour tribunal. Their main reservation was the time-consuming nature of some of these activities, particularly if the amount of assessment weighting was relatively low.

History students appreciated the fieldwork report on their museum visit because it helped them to understand how museums are arranged to highlight certain messages and downplay others. The scavenger hunt drew more mixed reactions from students in that for many of them, it was difficult for them to understand and appreciate. Those who participated in it did, however, report finding it rewarding. A related issue was that a wide variety of tasks, including some innovative ones, could be confusing to students, particularly those who are less committed or less able.

In architecture, students expressed appreciation of a number of elements which related to real-life applications of their discipline. Their work was contextualized around the design of houses for a mainland Chinese village a few hours away from the university which underscored the real-life nature of architectural design. Students also compared the processes of developing a design portfolio for assessment with how professional architects work. One of the students viewed his portfolio as representing how designs became real in that the portfolio is a kind of publication which synthesizes ideas. Another student emphasized the importance in a student's portfolio of showing the procedures used to arrive at the solution in contrast with the professional architect who is more focused on displaying various perspectives. The iterative process of developing designs, presenting them publically for critical review, receiving feedback and then revising was also seen by students as mirroring the way architecture operates in real life.

Flexibility and Choice

Students reported valuing assessments which involved some flexibility and choice. For example, in the history case, students had different options both in relation to the tasks and how they could be approached. For their individual project, they could choose from a long list of possible topics or suggest one of their own (see also

Carless, 2015a, Chap. 4). Students could produce conventional written assignments or more innovative ones, such as videos or podcasts. A recurrent theme in the student data was the term 'flexibility' which students viewed as offering the chance to work on something which had a personal meaning for them and also provided them with opportunities to produce their best performance or avoid tasks with which they are unfamiliar. Choice also facilitates learner autonomy which a number of students favoured.

In law, there was some choice of assessment tasks and also some flexibility in terms of weighting of assessment. For example, students who were less confident about their exam performance could reduce the exam weighting from 60 to 40 % by doing alternative tasks, such as a research essay. The idea of choice enabling students to diversify risk was particularly attractive to them. A number of students, however, expressed concerns about workload. On the one hand, the coursework options were attractive and a positive learning experience, whereas on the other hand, the processes engendered considerably more overall workload than those courses assessed entirely through examinations.

Although students expressed a wide range of positive views about choice and flexibility, there are some potential disadvantages of choice. Students may need some guidance on what kinds of choice are available and how they relate to the bigger picture of the learning outcomes they are trying to develop. Students might choose easy options, such as something similar to what they have done before. Even worse, there may be concerns that providing choice might make cheating or plagiarism easier.

Developing Understanding of Expectations

Students need to understand the goals of assessment tasks and the standards expected. In all the courses I observed, teachers provided details at the outset of assessment requirements and associated criteria in the form of rubrics. Somewhat to my surprise, the majority of students, reported a rather cynical view of rubrics or lists of criteria, using adjectives, such as 'vague', 'unclear', 'all the same' and 'inert' to describe them. Students also commented that they were not convinced that the criteria shown to them fully represented how teachers evaluated student work. Teachers' impressions and personal judgments were deemed by students to be more significant than what was stated in the criteria: a point also made in the literature (Bloxham, Boyd, & Orr, 2011).

Instead, students particularly valued exposure to exemplars which could help them understand what teachers were looking for in a particular assessment task. Students perceived that exemplars are more effective than rubrics at indicating how assessment requirements are operationalized. They reported them as being valuable in indicating what was required and also as a useful benchmark of the academic standards. Students felt this was particularly necessary in relation to innovative tasks

with which they are not familiar. These unfamiliar formats provide a sense of anxiety as students are not sure how to obtain a high grade, and exemplars can relieve some of these concerns.

Students were asked how they used exemplars to inform their own assignments and there were a variety of responses. Some students reported using the exemplars to help them understand the abstract assessment criteria in more concrete terms. Other students stated that they could obtain some ideas and inspiration from the exemplars. Students also reported using exemplars as a template for their own work, for example, by imitating the format and then adding some ideas of their own. Overall students were highly positive about being exposed to exemplars and expressed a wish that more teachers would make this a regular part of their practice.

Productive Feedback Processes

Students frequently expressed the view that they wanted to receive feedback during the process of their work so that they could act on it in a timely way for the teacher who had provided it. Timeliness was thus a key issue from their point of view. They did not perceive much scope for transferring end of module feedback from one course to another: they might not remember feedback; assignment formats and content varied; and they believed that different teachers had different requirements.

There were some feedback design elements in the case studies (see also Carless, 2015a, Chap. 11). In architecture, students presented their design work in progress for 'critical review'. These processes provided individual feedback for the presenter and also acted as 'on-display' assignments which could facilitate wider discussion of architectural issues. Particularly fruitful from the student perspective was the final review when all the student designs were displayed in the studio, and there were ample opportunities for peer feedback, comparison and discussion.

In the history case, a feedback design feature was that for their individual project, the teacher required a draft worth 10 % and final version worth a further 30 %. This was generally popular with students as it provided timely feedback which students could act on. Two drawbacks also emerged. First, there is a danger that feedback on drafts can create student dependency on the teacher and fail to develop self-evaluative capacities. Second, when feedback on a draft does not connect or is misunderstood, this can lead to frustrations: a particular issue for lower-achieving students.

In the business case study, a prominent feedback element was in-class dialogues facilitated by a small class size (see also Carless, 2013a). Students spoke positively of the teacher's skill in creating an interactive classroom with plentiful verbal feedback which challenged them to raise their thinking to a higher level. The amount of classroom time devoted to interaction also represented some tensions: some students would have preferred more content to be delivered; other students felt that discussions were sometimes tangential to core course content.

A useful feedback design feature in geology was early interaction around student topics for their group project which provided students with timely guidance that they were on the right track. For feedback on the oral presentation for the group project, students perceived some inconsistencies in standards between the different teachers. They generally seemed to prefer the more easy-going encouraging feedback of some tutors, rather than the more critical but perceptive analyses of others.

In the law case, a special feature was immediate interactive verbal feedback after the exam: students were invited to remain in the exam hall to discuss their answers and this was also supplemented by online discussion. Students expressed appreciation of the opportunity for prompt discussion of their exam performance. For some students, however, receiving immediate feedback about the exam was perceived as anxiety-inducing so they preferred not to join the discussion in case it revealed discouraging failings in their performance (Carless, 2015b).

Assessing Participation

Assessing participation was a significant feature of the business and history case studies. Both of these cases involved both oral and written participation: in business, verbal in class and written through a blog, and in history, small-group tutorial participation and written assessment of short in-class responses. The combination of verbal and written involvement represents a positive feature which allows students of different personalities and preferences to participate in alternative ways.

There was an atmosphere in the business case study which was quite different to other university classes I have observed. There was a kind of 'productive tension', a feeling that something interesting or provocative might happen and that participants should be well-prepared. The teacher might ask at any point a challenging individual question and may also interrogate the resulting answer. Accordingly, students reported that they were more concentrated and better prepared in this class as opposed to others. This challenging atmosphere was also balanced by warmth, empathy and trust between participants (Carless, 2013b). The participation grade was one of the factors that students reported as encouraging them to maintain their concentration and express their thoughts. They did, however, express the view that they contributed because they wanted to do so and had something to say, not merely to gain marks.

In the history case, an innovative strategy was the weekly assessed written in-class 'One sentence responses': concise answers to a question that related to the topic to be addressed in the following session. Students perceived this strategy as being novel and fun and providing a useful entry point to the content of the next session. In the following class, the teacher also carried out some follow-up on the student responses, including displaying examples of good contributions which acted as a form of feedback and clarification of expectations. Overall, the short-written responses seemed to bring a number of benefits as they encouraged student

expression of thoughts in a concise form and prompted them to reflect on what was coming next in the course.

Students were generally positive about the impact of assessed participation on their engagement and the classroom atmosphere. Some students even expressed the view that they might not attend class so regularly if participation was not assessed. Students evidenced, however, quite a lot of doubts and confusion as to how participation grades were awarded even though rubrics were made available to them. They generally perceived participation grades as being subjective and hard to judge reliably. As long as the assessment of participation did not count for too high a weighting, students were generally acquiescent of these limitations. A factor supporting their tolerance of assessed participation was their trust in the award-winning teachers.

How does the assessment of participation fare in relation to the four features of assessment valued by students discussed above? There are linkages with real life in that participation in discussion and debates, for example, in meetings, is an important part of the future workplace. The way assessing participation was implemented in the cases also resonates with the concept of choice in that students have some flexibility in devoting more effort to expressing their thoughts verbally or in writing. Clarifying expectations is a challenge because it is difficult to model or exemplify what good participation looks like, and rubrics may fail to do this adequately. Feedback on participation is also challenging and potentially time-consuming but, if carried out effectively, could contribute to clarifying expectations about what good participation entails.

Overall, the assessment of participation seems a somewhat contentious issue in terms of the challenge of reliable assessment of something as potentially vague as participation. What struck me was the positive impact it had on student engagement and classroom atmosphere. Assessing clearly defined contributions, with well-defined criteria, could form part of a worthwhile overall assessment design (Carless, 2015a).

Group Assessment

Group assessment was a core element of both the business and geology cases (see Carless, 2015a, Chap. 6). Here I summarize student perceptions and then conclude with some recommendations. Students expressed mixed reactions and a variety of experiences. Some groups were reported as operating cohesively and a capable coordinator was an important feature of that kind of group. In other teams, the group leader ended up doing most of the work: sometimes they seemed quite happy in that role because they perceived themselves as more capable, more motivated or with a higher intensity of desire for a high grade. On other occasions, there appeared to be simmering resentment at the failure of teammates to attend meetings, respond to electronic communication or produce work in a timely fashion. Procrastination was

a regular problem and often resulted in the most responsible team member ending up doing the work left by a free-rider (cf. Davies, 2009).

A key issue is the amount of guidance and support which is offered in relation to the processes of group assessment. In one of the introductory lectures in geology, the teacher explained the guidelines for the group assessment and shared some tips and experiences. He provided guidance on how to narrow down the focus of a topic and spoke about the need for some novelty in what was being investigated with an aim of generating insight. It was evident that some students were not fully concentrated on what the teacher was saying; rather, they were starting to negotiate with classmates on the composition of the groups. At the time, this seemed to me a missed opportunity to pick up on some important cues as to successful execution of the group project. On further reflection from the student perspective, this may be a logical response as the members of your team are a significant factor in completing a group project well and attaining a high grade.

Students perceived a number of benefits accruing from group assessment. These included learning from peers, developing teamwork and interpersonal skills and the social aspects of collaborating with peers. Peer support can provide a sense of safety and the feeling of not being on your own seemed to be particularly reassuring for first year students. In the business class, students generally seemed to relish working in groups. Some of these teams reported spending a lot of time discussing and negotiating and this seemed to be facilitating rich learning experiences.

Relating the assessment of group projects with the previous themes, work in teams mirrors the future workplace. It can involve some flexibility and choice in relation to topics and group membership. There seem to be some challenges for students in understanding expectations, especially in relation to balance between process and product. To what extent are both the processes of working in teams and the final product equally valued and rewarded? A useful strategy to enhance group assessment would be to enhance the integration of guidance and feedback processes. This might involve interim reports of work in progress which could be in the form of online reports or brief in-class progress updates. Requiring students to report on the progress of their projects reduces procrastination, encourages student accountability, discourages free-riding and provides opportunities for timely feedback and guidance, including student self-evaluation.

Discussion

The findings carry some similarities with previous literature and also add some further dimensions. For example, the data reinforce student enthusiasm for assessment mirroring real-life uses of the discipline and reiterate some of the challenges of group assessment. In comparison with these well-worn themes, student choice has been relatively modestly treated in the existing assessment literature. Choice can be a way of providing students with options to cater for their strengths and preferences or perhaps, more fundamentally, a means of generating some student

ownership of the assessment process. It is also possible that students may achieve superior outcomes when they have choice because we generally perform better when we possess some form of agency in relation to academic work. The concept of student choice also resonates with emerging recent trends in the development of personalized learning and new technologies where students co-construct their own learning pathways and learning environments.

Relatively few studies have elicited students' opinions of both rubrics and exemplars. The findings from this study strikingly indicate that students find exemplars more useful than rubrics (cf. Jonsson and Panadero, this volume). Exemplars are perceived as concrete illustrations of how an assignment can be tackled, whereas students perceived rubrics as vague. The findings can be contrasted with Lipnevich et al. (2014) where students welcomed both exposure to exemplars and the rubric, but it was the 'rubric only' group who improved more than 'exemplars only' or 'rubric and exemplars' groups in a quasi-experimental study. Whether and how exemplars support students to develop superior learning outcomes bears further investigation. The data also reinforce previous studies which report student enthusiasm for exemplars (Hendry et al., 2012). Perhaps it is not surprising that students are positive about exemplars because they make the processes of tackling assignments easier and also provide psychological reassurance. Ways in which exemplars can be used to illustrate the nature of quality without reducing student creativity and intellectual challenge require further research.

The findings in relation to feedback processes provide further evidence of key challenges and dilemmas which are also taken up elsewhere in this volume. What is a healthy balance between teacher guidance, peer feedback and the development of student self-evaluative capacities? How might students be enabled to transfer learning from feedback from one task to another or from one course to another? How can timely in-course feedback processes be developed within the structural challenge of assessment mainly coming at ends of semesters? How can feedback be honest and critical without upsetting student emotional equilibrium?

The generally positive student views on the assessment of participation might stimulate further attention to this topic which seems to have been more widely discussed in the North American literature than in analyses from Europe, Asia or Australia. The informants in this study mainly reported that participation grades prompted engagement, encouraged preparation before coming to class and played a role in enlivening classroom atmosphere. An important facilitating feature is for teachers to explain in the course documentation, during the first class meeting, and periodically thereafter, that good participation is not simply attending class or talking a lot, it depends on the quality of the contributions not their quantity (cf. Mello, 2010). A useful feature of the assessment of participation in the study was the use of both assessed verbal participation and assessed written participation: in other words, there was choice in mode of participation which may reduce anxiety and allow students some degree of flexibility.

The use of short in-class written responses in the history case seems to be a useful alternative to other types of personal response tasks, such as contributions posted on learning management systems, blogs or wikis or the use of electronic

voting systems. At the first opportunity, I experimented with this form of assessment in my own teaching, also eliciting positive student responses (Carless & Zhou, 2015). Further research into the value of assessing short in-class or online responses could be worthwhile. Specific avenues in relation to Chinese students might involve investigating the extent to which they may benefit from, or appreciate, incentives to participate actively in class and/or whether they may prefer to contribute their thoughts in writing rather than verbally.

A number of key implications for practice are worth summarizing. Effective assessment task design includes the development of participation in the discipline through mirroring real-life elements, permits some degree of student choice and flexibility, raises awareness of quality work through analysing exemplars and promotes various forms of guidance and feedback dialogues with peers and teachers. The scaling up of good practice in AfL could also be enhanced by integration of productive assessment tasks and the development of student understanding of the nature of quality and feedback designs (Carless, 2015b).

Conclusion

This chapter has discussed students' perspectives on various aspects of their assessment experience. A number of positive student perceptions are reported, including the use of exemplars to clarify expectations and the design of thoughtful feedback processes. As these are both key AfL strategies, I want to conclude by sketching some prospects for wider implementation and suggest some related avenues for further research.

First, there seems to be some lack of teacher appreciation of the value of exemplars in supporting student capacities to make evaluative judgments (Thomson, 2013). Given that exposure to exemplars is popular with students and has a persuasive academic rationale, this state of affairs needs challenging. Teacher concerns about the use of exemplars need to be interrogated and tackled. Larger-scale studies of the use of exemplars going beyond specific individual courses might provide further evidence of their value. The cumulative impact of exemplars on students over the duration of a programme is also worthy of investigation.

Second, effective feedback processes lie at the heart of AfL but are difficult to manage effectively within the structural challenges of modularized systems in which end-loaded assessment predominates. The need for new ways of thinking about feedback has been highlighted in recent literature (Boud & Molloy, 2013) but still has a long way to go to be scaled up (see also Ajjawi et al., this volume, Chap. 9). The development of teacher and student feedback literacy would be a key facilitating factor for more sophisticated approaches to feedback. Research focused on effective feedback designs at scale and the associated development of feedback literacy are sorely needed.

References

Armstrong, M., & Boud, D. (1983). Assessing participation in discussion: An exploration of the issues. *Studies in Higher Education, 8*(1), 33–44.

Bevitt, S. (2015). Assessment innovation and student experience: A new assessment challenge and call for a multi-perspective approach to assessment research. *Assessment & Evaluation in Higher Education, 40*(1), 103–119.

Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. *Studies in Higher Education, 36*(6), 655–670.

Boud, D. (1995). Assessment and learning: Contradictory or complementary. In P. Knight (Ed.), *Assessment and learning in higher education* (pp. 35–48). London: Kogan Page.

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

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

Brown, C., & McIlroy, K. (2011). Group work in healthcare students' education: What do we think we are doing? *Assessment & Evaluation in Higher Education, 36*(6), 687–699.

Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.

Carless, D. (2013a). Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education: Developing dialogue with students* (pp. 113–122). London: Routledge.

Carless, D. (2013b). Trust and its role in facilitating dialogic feedback. In D. Boud & L. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 90–103). London: Routledge.

Carless, D. (2015a). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.

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

Carless, D., & Zhou, J. (2015). Starting small in assessment change: Short in-class written responses. *Assessment & Evaluation in Higher Education.* doi:10.1080/02602938.2015.1068272.

Coutts, R., Gilleard, W., & Baglin, R. (2011). Evidence for the impact of assessment on mood and motivation in first-year students. *Studies in Higher Education, 36*(3), 291–300.

Craddock, D., & Mathias, H. (2009). Assessment options in higher education. *Assessment & Evaluation in Higher Education, 34*(2), 127–140.

Crook, C., Gross, H., & Dymott, R. (2006). Assessment relationships in higher education: The tension of process and practice. *British Educational Research Journal, 32*(1), 95–114.

Davies, W. (2009). Groupwork as a form of assessment: Common problems and recommended solutions. *Higher Education, 58*, 563–584.

Flint, N., & Johnson, B. (2011). *Towards fairer university assessment: Recognizing the concerns of students*. London: Routledge.

Frisby, B. N., Weber, K., & Beckner, B. N. (2014). Requiring participation: An instructor strategy to influence student interest and learning. *Communication Quarterly, 62*(3), 308–322.

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education, 36*(1), 95–108.

Hendry, G., 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.

Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research and Development, 27*(1), 55–67.

Lipnevich, A., McCallen, L., Miles, K., & Smith, J. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science, 42*, 539–559.

Lizzio, A., & Wilson, K. (2013). First-year students' appraisal of assessment tasks: Implications for efficacy, engagement and performance. *Assessment & Evaluation in Higher Education, 38*(4), 389–406.

McDowell, L., & Sambell, K. (1999). The experience of innovative assessment: Student perspectives. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education* (pp. 71–82). Maidenhead, UK: Society for Research into Higher Education and Open University Press.

Mello, J. (2010). The good, the bad and the controversial: The practicalities and pitfalls of the grading of participation. *Academy of Educational Leadership Journal, 14*(1), 77–97.

O'Keefe, P., & Linnenbrink-Garcia, L. (2014). The role of interest in optimizing performance and self-regulation. *Journal of Experimental Social Psychology, 53*, 70–78.

Pauli, R., Mohiyeddini, C., Bray, D., Michie, F., & Street, B. (2008). Individual differences in negative group work experiences in collaborative student learning. *Educational Psychology, 28*(1), 47–58.

Thomson, R. (2013). Implementation of criteria and standards-based assessment: An analysis of first-year learning guides. *Higher Education Research and Development, 32*(2), 272–286.

To, J., & Carless, D. (2015). Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education*. doi:10.1080/0309877X.2015.1014317.

Volet, S., & Mansfield, C. (2006). Group work at university: Significance of personal goals in the regulation strategies of students with positive and negative appraisals. *Higher Education Research and Development, 25*(4), 341–356.

Williams, J., & Kane, D. (2009). Assessment and feedback: Institutional experiences of student feedback, 1996-2007. *Higher Education Quarterly, 63*(3), 264–286.

Part III
Feedback for Learning

Chapter 9
Contextual Influences on Feedback Practices: An Ecological Perspective

Rola Ajjawi, Elizabeth Molloy, Margaret Bearman, and Charlotte E. Rees

Abstract Critique has been levelled at the use of models for feedback practices that ignore context in health professions education. Models such as the 'feedback sandwich' are often adopted as rules to be followed regardless of the situation. In this chapter, we utilise an updated version of the Bronfenbrenner ecological framework of human development to unpack contextual influences on feedback practices at different levels. The framework seeks to integrate and conceptualise the environment and other influences on behaviour. The implication of the interplay of these networked systems on feedback practices and consequences for learners is that a one-size feedback intervention is not suitable for all situations. Promoting feedback by design involves taking context into account for each of the systems. A step forward in terms of scaling up effective feedback practices would be through using this contextual mapping to improve feedback literacy of students and staff. On the basis of our mapping, we highlight the usefulness of ecological models for research and practice in assessment for learning in higher education and propose recommendations for future research.

Introduction

Feedback is important for learning and is valued by staff and students. Meta-analyses show a beneficial effect of feedback on learning with detrimental effects highlighted in a subset of learners (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Feedback from external sources such as teachers and peers is crucial to the development of learners during higher education (Hattie & Timperley, 2007; Sargeant et al., 2010), particularly given the lack of reliability of self-assessment (Eva & Regehr, 2005; Kruger & Dunning, 1999). Therefore, learners need external

R. Ajjawi (✉)
Deakin University, Burwood, VIC, Australia
e-mail: rola.ajjawi@deakin.edu.au

E. Molloy • M. Bearman • C.E. Rees
Monash University, Melbourne, VIC, Australia

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_9

feedback in order to help them improve their performance and to calibrate their evaluative judgement. Despite consensus in the literature on the *potential* for feedback to promote learning, there are also multiple reports on the problematic nature of feedback in higher and professional education, such as feedback as information transmission (e.g. Barton, Schofield, McAleer, & Ajjawi, 2016; Boud & Molloy, 2013; Carless, 2006; Urquhart, Rees, & Ker, 2014).

Studies within the higher and professional education literature indicate that feedback is most often 'delivered' to the learner without invitation for the learner to engage in the process (Molloy, 2009; Nicol, 2010), the information is focused on deficits rather than on strategies to improve subsequent performance or learning (Fernando, Cleland, McKenzie, & Cassar, 2008), and that the emotive potential of a feedback interaction can inhibit productive and meaningful conversations that promote extension of learning (Carless, Salter, Yang, & Lam, 2011; Sargeant, Mann, Sinclair, Van Der Vleuten, & Metsemakers, 2008; Urquhart et al., 2014). Overly critical feedback may have damaging impacts on the quality of learning in the moment, as well as into the future (Henderson, Ferguson-Smith, & Johnson, 2005). In addition, models that ignore context, such as the 'feedback sandwich' where feedback givers ask formulaic questions focusing on positive elements of performance followed by constructive elements followed by positive comments, are often adopted as rules to be followed regardless of the situation. In the health professions, there is an additional layer of complexity; students work in real practice environments where feedback is often informal and verbal and is given by clinical supervisors as well as university-based academics. Observational studies of feedback in clinical education have demonstrated that educators or feedback providers can be so nervous about providing honest performance information to learners that they talk around the problem thus obfuscating the message (Molloy, Borello, & Epstein, 2013).

A recent definition in higher and professional education, called 'Feedback Mark 2', is 'Feedback is a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work' (Boud & Molloy, 2013, p. 205). Some of the distinctive properties of this definition include foregrounding of learner engagement, acknowledging that learners require standards literacy as a baseline in order to evaluate the quality of their own work, recognising feedback is an iterative process not a one-off information exchange and, finally, that feedback leading to action is a key ingredient. This comprehensive definition is a good starting point for our investigation, although it does not explicate the role of context in feedback interactions.

The health professions domain, where feedback interactions occur across multiple settings and people, presents an opportunity to explore the impact of context upon feedback. The lack of satisfaction with feedback practices from the perspective of learners, educators and policy makers (Carless et al., 2011; Gibbs & Simpson, 2004; Williams & Kane, 2009) suggests that there is a need to apply

new frameworks to the feedback question. One such framework, Bronfenbrenner's (1979) ecological systems theory, helps to integrate and conceptualise the environment and other influences on behaviour. In this chapter, we use an updated version (Neal & Neal, 2013) to unpack contextual influences on feedback practices and student learning in health professions education.

Bronfenbrenner's Ecological Systems Theory

Bronfenbrenner's (1977, 1979) ecological systems theory posits that individuals are influenced by interdependent systems at multiple levels. Originating in child development as a backlash to the scientific and experimental development psychology of the day, he argued that the natural ecological environment must be examined as an interdependent whole to fully understand the forces surrounding a developing individual. The developmental status of the individual is reflected in the substantive variety and structural complexity of the activities which he/she initiates (Bronfenbrenner, 1979). To understand human development, one must consider the entire ecological system in which growth occurs. The development of an individual is influenced by five environmental systems: microsystem, mesosystem, exosystem, macrosystem and chronosystem.

The original ecological systems theory considered multiple systems as nested concentric circles around a focal individual, therefore obscuring the important relationships between them. We will take a more contemporary view of these systems as 'networked' rather than nested as advocated by Neal and Neal (2013) lending greater theoretical clarity. In this conceptualisation, 'each system is defined in terms of the social relationships surrounding a focal individual, and where systems at different levels relate to one another in an overlapping but non-nested way' (Neal & Neal, 2013, p. 723). This approach promotes an exploration of social interactions (and patterns of social interactions) that comprise the different systems, each directly or indirectly connected to the others through direct and indirect social interactions of their participants (see Box 9.1 for further explanation of each system).

Box 9.1 The five networked environmental systems (Neal & Neal, 2013, p. 724) and exemplars

Microsystem: a set of people engaged in social interaction in one setting that includes the focal individual, for example, a learner engaging with informal feedback following a case presentation or work-based assessment of an observed task

Mesosystem: a social interaction between participants in different settings that all include the focal individual (i.e. the interrelations between

(continued)

Box 9.1 (continued)
microsystems), for example, a learner needing to shift his/her mode of seeking feedback as he/she moves between classroom, clinical and simulation-based microsystems

Exosystem: a set of people engaged in social interaction that does not include, but whose participants interact directly or indirectly with, the focal individual, for example, assessment policies in a university that dictate blinded feedback information provision to the learner on assignments or examination boards that set arbitrary feedback deadlines not in relation to sequencing of assignments

Macrosystem: the set of social patterns that govern the formation and dissolution of social interactions between individuals and thus the relationship among ecological systems, for example, professional feedback cultures within the health professions where mentors are also assessors

Chronosystem: the observation that patterns of social interactions between individuals change over time and that such changes impact the focal individual, both directly and by altering the configuration of ecological systems surrounding him/her, for example, developing more sophisticated feedback literacy in the transition from the preclinical to the clinical years.

Feedback and the Networked View of Ecological Systems Theory

Let us apply this approach to the development of a healthcare student, in particular with regard to feedback interactions. We take the case of Sarah our fictitious medical student and the influence of the various systems on feedback interactions and resultant effects. During her medical training, she will move (more or less) seamlessly between a number of microsystems each contributing to her learning including her personal home environment and the classroom, simulated and workplace learning settings. Sarah will engage in feedback interactions in each of these microsystems. She will also need to learn to negotiate feedback interactions between microsystems (i.e. mesosystem) and across broader macrosystems (e.g. institutional assessment policies) which Sarah might not have direct interaction with but which will have an influence on her development. Further she will negotiate exosystems of multiple cultures and subcultures, for example, in moving between different disciplines and wards. Finally we consider how her feedback interactions might change across her years of experience within the curriculum as a result of maturation and prior feedback experiences (i.e. chronosystem). Figure 9.1 presents an illustration of these systems.

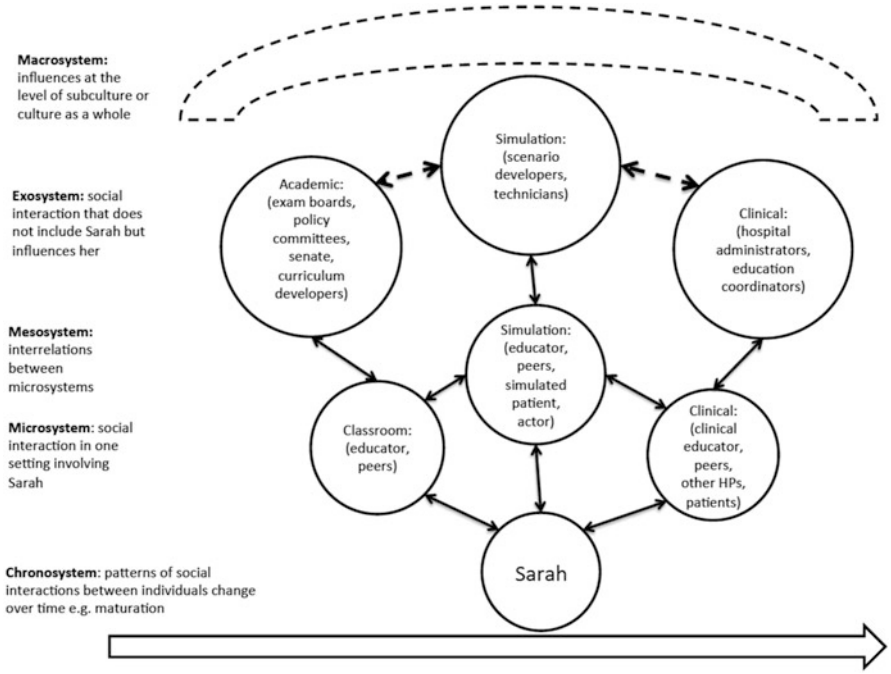

Fig. 9.1 Multiple networked systems that influence feedback interactions

Feedback Interactions Within the Microsystem

The microsystem is a 'pattern of activities, roles, and interpersonal relations experienced by the developing person in a given setting with particular physical and material characteristics' (Bronfenbrenner, 1979, p. 22). Sarah will experience her medical education in multiple settings, where she directly engages in feedback interactions with peers, patients (or care seekers), educators, assessors, tutors and administrators in classroom, simulated and clinical settings. In addition, she will have certain expectations and feedback experiences based on her life outside the medical school drawing from personal (e.g. work, hobbies, interactions with parents), work and prior educational experiences. Here we highlight findings from feedback research within the classroom, simulation-based and clinical microsystems in the health professions.

In the classroom/academic microsystem, feedback can be provided by peers and tutors, written or verbal, and is often included within formal summative assessment. Worryingly, medical students, like Sarah, generally understand feedback to be a one-way process of information giving rather than an active and collaborative process (Murdoch-Eaton & Sargeant, 2012; Urquhart et al., 2014). Sarah may come to view feedback as something 'done to her' and not 'with her'. This view of feedback as 'a destabilizing or debilitating act "done to them" by those in authority' (Molloy

& Boud, 2014, p. 422) goes against current recommendations that feedback be an active and collaborative process (Boud & Molloy, 2013) and certainly is not ideal as feedback should be deemed successful if shared understanding, learning or change in behaviour has been achieved.

In the clinical (workplace) microsystem, students learn through observing and participating in patient care. Students enter into the day-to-day work of healthcare environments, including hospital wards and general practices. The types of tasks undertaken are variable and depend both on the context and the students' capabilities and attitudes. For example, Sarah might follow a ward round, observing how patient care unfolds; or she may talk with a patient about their condition. Learning occurs through engagement with clinical supervisors, other health professionals, peers, patients and so on. Research on feedback interactions in the clinical microsystem has recently focused on the essential social and relational dimensions of feedback. Students make ongoing active judgements about the feedback source which influences their interpretations of, engagement with and future behaviours around feedback (Tai, Canny, Haines, & Molloy, 2015; Telio, Regehr, & Ajjawi, in press). Medical students make credibility judgements about their educators from the perspective of the educator's clinical credibility (Telio et al., in press). Others have described it in terms of the perceived beneficence of the feedback provider (Eva et al., 2012). Urquhart et al. (2014) additionally highlight how personal characteristics of the feedback provider (e.g. perception of authenticity) influence credibility judgements.

One way of conceptualising credibility judgements between learners and educators is through the lens of the educational alliance (Telio, Ajjawi, & Regehr, 2015). The educational alliance is derived from the concept of the 'therapeutic alliance' as evolved in psychotherapy (Telio et al., 2015). The quality of this alliance has been shown repeatedly to be the most robust predictor of therapy outcome, surpassing the impact of specific therapeutic techniques. In the same way that a patient can form a therapeutic alliance with the therapist, so a learner may be thought to form an educational alliance with their educator. The educational alliance is composed of:

1. The learner's belief that there is a mutual understanding of the purpose or goal of the relationship
2. The learner's belief that there is an agreement about how to work towards that goal and the activities involved
3. The learner's credibility judgements of the educator including liking, trusting, and valuing of the educator and belief that these feelings are mutual (Telio et al., 2015)

It is therefore Sarah's judgement about the quality of the educational alliance that matters here. Telio et al. (in press) found that feedback incorporation and the valence of emotion were related to the strength of the educational alliance rather than the direction of feedback. Indeed it is in the context of strong alliances that one can engage in 'negative' feedback with effective impact because this difficult feedback is likely to be received with the understanding that it is to help the learner improve rather than as an attack on or denigration of the individual. It is also in the context

of stronger alliances that learners are more likely to seek external feedback and to engage in open and constructive feedback encounters, which are the necessary conditions for the development of evaluative judgement. The educational alliance may help to reframe understandings of feedback from rules about content and delivery to a more nuanced appreciation of the role of relationships and feedback interactions in learning within the microsystem.

The simulation learning environment or microsystem can be thought of as a bridge between classroom and clinical environments, as this is where learners rehearse the practices required of them as professionals. Within this broad notion of a simulation microsystem, we include different simulation methodologies ranging from psychomotor skills development (e.g. learning to suture on foam pads) to immersive acute simulation (e.g. fully body mannequins) and communication skills training (e.g. working with simulated patients or actors). One of the contrasting features of the simulation microsystem compared to the clinical education microsystem is the deficit of real patients and real responsibility and the (rich) dynamics of a real clinical environment. In simulation, student learning is the primary focus of the activity, rather than patient care. This means that feedback time can be scheduled and prioritised. In general, simulation offers a relatively feedback-rich experience, although there are obviously variations across simulation methodologies and particular programmes. It is also important to note that just because there are many opportunities for feedback, it is not necessarily *effective* feedback. There is some indication that, as in the other microsystems, the one-way flow of information from educator to student persists (Dieckmann, Molin Friis, Lippert, & Østergaard, 2009).

The notion of credibility judgement necessarily shifts in the simulation microsystem. For example, in a usual patient-learner encounter, the learner can be considered to be positioned as the powerful presence in the duo; in a simulated patient-learner encounter, the simulated patient may be positioned as more powerful, particularly if they are providing a judgement about the learner's progress (Hanna & Fins, 2006). Furthermore, as feedback in the simulation setting may be provided by non-medical practitioners, such as nurses, who are no longer working in the clinical environment, this may influence students' credibility judgements of the feedback. This notion of credibility and how it transfers from simulated to clinical environments is particularly thrown into relief when considering the mesosystem.

Feedback Interactions Within the Mesosystem

The mesosystem is constituted in the interactions between intersecting microsystems (Bronfenbrenner, 1979). As described in the previous section, Sarah will learn to engage with and negotiate feedback interactions across multiple microsystems: classroom, simulation-based and workplace-based settings and various sub-settings within those settings (e.g. primary and secondary care workplace settings). She will learn that there are different feedback expectations and practices embedded within

each of those different settings. Disconnects between microsystems through mixed messages, lack of alignment and the hidden curriculum have significant implications for Sarah's development as a doctor. For example, we know that students learn to expect and demand feedback interactions within the simulation environment but that they feel a burden on their busy clinical educators who are first and foremost clinicians caring for patients within the clinical environment (Urquhart et al., 2014). This means that Sarah might appear to be actively seeking feedback in one microsystem but may be reluctant and passive in another, thus negatively influencing her learning opportunities.

The potential variations in feedback practices between clinical and simulated environments within medical schools have been described (Urquhart, Rees, & Ker, 2015). A video-reflexive ethnography study conducted at one UK medical school, for example, found that feedback practices differed between clinical and simulated environments in terms of who the feedback providers were, what feedback was given in terms of content and style and when and where feedback was given (Urquhart et al., 2015). The authors found that learners' and tutors' perceptions of feedback depended on their perceptions about the primary purpose of the contexts in which students received feedback, that is, patient care (clinical context) versus student learning (simulated context) (Urquhart et al., 2015).

The movement between microsystems can be challenging, not just in terms of what feedback is given but how feedback may be applied. Yardley, Irvine and Lefroy (2013, p. 506) describe how 'the student rejects learning constructed from simulation that appears to conflict with the practice he or she observes in authentic workplaces'. In their subsequent discussion, they propose that educators have to learn to highlight, manage and be mindful of 'the gap' between simulated and real environments. This has implications for Sarah and her ability to engage in feedback across all three microsystems.

Feedback Interactions Within the Exosystem

The exosystem refers to 'one or more settings that do not involve the developing person as an active participant, but in which events occur that affect, or are affected by, what happens in the setting containing the developing person' (Bronfenbrenner, 1979, p. 25). In the case of our medical student, there are various individuals within the university and clinical settings who Sarah might not interact with personally but who influence her development. For example, the academic exosystem would include curriculum developers, assessment designers, high-level policy committees, examination boards and so on. The clinical exosystem includes hospital administrators, deans of education and high-level policy committees. Another important exosystem in Sarah's growth is the regulatory body which sets standards for practice and writes the language around these competency frameworks. This may become the language of feedback interactions, as Sarah learns what professional standards, values and qualities are expected of her, the neophyte doctor.

Many decisions occur within the academic exosystem that impact Sarah's learning through feedback. Issues of curriculum design, feedback loops and opportunities to incorporate feedback into learning are important considerations that take place in the exosystem and are considerations that are frequently overlooked in the feedback discourse (Molloy & Boud, 2013). The dominant understanding of feedback in higher education is that it constitutes a teacher providing comments to a learner in relation to a task (e.g. a workplace procedure/task or an assignment). This limited view of feedback is challenged by looking to the engineering origins of the term where feedback necessarily requires action or change to occur (Boud & Molloy, 2013). To use an engineering example, a thermostat responds to a drop in temperature by generating heat to bring the room to a set and desired temperature. A thermostat flashing 'too cold' on the register screen is an example of information display, not a feedback process. It is the response of the system to the information that closes the loop and which meets the definition of feedback. Hence, how the curriculum is designed to promote further opportunities for Sarah to apply feedback to related tasks is crucial to her development.

Another example of the academic exosystem influencing student behaviour is through assessment policy and exam standard setting decisions that may seem arbitrary to Sarah but can have significant implications on her making it through the course. Furthermore, feedback role modelling by medical schools is an important consideration of the exosystem. We know from research that how a medical school role models feedback, through how student evaluation data is acted upon, influences learners' receptiveness to the feedback they receive (Urquhart et al., 2014). For example, if students see no action to improve teaching on the basis of their feedback, they may become cynical of the real value of feedback (Urquhart et al., 2014).

Feedback Interactions Within the Macrosystem

The macrosystem can be considered 'at the level of subculture or culture as a whole, along with any belief systems or ideology underlying such consistencies' (Bronfenbrenner, 1979, p. 26). There have only been a handful of studies in medical education that have aimed to explore the macrosystem in which feedback occurs. Watling and colleagues (Watling, Driessen, van der Vleuten, & Lingard, 2014; Watling, Driessen, van der Vleuten, Vanstone, & Lingard, 2013a, 2013b) explored and compared feedback experiences in three distinct learning cultures – medicine, education and music. Each of these cultures shaped learners' expectations of feedback in particular ways. Whilst music and education students expected constant observation and feedback, medical students felt a burden on their teachers who had to juggle patient care and student education and often received feedback on unobserved performance. The study highlighted how credible and constructive feedback is valued across all learning cultures but how that credibility and constructiveness are defined is culturally determined (Watling et al., 2013b). In both music and education, the prevalence of observation, feedforward and action plans acted to

improve the credibility of feedback. These were almost absent practices in medicine which compromised the value of feedback in the eyes of the learners.

In another study Watling et al. (2014) contrasted experiences of feedback by doctors who had expertise in music or sport. Participants explained the indispensable role provided by music teachers and sports coaches, yet medical teachers were described as role models who provide 'examples of desired performance rather than motivation and continuous guidance' (Watling et al., 2014, p. 717). This stemmed from recognition that the primary job of a clinical teacher is to treat patients, whereas in music and sports, teaching is a dedicated role. Trusting long-term teacher-learner relationships were much more readily identified within music and sport than in medicine. Worryingly doctors felt that although feedback was crucial to their development as musician and sportspeople, feedback was less central to their development as doctors. The elements described by Watling et al. (2014) as valued in sports and music yet missing from medicine echo the dimensions described by Telio et al. (2015) regarding the educational alliance. As Sarah will experience, the educational alliance between teacher and learner is fragmented within the medical macrosystem to the detriment of learning from feedback and the learning experience.

Feedback Interactions Across the Chronosystem

The chronosystem is the observation that patterns of social interactions between individuals change over time, and that such changes impact the focal individual, both directly and by altering the configuration of ecological systems surrounding him/her (Neal & Neal, 2013). In her journey to becoming a safe and competent doctor, Sarah will experience several key educational transitions. Factors that will influence Sarah's feedback interactions across these transitions include prior experiences with feedback and developments in her self-regulation capacities.

One study has highlighted maturational differences between junior and senior medical students' conceptualisations of feedback with senior students adopting more sophisticated understandings of the role of feedback in their learning, shifting to more active (rather than passive) utilisation and valuing informal and verbal feedback from senior clinicians (Murdoch-Eaton & Sargeant, 2012). This highlights a shift in feedback literacy as students experienced and engaged with the curriculum and feedback interactions resulting in adjusting their expectations of feedback and their role in it. Senior students were generally more aware of the important role of feedback in their learning, their need to adopt a more active stance in seeking and incorporating feedback into a longer-term change in learning approach (Murdoch-Eaton & Sargeant, 2012).

Returning to the educational alliance, there are further implications of the influence of the chronosystem on learners such as Sarah. Early findings suggest that evaluations about the strength of the educational alliance not only affect a learner's engagement with a particular piece of feedback at the moment of delivery but also

have consequences for future engagement in (or avoidance of) further learning interactions with the supervisor (Telio et al., in press). There is early indication that such conditions can be generated even within brief encounters if educators are willing to invest in discussions around feedback expectations, co-construction of goals and embedding of feedback loops (Farrell, Bourgeois-Law, Ajjawi, & Regehr, 2016).

Another important factor to consider in relation to the chronosystem is the emotional legacy that students are left with as a result of feedback interactions during medical school. Urquhart et al. (2014) in their narrative study of feedback in the workplace highlighted how students positioned themselves as passive recipients (or victims) within their feedback narratives, with their feedback providers constructed as villains utilising 'us and them' language. They demonstrated the real emotional toll of feedback practices and the prevalence of negative experiences including verbally abusive and humiliating feedback comments and adversarial relations between students and tutors (Urquhart et al., 2014).

Scaling Up: What Are the Implications of the Ecological Model?

Scaling up has been conceptualised in relation to four interrelated dimensions: spread, depth, sustainability and shifts in ownership (Coburn, 2003). We believe that a significant step forward in terms of scaling up effective feedback practices is through improving feedback literacy of students and staff. This relates to notions of depth and sustainability, which can be promoted through considered 'feedback by design' practices and through shifting the onus of responsibility towards students who are better at navigating the feedback landscape (as judge, seeker and user). Often interventions to improve feedback practices are unilateral, typically focusing on teacher behaviours, feedback content or feedback delivery within a single microsystem. This ecological view could explain why such a landscape is resistant to change and why feedback interventions can (and often do) fail (Ferrell, 2012). It also highlights the challenges involved in changing feedback practices at scale (see introductory chapter in this book). The learner moves through a range of different systems with different feedback practices, which on the one hand lack cohesion but on the other provide a vast range of different and important opportunities and experiences. How might we better prepare students to navigate these systems in efficient and informed ways that enable effective feedback interactions? How might academic staff design feedback interactions to establish conditions in which students can operate with agency? How might the enabling conditions of context be harnessed to promote the positive effects of feedback in sustainable ways?

The implication of this brief examination of networked systems for the scaling up of effective feedback practices is that a one-size intervention is not suitable for all. Promoting feedback by design involves taking into account the multiplicity of

factors for each of the interdependent systems. We may choose to intervene early by improving feedback literacy of students in the first year of professional programmes. This potentially sets up the student to understand the ecological landscape of their professional formation so that future encounters in the curriculum build on realistic expectations and healthy feedback practices (e.g. seeking feedback, active self-evaluation and mindful development of evaluative judgement) when the tasks get increasingly complex. Perhaps if learners are socialised into this feedback landscape early, their future roles as feedback users and providers might look different.

Changing beliefs and practices of staff through improving feedback literacy would require teachers to work differently. There are particular interactional considerations such as establishing trust in the educational alliance, explicitly agreeing on the purposes of feedback and goals of the interaction and structuring dialogue and linguistic and non-verbal choices in the feedback episode to actively include the learner (Farrell et al., 2016). Collaborative models, such as Feedback Mark 2, place less emphasis on telling and more focus on designing of experiences across a programme of study and, in particular, nested tasks that give learners a chance to respond to previous performance information exchanges and put new strategies into practice. It is understandable that educators should wish to focus on the immediate microsystem within faculty development workshops. Yet it would be beneficial for them to also consider the students' journey through the broader landscape. Feedback on concepts learned in the classroom may be applied within a simulated environment and feedback given on a simulated performance may be applied within a 'real' clinical environment. Feedback givers can specifically highlight the challenges which may be experienced in the movement from working with a paper problem to working with a simulated patient or mannequin to providing supervised care to a real patient but also to consider alignment and graded complexity in the design of tasks across these microsystems. Effective feedback cultures may be promoted through the engagement of higher education leaders and policy developers in examining the effects of their policy and infrastructure decisions on feedback cultures, learners and learning. Another strategy might be in facilitating different stakeholders (from the different systems) to come together to collaborate in seeking understanding of synergies and tensions across the networked systems and to use this understanding to inform change strategies. It behoves all those invested in the development of students to consider the emotional legacy of assessment and feedback interactions on learners and their developing professional identity.

Gaps and Recommendations for Research

Based on this conceptual framing of feedback and the ecological systems theory, we have highlighted gaps in the literature from which we draw some recommendations for future research. Whilst much of the research on feedback has occurred within the *microsystem*, opportunities exist in understanding the value of the educational alliance to the broader spectrum of medical education. The applicability of this

concept to the undergraduate arena and to other health professions (and indeed beyond the health professions) is unknown. An interesting line of inquiry would be to identify the types of credibility judgements that different health professions learners make and how these influence future feedback behaviours (seeking, utilising and designing feedback). In addition, exploring the conditions that strengthen the educational alliance would be profitable to pursue. Research is also needed to better understand how students learn to navigate the *mesosystem* and how they calibrate their expectations of feedback flexibly within and across different microsystems. Exploring synergies, tensions and contradictions in feedback practices between microsystems (i.e. identifying the hidden curriculum of feedback in the mesosystem) and how this may be used to improve feedback literacy would be valuable.

Further research on collaborative models of feedback and implementation on a large scale is needed to identify key design features that promote learning beyond the immediate task (exosystem). Effecting culture change within a macrosystem is not easy, and research shows that feedback cultures within medical education can act as a hindrance. Interdisciplinary work is needed to better understand the effects of feedback cultures on learners and to dismantle some of the structures that act to fragment feedback practices. Within the chronosystem it is not clear if improvements in feedback literacy (and resultant improvements in self-regulation) could be achieved through explicitly educating students about feedback and their role in seeking and using it, early in a curriculum. This could be one area of future research. How trust evolves over time, the establishment of strong educational alliances and the influence of multiple feedback sources (patient, educator, peer) on building pictures of learner performance in complex systems are other areas for future research.

Conclusion

We have highlighted how feedback interactions occur through our student Sarah's journey through multiple networked systems. Promoting feedback by design involves taking account of the contextual factors relevant to each system. As we have explored, this may be at the microsystem (e.g. reflecting on the educational alliance, establishing trust), at the mesosystem (e.g. setting up expectations for effective feedback behaviours for students and staff to navigate across microsystems), at the exosystem (e.g. designing curricula), at the macrosystem (e.g. critically examining feedback cultures) and at the chronosystem (e.g. explicitly promoting feedback literacy aligned with key transitions). The key message here is that feedback is influenced by individual, interpersonal, social, contextual and cultural factors. Educational interventions that only take into account the individual are bound to be less effective and may explain the wave of feedback dissatisfaction in the higher education literature. On the basis of mapping the ecological systems theory with feedback practices, we highlight the potential usefulness of ecological models for research and scaling up practice in assessment for learning in higher education.

References

Barton, K. L., Schofield, S. J., McAleer, S., & Ajjawi, R. (2016). Translating evidence-based guidelines to improve feedback practices: The interACT case study. *BMC Medical Education, 16*(1), 1–12.

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

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*, 513–531.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12.

Dieckmann, P., Molin Friis, S., Lippert, A., & Østergaard, D. (2009). The art and science of debriefing in simulation: Ideal and practice. *Medical Teacher, 31*(7), 287–294.

Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., et al. (2012). Factors influencing responsiveness to feedback: On the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education, 17*(1), 15–26.

Eva, K. W., & Regehr, G. (2005). Self-Assessment in the health professions: A reformulation and research agenda. *Academic Medicine, 80*(Suppl), S46–S54.

Farrell, L., Bourgeois-Law, G., Ajjawi, R., & Regehr, G. (2016). An autoethnographic exploration of the use of goal oriented feedback to enhance brief clinical teaching encounters. *Advances in Health Sciences Education*, 1–14.

Fernando, N., Cleland, J., McKenzie, H., & Cassar, K. (2008). Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. *Medical Education, 42*(1), 89–95.

Ferrell, G. (2012). *A view of the assessment and feedback landscape: Baseline analysis of policy and practice from the JISC Assessment & Feedback programme*. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/elearning/Assessment/JISCAFBaselineReportMay2012.pdf

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Hanna, M., & Fins, J. (2006). Power and communication: Why simulation training ought to be complemented by experiential and humanist learning. *Academic Medicine, 81*(3), 265–270.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Henderson, P., Ferguson-Smith, A. C., & Johnson, M. H. (2005). Developing essential professional skills: A framework for teaching and learning about feedback. *BMC Medical Education, 5*(1), 1–6.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

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.

Molloy, E. K. (2009). Time to pause: Feedback in clinical education. In C. Delaney & E. K. Molloy (Eds.), *Clinical education in the health professions*. Sydney, Australia: Elsevier.

Molloy, E. K., & Boud, D. (2013). Seeking a different angle on feedback in clinical education: The learner as seeker, judge and user of performance information. *Medical Education, 47*(3), 227–229.

Molloy, E. K., & Boud, D. (2014). Feedback models for learning, teaching and performance. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 413–424). New York: Springer.

Molloy, E. K., Borello, F., & Epstein, R. (2013). The impact of emotion in feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher education* (pp. 50–71). London: Routledge.

Murdoch-Eaton, D., & Sargeant, J. (2012). Maturational differences in undergraduate medical students' perceptions about feedback. *Medical Education, 46*(7), 711–721.

Neal, J. W., & Neal, Z. P. (2013). Nested or networked? Future directions for ecological systems theory. *Social Development, 22*(4), 722–737.

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

Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., et al. (2010). The processes and dimensions of informed self-assessment: A conceptual model. *Academic Medicine, 85*(7), 1212–1220.

Sargeant, J., Mann, K., Sinclair, D., van der Vleuten, C., & Metsemakers, J. (2008). Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education, 13*(3), 275–288.

Tai, J. H.-M., Canny, B. J., Haines, T. P., & Molloy, E. K. (2015). The role of peer-assisted learning in building evaluative judgement: Opportunities in clinical medical education. *Advances in Health Sciences Education.* doi:10.1007/s10459-015-9659-0.

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.

Telio, S., Regehr, G., & Ajjawi, R. (in press). Feedback and the educational alliance: Examining credibility judgments and their consequences. *Medical Education*.

Urquhart, L. M., Rees, C. E., & Ker, J. S. (2014). Making sense of feedback experiences: A multi-school study of medical students' narratives. *Medical Education, 48*(2), 189–203.

Urquhart, L. M., Rees, C. E., & Ker, J. S. (2015). *Exploring feedback in context at medical school through video-reflexive ethnography*. Paper presented at the Association for Medical Education in Europe, Scottish Exhibition and Conference Centre, Glasgow, Scotland. Paper retrieved from https://www.amee.org/getattachment/amee-news/AMEE-2015-Abstract-Book/Final-Abstract-Book-as-at-3-SEptember-2015.pdf

Watling, C., Driessen, E., van der Vleuten, C. P. M., & Lingard, L. (2014). Learning culture and feedback: An international study of medical athletes and musicians. *Medical Education, 48*(7), 713–723.

Watling, C., Driessen, E., van der Vleuten, C. P. M., Vanstone, M., & Lingard, L. (2013a). Beyond individualism: Professional culture and its influence on feedback. *Medical Education, 47*(6), 585–594.

Watling, C., Driessen, E., van der Vleuten, C. P. M., Vanstone, M., & Lingard, L. (2013b). Music lessons: Revealing medicine's learning culture through a comparison with that of music. *Medical Education, 47*(8), 842–850.

Williams, J., & Kane, D. (2009). Assessment and feedback: Institutional experiences of student feedback, 1996–2007. *Higher Education Quarterly, 63*(3), 264–286.

Yardley, S., Irvine, A. W., & Lefroy, J. (2013). Minding the gap between communication skills simulation and authentic experience. *Medical Education, 47*(5), 495–510.

Chapter 10
Student Utilisation of Feedback: A Cyclical Model

Edd Pitt

Abstract In this chapter I discuss research which explored student's experiences of assessment and feedback from a phenomenographic perspective and propose a conceptual six-stage cyclical assessment and feedback model. Data indicated a multifaceted interpretation of the student experience suggesting that a student's achievement outcome, relative to their predetermined expectation level, regulates their emotional reaction and subsequent feedback utilisation. The phenomenographic outcome space (Åkerlind GS, High Educ Res Dev 24(4):321–334, 2005) revealed five categories of description (rationalising, needy, low achiever, emotionally changeable and high achiever). The chapter explores the implications of this outcome space in relation to how students utilised the feedback they received. In light of such findings, implications for practice are discussed indicating that grade outcome was an extremely powerful construct which seemed to foster both adaptive and maladaptive emotions and subsequent assessment-related behaviours. In conclusion, I suggest that understanding students' individual needs through fostering lecturer and student relationships, alongside dialogic feedback opportunities, may help to improve a student's propensity to utilise the feedback received.

Introduction and Literature Review

The literature relating to feedback has seen many shifts in supported conceptual and theoretical understanding in recent years. In particular, there are current debates relating to the exact purpose of feedback. Quite often feedback is facilitated by academic lecturers via a monologic transmission process, in the hope that the student will utilise this and improve in their next assessment (Handley, Price, & Millar, 2011; O'Donovan, Rust, & Price, 2015). Students' emotional response, motivation, self-confidence and subsequent effort deployment in future assessments following feedback are unpredictable and warrant further consideration. The research I report on in this chapter explored how social science students appraise, comprehend and

E. Pitt (✉)
University of Kent, Canterbury, UK
e-mail: E.Pitt@kent.ac.uk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_10

subsequently utilise feedback received from lecturers during their undergraduate degree. In particular the focus was upon students' emotional processing following both a good or poor grade and how this affected their use of feedback. A number of studies have suggested that the feedback students are receiving is doing little to improve their learning (Bailey & Garner, 2010; O'Donovan et al., 2015; Sadler, 2010). Research needs to question the established mechanism of simply giving students' feedback and expecting them to attend and adjust (Handley et al., 2011). This seems to be particularly prudent if one appreciates Beard, Clegg and Smith's (2007) contentions that in the first year of university, students experience an emotional roller coaster which transcends many aspects of their lives. Emotional engagement is of interest to university lecturers, if we appreciate that emotions could last for a sustained period of time and have a long-term effect upon students' learning. Such emotions can either obstruct or stimulate learning by affecting attention deployment, memory and problem-solving performance (Boud & Falchikov, 2007). Encouraging comments reinforce positive reactions as well as moderate the effect of negative comments (Fredrickson & Cohn, 2008; Lizzio, Wilson, Gilchrist, & Gallois, 2003). Further, positive emotions seem to encourage self-regulation and learning strategies aligned to deep learning (Fredrickson & Cohn, 2008; Pekrun, Goetz, & Titz, 2002). However, negative emotions seem to foster an over-reliance upon the lecturer and surface learning (Pekrun et al., 2002). Recent research has even suggested that negative emotions, such as fear, can reduce students' already low self-efficacy causing academic paralysis (Nash, Crimmins, & Oprescu, 2016).

The literature has also considered students' achievement expectations suggesting that grade outcome expectations promote pride or disappointment (Kahu, Stephens, Leach, & Zepke, 2015). When receiving good grades, students do not read or attend to the feedback if they feel satisfied with that grade (Lipnevich & Smith, 2009; Vardi, 2009). The debilitating effect negative feedback has upon students' behavioural response has been highlighted as a cause for concern for lecturers (Vardi, 2009). Pitt and Norton (2016) indicate that when receiving negative feedback, and/or low grades, students can react either positively or negatively. High-achieving students demonstrated strong self-assessment skills alongside an ability to distinguish their current level of learning from the feedback received and how this may be used in future assessments. Conversely, lower-achieving students struggled to understand feedback language and carry out self-assessment and were therefore unable to regulate their learning (Orsmond & Merry, 2013; Pitt & Norton, 2016). In this regard Beard, Humberstone and Clayton (2014) have called for researchers to view students as affective and embodied individuals, concluding that in order to understand this phenomenon, clearer theorisation of students' emotional experiences is needed. In this chapter, I explore how students' pre-assessment dispositions and grade expectations affect their emotional reactions, behavioural actions and feedback utilisation in subsequent assessment opportunities.

Methodology

The methodology used in this research was phenomenography. A structural framework is provided within phenomenography to understand others' experience from their perspective. Such a method allowed me to report the variation and internal relations of the collective students rather than the individual students' experiences of assessment and feedback (Åkerlind, 2005; Pitt, 2014; Trigwell, 2006).

Participant Description

Twenty final year undergraduate students studying in the Science and Social Science Faculty at a university in the north-west of England took part in a student-generated drawing and subsequent one-to-one interview relating to their assessment and feedback experiences in higher education. The procedure lasted 90 min. An important consideration in this research was involving students across differing levels of achievement. Students at the higher end of the grade point average (i.e. traditionally keen to be involved in research projects and well-represented in the literature) but also those at the lower end (i.e. traditionally not so keen to be involved in research projects and under-represented in the literature) were selected for the research. The breakdown of gender reflected male ($n = 9$, average age 22.66 years) and female ($n = 11$, average age 21.66 years). Ethical procedures were followed rigorously within this study and clearance gained from the university.

Data Gathering

A concern in this study centred upon asking students to articulate their emotional experiences at times that I perceived were stressful or upsetting. Student-generated drawings allowed students time to really understand and formalise their responses and articulate and emphasise implicit emotional and relational aspects surrounding assessment and feedback that may have been missed out had straightforward pressurised one-to-one interviews been used alone (see Pitt, 2014 and 2016 for a detailed review of this methodology). Students were given large pieces of A2 flip chart paper and a selection of coloured marker pens. The drawing process began with three warm-up tasks to allow students to move into a cognitive process of thinking in visual terms (Stiles, 2004). In the main drawing task, the students visually represented their thoughts and feelings regarding their assessment and feedback experiences to date in higher education. In order to avoid misinterpretation of the student-generated drawing, I asked the students to describe to me what their drawings represented in a one-to-one interview. As such the flow and content of the interviews were entirely dependent upon the student-generated drawing.

Data Analysis

The interviews yielded a significant number of statements which were phenomenographically analysed. Through an iterative process, meanings emerged from the data related to the context of the collective group of students and were subsequently categorised (Åkerlind, 2005, 2008). The analysis in this chapter is referred to as the 'outcome space' where common themes in the meanings expressed by the students are explained (Åkerlind, 2008). It is important to stress here that the categories do not reflect individual students, rather composites of many utterances statements from within the data which I have constructed into categories of description.

Results

The assessment and feedback cycle is an analytical frame through which I explore differences between the students interviewed in this study. The cycle proposes that the students' experiences of assessment and feedback impact upon how they use the feedback received in future assessments. Figure 10.1 (Pitt, 2014) visually depicts the cyclical nature which begins at stage one and continues through to stage six whereby the process restarts at the next assessment opportunity:

The students' predispositions start the assessment and feedback cycle. The predispositions are not personality traits or characteristics; rather they reflect student's current cognitions in relation to the forthcoming assessment task (Pitt & Norton, 2016). The students' disposition could have been informed by previous

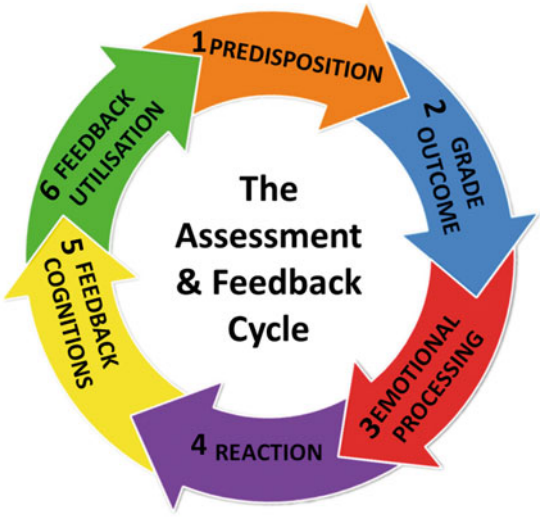

Fig. 10.1 Assessment and feedback cycle (Pitt, 2014)

assessment and feedback situations within higher education or from school or college (Beaumont, O'Doherty, & Shannon, 2011). The lecturer is presented students who hold certain beliefs about assessment and feedback. The decisive factor I have used here is the distinction the students made about the quality of their work, that is to say students reflected upon times when they perceived their work to be either good or poor. The interviews also revealed that all students held a predetermined expected level of achievement prior to commencement of each assessment (see Pitt & Norton, 2016 for a detailed explanation). The literature reports the value that students place upon grade outcome (Carless, 2006; Orsmond, Merry, & Reiling, 2005). The data in this study reflected periods when the students discussed both their good grades and poor grades (further explored later in this chapter). Following the grade outcome in the cycle, the students' immediate emotional reactions are reported; subsequent to these, non-emotional reactions follow. The final two stages are where the student experiences cognitions related to the feedback (where processing of the feedback occurs) and utilisation of the feedback in the next assessment, which completes the cycle. It is here that the students' assessment and feedback behaviour could have been potentially modified by the feedback they had received and the cycle begins again for the next assessment situation. The assessment and feedback cycle is illuminated in the following section of this chapter where I detail the outcome space which contains the categories of description.

It is important at this junction that I highlight the fact that lecturers' feedback is designed to initiate change within the student. However, the literature has demonstrated this is as a rather simplistic view (see Nicol, 2010). In this present research, I sought to further understand how multifaceted constructs such as emotional processing, grade outcome, motivation and self-regulation interact to affect the student's utilisation of feedback in the next assessment opportunity. The assessment and feedback cycle used as an analytical frame alongside the outcome space I have proposed here in this chapter attempts to demonstrate the adaptive and maladaptive processes that students in this study experienced. As such I propose that subsequent assessment situations are informed by previous experiences and those predispositions could change and therefore positively or negatively affect the subsequent student performance. The outcome space demonstrates the qualitatively different ways that the students in this study experienced the phenomenon. The variation in meaning the students in this study attached to their experiences of assessment and feedback is also highlighted by this outcome space.

Categories of Description

I propose five categories of description within the outcome space. The variation within this sample is reflective of the variation of meaning that one could conceivably expect within other similar population groups, although I do acknowledge that it is somewhat limited to one group of social science students in one institution.

The five categories of description representing key elements of variation in meanings and experience which were constructed from the data are:

1. Rationalising
2. Needy
3. Low achiever
4. Emotionally changeable
5. High achiever

These categories are now discussed in terms of their predispositions, grade outcome, emotional reactions, feedback cognitions and feedback utilisation in the next assessment. I make a distinction within this discussion in relation to the students' grade outcome in order to highlight the adaptive and maladaptive effect grades had upon the interviewed students in this study.

Rationalising

The students in this study categorised as rationalising showed more introverted characteristics and avoided interpersonal grade comparisons, attributing their performance outcome to the lecturer but still sought feedback on draft work from them. They had an aspirational level of achievement but hardly ever achieved this, rationalising their high effort deployment should have resulted in a successful grade outcome. Receiving a poor grade negatively affected their emotions and motivation for a considerable length of time and negative thoughts dominated their cognition both inside and outside university. Performance outcomes were attributed to their lecturer as they did not help them enough through draft work and teaching situations or had not marked fairly. Good grades were achieved occasionally, initiating emotions such as elation and pride which lasted for a sustained period as they felt they had proved the lecturer wrong. Aspirations to maintain this grade outcome lead to an increase in self-confidence and motivation going into the next assessment. The students categorised in this way in the analysis struggled to understand feedback language, believing that it did not always match the grade awarded and resulted in their relationship with the lecturer breaking down. Regardless of the grade awarded, the feedback was not very useable in the next assessment, as it related to effort deployment rather than content improvement. They did not value their lecturers' perceptions or judgements and decided not to utilise the feedback in the next assessment.

Needy

In this analytical category, the students discussed low ability conceptions and low self-confidence. They expected feedback on multiple drafts and support from the lecturer. Regardless of the direction of effort, they expected to do well in

assessments. They avoided interpersonal grade comparisons as these made them feel inferior. Receiving a poor grade, for the students in this study categorised as needy, resulted in emotional reactions of feeling demoralised and annoyed, leading to questioning their ability to succeed at university. They attributed the poor grade to their lecturer, citing poor teaching, and as such motivation to engage with their studies and feedback in the next assessment opportunity was diminished significantly. The poor grade seemed to act as a barrier to processing the feedback as they were emotionally so demoralised and annoyed immediately at the time of receiving it. Despite their emotional reaction, they were able to articulate that they wanted diagnostic feedback, which informed them what they had done wrong and what they needed to do to correct it. They viewed counterproductive feedback as reflecting effort expenditure judgements made by the lecturer or overly negative feedback. They also believed that feedback in one assessment could not be applied or even transferred to any other assessments. Receiving a good grade for students in this study categorised as needy lead led to emotions of happiness alongside increased motivation, self-confidence and positive perceptions about achieving higher in the next assessment. Their first pragmatic reaction (aside from the emotional reaction) was that the feedback was not needed as it just confirmed they had done well and did not include anything which would help them in the next assessment. They understood the positively worded feedback but they only looked at it once believing it wouldn't help them next time. They were happy with the grade and read the feedback but did not use it as they did not see the transferability to their next assessment. Last two sentences are slightly repetitive.

Low Achiever

In this analytical category, the students had low self-confidence, had ability conceptions and questioned if they were cut out for university. They did not seek feedback on draft work; rather they strategically asked assessment-related questions in lectures, circumventing one-to-one situations to avoid looking stupid. They had low grade expectations when they knew they had not put in a lot of effort. Receiving a poor grade, for the students in this study categorised as low achievers, initiated sustained emotions of anger and frustration which they attempted to mask in front of peers and the lecturer. They understood the marking process and feedback language, appreciating why the grade was given. However, they were unable to accept criticism so the negative feedback was unusable. Consequently, they did not seek lecturer assistance as this would have revealed a weakness to the lecturer. Constructive feedback was preferred, although they did not experience this very frequently, so the feedback was not used in the next assessment. Receiving a good grade for the students categorised as low achiever fostered sustained euphoric emotions, increased confidence, motivation and positive intrapersonal feelings. They viewed the good grade as exceeding normal achievement level, believing it did not reflect actual ability; rather it was a lucky occurrence. The positive grade outcome, alongside euphoric emotions, initiated engagement with the feedback but

the contents were almost immediately forgotten. The lecturer was used to clarify the feedback as they were proud of the outcome and interpreted this as the lecturer holding a positive perception of them. They saw feedback as isolated to only this piece of work and not transferable to other modules so it is not utilised in the next assessments.

Emotionally Changeable

The students categorised as emotionally changeable in the analysis were governed by their emotions. They were engaged in their studies, believed their ability was fixed and had fragile confidence, and their expected achievement level was modest. They sought feedback on draft work but negative feedback reduced their self-confidence. Overwhelmingly they did not want to let anyone down especially not the lecturer. For the students categorised as emotionally changeable, receiving a poor grade could initiate emotions of anger, disappointment and periodic tearfulness. Negative ability perceptions led them to question their future involvement in the degree. The strength of their emotions made it difficult for them to engage with the feedback, and they attributed the outcome to the lecturer which initiated feelings that the marking may have been unjust. After the emotional reaction had passed, they tried to forget about the outcome, increasing their level of motivation, and endeavoured towards improving their grade outcome in the next assessment. When the emotional reaction had passed, they returned to the feedback and attempted to engage with it; in particular, written comments were the most useful for them in the next assessment. Receiving a good grade, for the students in this study categorised as emotionally changeable, resulted in emotions of happiness, increased motivation and self-confidence for a sustained period of time. They were keen not to let their lecturer down as they felt that their good grade would make the lecturer proud. Receiving a higher grade dominated their cognition for a long period of time as they were so pleased with this achievement. The grade outcome was the most important factor; they liked to receive written feedback as this was the most usable in the next assessment. Group-based feedback was also used as they liked to work through it with their peers in order to use it in the next assessment.

High Achiever

The students categorised in this way in the analysis had positive ability conceptions and high self-confidence, which was increased following positive feedback on draft work. They had a high predetermined achievement level in their mind prior to submission. They also had self-imposed pressure to perform and equated that the more effort they put into their work, the higher grade they should have achieved. Receiving a poor grade, for the students categorised as high achiever, resulted in emotions such as disappointment, frustration and a deflated mood. Emotional reactions and the lower than normal grade outcome could increase personal pressure

to compensate by performing better in the next assessment. Consequently they felt more motivated and increased effort in the next assessment. Feedback on poor work was interpreted as a reflection of the actual work and not themselves. The feedback language was understood, and although at times the grade outcome and emotional reaction prevented them from engaging with the feedback immediately, they eventually were able to use the negative feedback to promote improvement in the next assessment. They used one-to-one meetings with their lecturer to clarify the feedback messages to ensure improvement occurred. When receiving a grade which they considered to be good, these students experienced jubilant emotions and were motivated to maintain this level. Their feedback-related thoughts did not reflect a high level of engagement. They understood the language used in the feedback, but in most cases, it was only confirming they had done well and developmental feedback towards the next assessment was absent. They did appreciate the feedback, but their confidence level was so high they vowed to replicate their approach in the next assessment.

Discussion

In this chapter the outcome space displayed five categories of description which represent the qualitatively different ways the students experienced and utilised feedback. The outcome space represents their multifaceted experiences of assessment and feedback. Frequently the feedback literature articulates research which at times views students' experiences as rather one-dimensional in nature. The research reported within this chapter has demonstrated firstly the interactional nature of the student's feedback experiences and secondly how such experiences impacted upon their future assessment practices. Previous literature has reported such notions as students not adhering to feedback, not picking up feedback and not understanding the language of feedback (Bloxham & Campbell, 2010; Fisher, Cavanagh, & Bowles, 2011; Higgins, Hartley, & Skelton, 2001). In this chapter I have holistically explored the underlying more complex constructs which have been promoting such behaviours. The outcome space in this chapter illuminates how the constructs interacted in order to influence the student's tendency to engage in adaptive or maladaptive feedback practices. The outcome space deviates from the traditional representation of feedback research-related data, not least in the fact that the student's emotional and cognitive processing was aligned to their behavioural actions in a way that allowed one to understand the sometime nuanced differences between groups of students as depicted by the five categories of description. The categories of description indicated findings which supported contentions within the existing literature. For example, findings for students who typified what I categorised as low achievers, rationalising and needy aligned with previous literature suggestions that when students are satisfied with the grade outcome, they either do not read or attend to the feedback messages (Lipnevich & Smith, 2009; Pitt & Norton, 2016; Vardi, 2009). Conversely, students who I identified as high achiever

and emotionally changeable were able to use the feedback despite achieving above their predetermined grade level (Pitt & Norton, 2016).

Maladaptive behaviours can be seen with regard to emotional reaction in the case of the needy, rationalising and low achiever categories; the positive emotional reactions they experienced did not positively influence feedback use, conflicting with previous literature suggesting that positive emotions produce more resilient students (Beard et al., 2014; Rowe, Fitness, & Wood, 2013). This also seems to contradict both Pekrun et al. (2002) and Fredrickson and Cohn (2008) who argued that positive emotions can act to enhance a student's learning and achievement due to their inherent propensity to assist self-regulation and motivation. Although this did appear to be the case for the high achiever category, as the presence of positive feedback enabled them to utilise the feedback in order to improve, in this regard Boud and Falchikov's (2007) and Nash et al.'s (2016) findings seem to be corroborated in that the student's emotional reactions obstructed their cognitive processing of the feedback. However, in the case of the emotionally changeable category, they did return to the feedback once the emotional reaction had passed, suggesting a more developed ability to self-regulate. When poor grades were received, it appeared that predispositions were affected by previous assessment experiences, lending support to my contention that feedback is cyclical in nature. It was apparent that many constructs, such as emotions, grades and motivation, were interacting with the student's predispositions in order to affect their subsequent feedback utilisation.

Negative feedback produced a discernible dichotomy of reactions between the categories of description. In an adaptive sense, the high achiever and emotionally changeable categories appeared to support the notion that negative feedback is seen as motivational. However, some students reported that negatively phrased feedback appeared to cause them, especially those who are already low in confidence, to react in a very negative manner which has been frequently reported in the literature (Poulos & Mahony, 2008; Weaver, 2006). This was especially noticeable with the needy and low achiever categories, which reinforces my earlier argument in this chapter that such instances served to influence their subsequent predispositions in the next assessment.

Wider Implications of the Findings

Nicol (2009) has argued that one issue relating to feedback practice is the culture which students bring with them when entering university. In a sense what Nicol (2009) is suggesting is that some students may enter university without the necessary self-regulatory skills in order to successfully engage with a change in feedback practices. This does appear to be particularly the case for the categories labelled as low achiever and needy, suggesting that they struggle with self-assessment skills and are unable to regulate their learning by using the feedback provided (Orsmond & Merry, 2013). Monologic transmission feedback for these categories appears to

be ineffectual, and other ways of engaging students in self-reflection perhaps need to be explored. More dialogic forms of feedback consisting of elements of peer learning and draft formative submissions alongside exemplars of what constitutes work awarded at the various grade point intervals (see, e.g. Beaumont et al., 2011) may help those in these categories.

The majority of researchers within the realm of feedback research are in agreement that if students are to learn from feedback dialogue, opportunities to act upon the feedback received must be afforded to them (Carless, Salter, Yang, & Lam, 2011; Nicol, 2010). If one is to follow the line of logic that proponents of dialogic feedback articulate, then it does not appear surprising than only a small proportion of students (high achievers) demonstrated an ability to self-regulate their emotions, motivational state and subsequent behavioural actions in order to use the feedback in the next assessment. Within dialogic feedback, students are encouraged to engage in self-judgement and self-regulation (Blair & McGinty, 2012; Carless et al., 2011; Nicol, 2010; Sadler, 1989, 2010). The overwhelming majority of students in my research failed to demonstrate self-regulatory behaviour when receiving a poor grade alongside negative feedback and thus did not utilise the feedback in the next assessment. Changing to a feedback practice which reflects increased dialogue means that students need time to develop skills alongside multiple opportunities in which to engage with this (Carless et al., 2011). However, it does appear that some of the literature and especially Yang and Carless (2013) suggest that students who engage in dialogue will become better equipped to self-regulate. However, my findings suggest that for many students, such as those in the categories of needy and low achiever, this could be a step too far in the short term as their predispositions coupled with their propensity for emotional backwash (Pitt & Norton, 2016) mean they are not in the position to self-regulate. It therefore seems prudent to suggest that any intervention designed to facilitate self-regulation may need to consider the potential maladaptive effects of student predispositions and emotional backwash.

It follows that further research in relation to how students engage with dialogic feedback is needed. This research needs to continue to investigate lower-achieving students within a dialogic feedback framework, as they still remain an under-researched population group. In particular the research needs to progress what Orsmond and Merry (2013) report, alongside the research in this chapter in order to further understand the adaptive and maladaptive approaches of lower-achieving students. In order to scale this up, such practices require a cultural shift at not only modular level but also at the institutional level. It is apparent that not only the students who are less able to self-regulate but also those who are able to self-regulate might benefit from such practice. As this chapter has indicated, even the higher achievers sometimes think they have reached a ceiling when they achieve a high grade, yet at times the submission could still have been improved beyond that. The key seems to be that in order for students to really understand, appreciate and utilise feedback, a consciousness-raising exercise about the different categories might help tutors tailor their feedback comments. However, it is clear that in order to achieve this at institutional level, current feedback practices may need to be

modified to reflect more developed assessment pedagogy. For example, and this was certainly the case for the students involved in my research, content-driven lectures and seminars followed by summative assessments designed to 'test' the learning outcomes of a module are commonplace. This type of pedagogy will result in the same issues relating to assessment and feedback that I and many others have discussed over the last 30 years. Embedding assessment and feedback literacy development into teaching sessions may provide a useful approach to tackling the apparent issues discussed in this chapter. Over time getting students to appreciate the purpose of assessment and feedback and how it can improve their work and by inference their performance outcome needs to be fostered (Price, Rust, O'Donovan, Handley, & Bryant, 2012). This could be achieved by developing students' ability to self-regulate emotions and feedback by exposing them to multiple formative opportunities, facilitated by peer learning, exemplars and reflection. This may instigate a movement away from such a heavy focus upon grade outcome, which needs to be achieved so that students are engaged in the discourse surrounding academic work far earlier and experience more frequent, higher in quality and supportive feedback prior to the summative submission.

This chapter has demonstrated how students' assessment predispositions interacted with their predetermined grade expectations to determine how they perceived their grade outcome. Their subsequent adaptive or maladaptive emotional reactions influenced their behavioural actions and feedback utilisation in subsequent assessment opportunities. It is apparent therefore that if lecturers can positively affect students' assessment and feedback cognitions at the earliest stage possible, the categories of needy, low achiever, rationalising and emotionally changeable may be less commonplace promoting more students reflecting the assessment and feedback behaviours of the high achiever category. In an applied sense, lecturers could attempt to help students appreciate how to adaptively utilise the feedback they receive by sharing with them how they deal with negative feedback on, for instance, journal submissions or grant applications.

References

Åkerlind, G. S. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research and Development, 24*(4), 321–334.

Åkerlind, G. S. (2008). An academic perspective on the nature of research: A review and empirical extension of the literature. *Studies in Higher Education, 33*(1), 17–31.

Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflection on their practices. *Teaching in Higher Education, 15*(2), 187–198.

Beard, C., Clegg, S., & Smith, K. (2007). Acknowledging the affective in higher education. *British Educational Research Journal, 33*(2), 235–252.

Beard, C., Humberstone, B., & Clayton, B. (2014). Positive emotions: Passionate scholarship and student transformation. *Teaching in Higher Education, 19*(6), 630–643.

Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education, 36*(6), 1–17.

Blair, A., & McGinty, S. (2012). Feedback-dialogues: Exploring the student perspective. *Assessment & Evaluation in Higher Education, 38*(5), 554–566.

Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheet. *Assessment & Evaluation in Higher Education, 35*(3), 291–300.

Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. London: Routledge.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407.

Fisher, R., Cavanagh, J., & Bowles, A. (2011). Assisting transition to university: Using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education, 36*(2), 225–237.

Fredrickson, B. L. & Cohn, M. A. (2008). Positive emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett, L. F. (Eds.). Handbook of emotions (3rd ed.) (pp. 777–796). New York: Guilford Press.

Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': Investigating the concept of student engagement with feedback. *Oxford Review of Education, 37*(4), 543–560.

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education, 6*(2), 269–274.

Kahu, E., Stephens, C., Leach, L., & Zepke, N. (2015). Linking academic emotions and student engagement: Mature-aged distance students' transition to university. *Journal of Further and Higher Education, 39*(4), 481–497.

Lipnevich, A. A., & Smith, J. K. (2009). I really need feedback to learn: Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability, 21*(4), 347–367.

Lizzio, A., Wilson, K., Gilchrist, J., & Gallois, C. (2003). The role of gender in the construction and evaluation of feedback effectiveness. *Management Communication Quarterly, 16*(3), 341–379.

Nash, G., Crimmins, G., & Oprescu, F. (2016). If first-year students are afraid of public speaking assessments what can teachers do to alleviate such anxiety? *Assessment & Evaluation in Higher Education, 41*(4), 586–600.

Nicol, D. J. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education, 34*(3), 335–352.

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

O'Donovan, B., Rust, C., & Price, M. (2015). A scholarly approach to solving the feedback dilemma in practice. *Assessment & Evaluation in Higher Education.* doi:10.1080/02602938.2015.1052774.

Orsmond, P., & Merry, S. (2013). The importance of self-assessment in students' use of tutors' feedback: A qualitative study of high and non-high achieving biology undergraduates. *Assessment & Evaluation in Higher Education, 38*(6), 737–753.

Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilisation of tutors' formative feedback: A qualitative interview study. *Assessment & Evaluation in Higher Education, 30*(4), 369–386.

Pekrun, R., Goetz, T., & Titz, W. (2002). Academic emotions in students' self-regulated learning and achievement: A programme of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105.

Pitt, E. (2014). *Feedback in higher education: Exploring students' appraisal, comprehension and utilisation*. Doctor of Philosophy (PhD) thesis, University of Liverpool. Retrieved from https://kar.kent.ac.uk/cgi/users/home?screen=EPrint::View&eprintid=47524#t

Pitt, E. (2016). Drawing: A visual method as an expressive data collection technique. In S. Watt & C. Wakefield (Eds.), *Teaching visual methods in the social sciences*. London: Routledge.

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*. doi:10.1080/02602938.2016.1142500.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143–154.

Price, M., Rust, R., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford, UK: Oxford Centre for Staff and Learning Development.

Rowe, A., Fitness, J., & Wood, L. (2013). University student and lecturer perceptions of positive emotions in learning. *International Journal of Qualitative Studies in Education, 28*(1), 1–20.

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

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550.

Stiles, D. R. (2004). Pictorial representation. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 123–139). London: Sage Publications.

Trigwell, K. (2006). Phenomenography: An approach to research in geography education. *Journal of Geography in Higher Education, 30*(2), 367–372.

Vardi, I. (2009). The relationship between feedback and change in tertiary student writing in the disciplines. *International Journal of Teaching and Learning in Higher Education, 20*(3), 350–361.

Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379–394.

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

Chapter 11
Feelings About Feedback: The Role of Emotions in Assessment for Learning

Anna D. Rowe

Abstract Feedback is a key element of quality teaching, which both evaluates and supports student learning. The role of emotions in assessment and feedback is less well understood than cognitive and motivational components. However, emotion is just as important – a student's willingness to engage with feedback is determined in part by its' emotional impact. This chapter provides an overview of key developments in the area of university students' emotions pertaining to feedback drawing on recent research from social psychology and education. Given that emotions and feedback are an under-explored area of research, the question of how understanding emotions can contribute to the scaling up of assessment for learning is problematised. Potential strategies for learners, teachers and faculty-student partnerships are discussed. Depth and sustainability are presently the most relevant dimensions of scaling up for emotions, assessment and feedback due in part to the psychological processes underpinning the self-management skills needed to regulate emotions.

Introduction: Feedback in Higher Education

Feedback is a key element of quality teaching (e.g. Black & Wiliam, 1998), serving a variety of purposes, which fall under the two broad functions of evaluation and the support of student learning (Rowe, Fitness & Wood, 2014). Purposes may include grading achievements, clarifying instructional expectations, developing students understanding, motivating students and communicating praise in student's work (e.g. Sadler, 2010; Shute, 2008).

Research shows that students want quality feedback, e.g. timely, comprehensive feedback that explains their performance against assessment criteria and identifies areas for improvement (e.g. Rowe, 2011). However, conceptions of quality have changed in recent years following the re-examination of feedback in teaching and learning scholarship (e.g. Nicol, 2010; Sadler, 2010). Previously, the focus was

A.D. Rowe (✉)
Macquarie University, Sydney, NSW, Australia
e-mail: anna.rowe@mq.edu.au

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_11

on the practicalities of feedback delivery, i.e. 'its' promptness, level of detail, clarity, structure and relevance' (Nicol, Thomson & Breslin, 2014, p. 102). Recently attention has shifted to the individual meaning of feedback for the learner and learner interactions with feedback and teachers (Boud & Molloy, 2013a). This view of feedback as a process of two-way communication and dialogue within specific social contexts, rather than one-way transmission of information from teacher to student, offers a new model positioning students as active learners (Dowden, Pittaway, Yost & McCarthy, 2013; Nicol, 2010; Nicol et al., 2014; Sadler, 2010; Yang & Carless, 2013).

Despite its importance, dissatisfaction with feedback is widely reported both within Australia and the UK (e.g. ACER, 2009; Brinkworth, McCann, Mathews & Nordström, 2009), and as noted by Carless (this volume, Chaps. 1 and 8), such findings underpin the recent scaling up of attention to feedback in assessment for learning (AfL). This chapter shows how better understanding the role of emotions in AfL, and feedback in particular, may provide valuable insights into the contribution of feedback to student learning and offer new directions for improving feedback practices in higher education.

Whilst it is acknowledged that the high-stakes nature of assessment creates the potential for strong feelings (Falchikov & Boud, 2007), until recently little was known about the role that such feelings play in how students respond to assessment and feedback in higher education. Indeed, emotions and learning is an emerging area of research and scholarship, which hasn't attracted the same level of attention as feedback in AfL. Once thought of as a barrier to rational and objective thinking, the role of emotions in promoting learning has only recently been recognised. This may be due in part to emerging work which supports the idea that learning involves the 'whole person', of which emotions are part, and also includes cognitive, neural, motivational and social factors (e.g. OECD, 2015).

In this chapter key current developments in the area of university students' emotions pertaining to feedback are explored. Specifically, the latest empirical research is synthesised, and a proposal for how emotion theory can inform our understanding of assessment and feedback in higher education offered. Finally, attention is given to how understanding emotions can contribute to the scaling up of AfL.

Why and When Do Students Experience Particular Emotions?

Much previous educational research has acknowledged the presence of affect and emotions generally (e.g. Crossman, 2007; Poulos & Mahony, 2008) but has not 'unpacked' or examined these concepts in relation to learning in a systematic way. Further, much scholarship has largely been based on dimensional approaches that effectively treat all discrete emotions (e.g. anger, sadness, joy) as globally 'negative' or 'positive' affective states. When discrete emotions are combined into

overall negative or positive dimensions for analytical purposes, we can overlook the different antecedents, functions and outcomes associated with discrete emotions (Gooty, Gavin & Ashkanasy, 2009). That is, emotions are preceded by an event (student receives extensive praise for their work); they serve a particular function (feelings of pride lead the student to desire obtaining further praise in the future) and lead to outcomes (increased effort for the next assessment).

The field of social psychology provides a useful conceptual framework from which to understand why and when particular emotions are experienced by students and what happens as a result of such experiences. Several concepts can be applied; firstly social factors 'are implicated in emotion in many ways and at many levels' (Parkinson & Manstead, 2015, p. 372), and feedback occurs within a social context. A substantial body of evidence supports the importance of relationships (both student–teacher and peer relationships) to learning broadly (e.g. Cornelius-White, 2007; Crossman, 2007) and to feedback (Rowe, 2011). Indeed, some evidence suggests that students view relationships with teaching staff as a factor mediating their perceptions of and responses to feedback (Dowden et al., 2013; Rowe et al., 2014). Supportive student–teacher relationships are fundamental to the establishment of dialogue which is central to feedback, and feedback is reciprocal in the sense that it involves both giving and receiving.

Secondly, emotions are notoriously difficult to conceptualise, with terms such as 'emotion', 'affect', 'mood' and 'feelings' often used interchangeably and inconsistently in the literature. Cognitive approaches to emotions, whilst not accepted by everyone, offer 'a clarifying perspective because they focus on the fundamental issues of how emotions are caused and what their effects are' (Oatley & Johnson-Laird, 2014, p. 134). Social functional and cognitive appraisal theories are two approaches offering explanations that can help to elucidate reasons for and likely timing of specific emotional reactions in response to different teaching and learning contexts, such as feedback.

Of relevance is the strong body of empirical evidence supporting the notion that emotions serve a range of intra- and interpersonal problem-solving functions, i.e. they motivate, organise and regulate behaviour (e.g. Oatley & Johnson-Laird, 2014). For example, emotions motivate adaptive behaviours in order to avoid negative outcomes (e.g. conflict, failure) and promote positive ones (e.g. achievement, social affiliation). Within feedback contexts, emotions may, for example, help students to recover from negative evaluations, provide protection against rejection, make sense of feedback and promote cooperative relationships with teaching staff and peers (Rowe et al., 2014). The social functional approach thus provides a framework from which to understand why particular emotions occur.

Cognitive appraisal models propose that emotions arise as a function of people's appraisals, or interpretations, of particular situations/events and concerns such as success, failure or perceived criticism (Lazarus, 1991; Moors, 2014; Oatley & Johnson-Laird, 2014). It is the appraisal, not the event itself that elicits an emotion. Appraisal theory has gained substantial empirical support and a high degree of consensus exists about the kinds of appraisals underpinning emotions such as fear,

anger and sadness (Scherer, 2001), i.e. whether the situation/concern aligns with the person's goals, the extent to which they feel able to control and cope with the situation and who or what is perceived to be the cause (another person, themselves or random circumstances). Within feedback contexts, a student may experience anger if they assign blame for a critical evaluation to their teacher or experience sadness/disappointment if they perceive the feedback to be caused by their own shortcomings. The different emotions experienced have varying implications for behaviour; an angry student may complain to their lecturer rather than engaging with feedback, whilst a sad student might withdraw from the course if they feel they weren't good enough to pass. The experience of emotion can prompt both approach and withdrawal behaviours which have potentially different outcomes for student learning (see also Pitt, this volume, Chap. 10).

Thus, cognitive appraisal theory helps us as teachers, academic developers and researchers to understand when particular emotions are experienced, why different emotions can be experienced in response to the same event and what behaviours are likely to result. Pekrun, Frenzel, Goetz and Perry (2007), for example, have proposed a control–value theory to account for the role of appraisals and emotions in educational contexts. This chapter similarly adopts a cognitive approach, conceptualising emotions as responses to appraisals of situational events or concerns, which are associated with various response tendencies involving coordinated changes in physiology, motor readiness, behaviour, cognitive processing and subjective experience (e.g. Lazarus, 1991). Emotions are experienced for short periods of time and viewed as fitting into discrete categories of emotion families (e.g. fear, joy), rather than more general or dispositional feeling states.

What Emotions Are Important in Feedback? Do They Enhance or Hinder Learning?

Whilst it is recognised that assessment is 'deeply personal' (Crossman, 2007, p. 322), creating the potential for strong feelings, less is understood about how discrete emotions affect the way that students are able to receive and process feedback (Dowden et al., 2013; Rowe et al., 2014). Several achievement emotions have been identified as being important to learning generally, including enjoyment, hope, pride, anger, anxiety, shame, hopelessness, relief, boredom and interest (e.g. Pekrun & Stephens, 2010; Rowe, Fitness & Wood, 2015; White, 2013). Less is known about social emotions such as gratitude in learning contexts. We know from empirical evidence that positive emotions generally enhance academic learning and achievement, particularly in areas of metacognition, effort on task, self-regulation, strategy use and motivation (Pekrun & Stephens, 2010). Negative emotions, whilst being detrimental to motivation and performance in many cases (e.g. the role of fear in avoidance behaviour), can also be beneficial (Pekrun & Stephens, 2010).

For example, anger might prompt a student to initiate a conversation with their teacher regarding feedback perceived to be undeserved. The promotion of such dialogue (provided it is constructive) is important for students' understanding of and engagement with feedback (Nicol, 2010; Nicol et al., 2014).

Positive and negative achievement emotions are experienced in assessment contexts including feedback, although less is known about the role they play (Peterson, Brown & Jun, 2015; Rowe et al., 2014). Educational literature has pointed to feedback as an important source of confidence and pride (Beard, Humberstone & Clayton, 2014; Pitt and Norton, 2016; Shields, 2015; Theising, Wu & Sheehan, 2014), as well as anxiety and shame (Cramp, Lamonda, Coleyshawa & Beck, 2012; Shields, 2015; Peterson et al., 2015; Rowe et al., 2014). Some studies suggest negative affect at the prospect of assessment and feedback (e.g. feelings of anxiety and doubt) may even heighten subsequent positive affect, e.g. experiences of euphoria (Beard et al., 2014; Cramp et al., 2012). Recent research by Rowe et al. (2014) point to joy/happiness, love (gratitude), anger, fear, pride, embarrassment and guilt, as playing a key role in how students respond to feedback. These emotions are experienced in response to various achievements and academic and social situations/concerns which relate to feedback's dual role as an evaluation tool and source of support (Rowe, 2011; Rowe et al., 2014). 'Self-conscious' emotions – which include pride, guilt and embarrassment – appear to be particularly salient responses to the evaluative component of feedback. This is likely because self-conscious emotions are strongly tied to social concerns (e.g. status, power, attachment) (Hareli & Parkinson, 2008). There is a more pronounced power dynamic between students and lecturers in relation to evaluation than there is to feedback's supportive mechanism, so it is not surprising self-conscious emotions are key here, as they signify the importance of certain relationships, helping to maintain and restore such relationships (Rowe et al., 2014). For example, when guilt is experienced as a result of self-blame for a poor outcome such as a critical evaluation, it might stimulate reparative action by the student to repair perceived damage to the relationship with their teacher, leading to an apology and resolution to work harder in the future (e.g. Hareli & Parkinson, 2008).

Broader social emotions (especially gratitude) and other discrete emotions such as happiness, interest/excitement, anger, fear and sadness also play an important role in how students respond to evaluative and supportive components of feedback (Rowe et al., 2014). For example, memory is enhanced when events have an emotional aspect (Talmi, 2013), and attention is broadened by positive emotions such as happiness and narrowed by anxiety (Huntsinger, 2013). Thus, the type and intensity of emotions experienced by students when receiving feedback may have implications for how well they remember and respond to it. Memory and attention are just two examples of cognitive processes affected by emotion; others include motivation, problem solving and information processing (e.g. Pekrun & Stephens, 2010). In terms of feedback's supportive function, gratitude, for example, has been found to motivate a willingness to forgive and sympathise, attribute positive

outcomes to the actions of others, notice and reciprocate help and act as an antidote to aggression (e.g. Wood, Froh & Geraghty, 2010). It could be inferred from such findings that when students experience gratitude in response to feedback, they might be more likely to approach teachers for help or be less inclined to complain about feedback if it does not meet their expectations in some way, i.e. it sets up a positive basis for communication. Such responses can also be influenced by student's emotional maturity and perceptions/expectations of grading and feedback (Pitt & Norton, 2016).

How Can Understanding Emotions Inform the Scaling Up of Feedback Practices in AfL?

The task of translating research evidence into effective instructional practice at scale is especially difficult for feedback and emotions at individual, group, discipline or organisational levels. Surprisingly, feedback practices have remained 'largely uninfluenced by ideas, or practices, or research on feedback from outside the education sector' (Boud & Molloy, 2013b, p. 698). Further, personal dimensions of feedback, including emotions, remain largely unacknowledged and absent from institutional level governance documentation such as assessment policies (Rowe, 2013). On the positive side, there is now recognition (at least in the literature) that timely and detailed feedback, whilst important, are not enough to promote the kinds of lasting developmental changes (such as building of student's self-regulation skills) (Boud & Molloy, 2013b; Carless, Salter, Yang & Lam, 2011; Sadler, 2010) and/or emotional outcomes (such as alleviating concerns about failing) (Shields, 2015) needed to improve student learning. In the following section, emotions are thus problematised as they relate to the scaling up of assessment. Key challenges and tensions are discussed and suggestions offered for scaling up feedback practices in AfL.

The preceding review suggests two key areas in need of further consideration with regard to scaling up. Firstly, that feedback will evoke emotions in students and teachers alike; hence, it is important both understand the roles that positive and negative emotions play in learning, teaching and assessment. Secondly, although perhaps less pertinent to assessment and feedback practices is that some kinds of learning may bring emotion to the fore (e.g. reflection, group work, challenging student assumptions and the development of graduate attributes such as resilience). The first, the evocation of emotion, forms the focus of the following discussion. It is posited that at present, depth and sustainability are the most relevant dimensions of scaling up for emotions, assessment and feedback due in part to the psychological processes underpinning the self-management skills needed to regulate emotions. As such, the following discussion is framed around these two dimensions.

The Challenge of Achieving Depth

What the Teacher Does

Much of the focus of recent educational scholarship on emotions emphasises the need for teachers to understand student emotions and responses to feedback (Rowe, 2013; Storrs, 2012). Specific practical approaches that have been offered include encouraging teachers to 'consider teaching strategies that elicit enhanced positive emotions, as these experiences help to broaden students' cognitive functioning and improve their ability to learn course content' (Goldman & Goodboy, 2014, p. 272), and strategies to reduce anxiety/negative emotions in students which interfere with their ability to engage with feedback (Rowe, 2013; Shields, 2015). Shields (2015), for example, advocates for the incorporation of more low-stakes assessment and/or assessment which offers students a 'second chance' as a way of reducing anxiety, given the links between student interpretations of feedback comments and their beliefs about themselves as learners (i.e. being 'good' enough, 'being wrong'). Other research emphasises the need for teachers to devote additional time to improving dialogue with students, as well as the importance of making changes at a curriculum level, and to individual teaching practice:

> Instructors committed to an emotional curriculum must be informed of the high degree of engagement and time investment required to read and respond to journals and meet individually with students. (Storrs, 2012, p. 10)

Whilst such strategies are undoubtedly crucial, they are associated with a number of tensions. Firstly, they are based on the assumption that 'responsibility for change is in the hands of academics who plan assessment' (Cramp et al., 2012, p. 517). Their focus is on what the teacher does and thus remains grounded within transmission approaches to AfL. Secondly, Nicol et al. (2014) are wary of interventions which potentially increase academic workload and can be seen as 'problematic given current resource constraints and rising student numbers in higher education' (p. 103). A further point to consider is the extent to which teachers in higher education 'feel equipped to facilitate the creation of emotionally positive and emotionally aware learning environments (or indeed the extent to which they see this as part of their role)' (Moore & Kuol, 2007, p. 95). Indeed, there is a perception by some lecturers that student engagement with feedback is limited (Price, Handley, Millar & O'Donovan, 2010), and such perceptions are likely to reduce the amount of thought and effort invested by teachers in providing feedback to students. Formal and informal communities of practice and other teaching peer or personal networks are one potential strategy to support staff towards implementing change by building knowledge and skills and offering emotional support (e.g. Storrs, 2012).

What the Learner Does

Teaching students' self-management skills to more effectively respond to the strong emotions evoked by feedback is another area of focus. This is more in line with student-centred approaches to learning (Rowe, 2013), reflecting the current emphasis on promoting self-regulation in feedback practices, 'the essence of sustainable feedback' (Carless et al., 2011, p. 398; see also Boud & Molloy, 2013b). Encouraging students to take a more active role through monitoring their own work and increased dialogue around feedback practices (Carless et al., 2011; Nicol, 2010) has already been taken up by some disciplines, particularly in the first year. For example, self-management skills are being incorporated into the first year law curriculum as a way of promoting independent learning and reflective/self-assessment capabilities in law graduates, which includes using feedback effectively particularly when it is 'negative' (James & Field, 2013). Such changes have emerged in response to mental health and wellbeing concerns for law students. Emotions are a key component of self-regulation (Pekrun, Goetz, Titz & Perry, 2002), and self-regulation in coping and learning styles can be taught and promoted. For example, teaching learners to reframe (or reappraise) negative feedback in a positive way through cognitive restructuring or cognitive reappraisal techniques (e.g. Gross & John, 2003) may be potentially useful for managing and regulating emotions in feedback contexts, leading to enhanced cognitive performance (Raferty & Bizer, 2009). However, achieving such changes on a large scale presents challenges, e.g. workload implications for lecturers.

Peer and self-assessment are also promoted as a way of increasing student responsibility for learning, developing students' abilities to make judgments about their own and others work and addressing a number of issues including staff workload (e.g. Nicol et al., 2014). Giving and receiving feedback provides opportunities for students to develop different types of skills, than through receiving feedback alone (Nicol et al., 2014), as well as potentially contributing to emotional ones. For example, anxiety can be reduced through supportive learning groups (Cramp et al., 2012). Peer feedback may offer similar potential, with peer assessment found to positively impact on perceptions of self-confidence in some circumstances (Theising et al., 2014). Whilst such strategies seem promising, only students can act on feedback, so obtaining their buy-in is crucial. Some students may be reluctant to adopt practices such as self-evaluation, which require autonomous learning; and teachers similarly may be reluctant to engage in such practices out of concern for negative impacts on their teaching evaluations as well as the consequent loss of time to cover disciplinary content (Carless et al., 2011). Institutional level resourcing to support professional development activities targeted at changing academic attitudes and behaviours is needed (Carless et al., 2011).

Teacher and Learner Partnerships

The third set of strategies extends the student's position from active learner to partners in learning and teaching. Cramp et al. (2012), for example, call for lecturers and learners 'to reflect on experiences of schooling together and anticipate reactions to future assessment judgments' (p. 518), and Boud and Molloy (2013b) call for feedback processes that are 'mutually constructed and co-dependent' (p. 711). Such discussions could be undertaken either at course level (e.g. providing opportunities for students to reflect and respond to feedback with teachers and peers, including staff talking with students about their own experiences of receiving feedback) or at institutional level (i.e. broader high-level discussions about the provision of feedback at university). Clarifying expectations and building shared understandings of assessment and feedback between academics and students are important for developing trust (Carless, 2009) and for cultivating relationships that promote the effective use of feedback (Price et al., 2010). It goes to follow that such experiences would likely promote positive feelings. Emotions such as empathy and compassion are potentially important here because of their relationship with shared meaning and goals and perceived similarities with others (Gibbons, 2011; Goetz, Keltner & Simon-Thomas, 2010), although further work is needed to investigate such links in educational contexts.

There are a number of tensions associated with this approach. Firstly, in order to effectively promote students as 'key players in the educational process', such dialogue must be 'genuine' (Carey, 2013, p. 257). There is also the question of the extent to which learners will want to be involved, and their reluctance to participate can be attributed in part to the way institutions frame their engagement:

> Partnership is not a one-off exchange, but an ongoing process that should characterise the whole student experience. There is little value waiting until the curriculum needs an overhaul before inviting students. (Carey, 2013, p. 258)

Finally, any scaling up of dialogue needs to be supported by resourcing and contained by clear boundaries and expectations, e.g. progress reports are useful for providing students with frequent, timely feedback and opportunities to feedforward but can also increase student enquiries, which may contribute to staff workload.

Student–faculty partnerships are ideally based on principles of respect, reciprocity and shared responsibility between learners and teachers (Cook-Sather, Bovill & Felten, 2014), each of which has a strong emotional underpinning. For example, Rowe et al. (2015) and Beard et al. (2014) have reported both students and teachers feel a duty to reciprocate energy/efforts invested by each other, and this appears to be tied to notions of respect. Feedback is an exchange; thus, it has the potential to promote reciprocity and other prosocial behaviours through feelings of gratitude and appreciation (Rowe, 2013). Involving students as partners in the development of feedback practices, whether at course or institutional level, could contribute to an enhanced sense of belonging (Cook-Sather et al., 2014), with affiliation linked to calming emotions (Gilbert, 2009). Finally, as Beard et al. (2014) notes, 'this emotional reciprocity is activating, and suggestive of a moral imperative'

(p. 638). The effects of an environment promoting the wellbeing of students and mutual care and respect need to be better understood.

The Challenge of Sustainability

Any enactment of sustainable changes in AfL needs to firstly be underpinned by a conceptual shift in how feedback is viewed. Boud and Molloy's (2013b) recent call for a 'fundamental rethinking of the place of assessment and feedback within the curriculum', to enable 'a more robust view of feedback: one that focuses primarily on the needs of learning rather than the capacities of the teacher' (p. 698), is essential here. In practice such rethinking could translate to a shift in focus from providing feedback to the embedding of feedback within learning design and emphasising interactions between students and lecturers (Boud & Molloy, 2013b). Cramp et al. (2012) aptly draw the link between the potential broader impact of such changes on how emotions are viewed in assessment, by observing, 'it is acknowledged that assessment in higher education causes anxiety, but this is often regarded as a 'problem' sited in the individual, not in the pedagogy' (p. 519). Such reconceptualisations have the potential to address critical issues in feedback by focussing on opportunities for teachers and students to communicate, in turn allowing for more depth, meaning and creation with the curriculum.

Conceptual changes need to occur at an institutional level before practice-based changes can be enacted. Ideally institutions would provide 'structural opportunities in terms of adequate time and smaller class sizes to allow students and faculty to experience the learning benefits of such pedagogy [i.e. emotional]' (Storrs, 2012, p.10). In actuality, however, such changes may not be possible nor indeed likely, with contextual factors such as the overloading of teaching staff and increasing class sizes an ever present reality (e.g. Gibbs, 2010). However, changes could potentially be achieved through partnerships between academics, students and professional staff fostered at an institutional level. Kift, Nelson and Clark's (2010) notion of a 'transition pedagogy' for the first year experience is useful here. Initiatives that enable first year students to 'achieve engagement, timely access to support and the development of a strong sense of belonging' require:

> The bringing together of co-curricular and curricular strategies into an intentionally designed and broadly conceptualised curriculum; one which is implemented through the shared knowledge and skills of partnered academic and professional staff in an institutional environment that is committed to an optimal first year experience both at the policy and practice levels. (Kift et al., 2010, p. 10)

Similar approaches could be implemented in feedback practices, e.g. assessment and feedback policies could be reframed to recognise personal, social and emotional factors, as well as the more practical elements of timing. Constructive feedback is particularly important in the first year, orientating learners to the expectations of higher education, as well as supporting their transition generally, with high

anxiety reported amongst many first year students (e.g. Brinkworth et al., 2009; Rowe, 2011). Policies such as graduate attribute statements could also be better integrated, e.g. intentionally align the development of student emotional capabilities such as interpersonal skills and resilience, with the development of students' self-management skills in relation to feedback. Indeed, the requirement to develop emotional and interpersonal capabilities as articulated through graduate attribute statements will likely contribute to an enhanced awareness by staff of emotions in learning more broadly.

Conclusion

Emotion presents a number of challenges to scaling up feedback practices in AfL. Social psychology has established that emotions serve particular functions and are related to student perceptions (appraisals, beliefs about self). Further scientific evidence is needed to determine the precise mechanisms of the processes involved and contribute to the development of new frameworks which better explain the relationship of emotions to cognitive, motivational, neurological and social dimensions of feedback and assessment. At the same time, feedback needs to be 'repositioned as a fundamental part of curriculum design, not an episodic mechanism delivered by teachers to learners' (Beard et al., 2014, p. 698). Theoretical and empirical advances, coupled with a deeper level of engagement and strengthened relationships between academics, professional staff and students at individual, group and institutional levels, are needed to advance feedback practices in AfL.

Acknowledgements The author wishes to express gratitude to Felicity Martin for assisting with the literature review and Teresa Dluzewska for providing valuable feedback on drafts.

References

Australian Council for Educational Research (ACER). (2009). Engaging students for success: Australasian student engagement report/Australasian survey of student engagement. Retrieved from www.acer.edu.au/documents/AUSSE_2008-StudentEngagementReport.pdf

Beard, C., Humberstone, B., & Clayton, B. (2014). Positive emotions: Passionate scholarship and student transformation. *Teaching in Higher Education, 19*(6), 630–643.

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

Boud, D., & Molloy, E. (2013a). What is the problem with feedback? In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 1–10). London: Routledge.

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

Brinkworth, R., McCann, B., Mathews, C., & Nordström, K. (2009). First year expectations and experiences: Student and teacher perspectives. *Higher Education, 58*, 157–173.

Carey, P. (2013). Student as co-producer in a marketised higher education system: A case study of students' experience of participation in curriculum design. *Innovations in Education & Teaching International, 50*(3), 250–260.
Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79–89.
Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407.
Cook-Sather, A., Bovill, C., & Felten, P. (2014). *Engaging students as partners in learning and teaching: A guide for faculty*. San Francisco: Jossey-Bass.
Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113–143.
Cramp, A., Lamonda, C., Coleyshawa, L., & Beck, S. (2012). Empowering or disabling? Emotional reactions to assessment amongst part-time adult students. *Teaching in Higher Education, 17*(5), 509–521.
Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research & Development, 26*(3), 313–327.
Dowden, T., Pittaway, S., Yost, H., & McCarthy, R. (2013). Students' perceptions of written feedback in teacher education: Ideally feedback is a continuing two-way communication that encourages progress. *Assessment & Evaluation in Higher Education, 38*(3), 349–362.
Falchikov, N., & Boud, D. (2007). Assessment and emotion: The impact of being assessed. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 144–156). London: Routledge.
Gibbons, S. B. (2011). Understanding empathy as a complex construct: A review of the literature. *Clinical Social Work Journal, 39*, 243–252.
Gibbs, G. (2010). *Dimensions of quality*. York: The Higher Education Academy Retrieved from https://www.heacademy.ac.uk/resource/dimensions-quality.
Gilbert, P. (2009). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment, 15*, 199–208.
Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin, 136*(3), 351–374.
Goldman, Z. W., & Goodboy, A. K. (2014). Making students feel better: Examining the relationships between teacher confirmation and college students' emotional outcomes. *Communication Education, 63*(3), 259–277.
Gooty, J., Gavin, M., & Ashkanasy, N. M. (2009). Emotions research in OB: The challenges that lie ahead. *Journal of Organizational Behavior, 30*, 833–838.
Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality & Social Psychology, 85*(2), 348–362.
Hareli, S., & Parkinson, B. (2008). What's social about social emotions? *Journal for the Theory of Social Behaviour, 38*(2), 131–156.
Huntsinger, J. R. (2013). Does emotion directly tune the scope of attention? *Current Directions in Psychological Science, 22*, 265–270.
James, N., & Field, R. (2013). *The new lawyer*. Milton, Australia: Wiley.
Kift, S., Nelson, K., & Clarke, J. (2010). Transition pedagogy: A third generation approach to FYE – A case study of policy and practice for the higher education sector. *The International Journal of the First Year in Higher Education, 1*(1), 1–20.
Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford, UK: Oxford University Press.
Moore, S., & Kuol, K. (2007). Matters of the heart: Exploring the emotional dimensions of educational experience in recollected accounts of excellent teaching. *International Journal for Academic Development, 12*(2), 87–98.
Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review, 6*(4), 303–307.
Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*(5), 501–517.

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.

Oatley, K., & Johnson-Laird, P. N. (2014). Cognitive approaches to emotions. *Trends in Cognitive Sciences, 18*(3), 134–140.

Organisation for Economic Co-operation and Development (OECD). (2015). *Skills for social progress: The power of social and emotional skills*. Paris: OECD Skills Studies, OECD Publishing.

Parkinson, B., & Manstead, A. S. R. (2015). Current emotion research in social psychology: Thinking about emotions and other people. *Emotion Review, 7*(4), 371–380.

Pekrun, R., & Stephens, S. J. (2010). Achievement emotions in higher education. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 257–306). New York: Springer.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). San Diego, CA: Academic Press.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105.

Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology, 42*, 82–96.

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*. doi:10.1080/02602938.2016.1142500.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143–154.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education, 35*(3), 277–289.

Raferty, J. N., & Bizer, G. Y. (2009). Negative feedback and performance: The moderating effect of emotion regulation. *Personality & Individual Differences, 47*, 481–486.

Rowe, A. (2011). The personal dimension in teaching: Why students value feedback. *International Journal of Educational Management, 25*(4), 343–360.

Rowe, A., Fitness, J., & Wood, L. (2014). The role and functionality of emotions in feedback at university: A qualitative study. *The Australian Educational Researcher, 41*(3), 283–309.

Rowe, A., Fitness, J., & Wood, L. (2015). University student and lecturer perceptions of positive emotions in learning. *International Journal of Qualitative Studies in Education, 28*(1), 1–20.

Rowe, A. D. (2013). *Feedback in higher education: Personal, relational and emotional dimensions*. Unpublished doctoral dissertation, Macquarie University, Australia.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford, UK: Oxford University Press.

Shields, S. (2015). 'My work is bleeding': Exploring students' emotional responses to first-year assignment feedback. *Teaching in Higher Education.* doi:10.1080/13562517.2015.1052786.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Storrs, D. (2012). 'Keeping it real' with an emotional curriculum. *Teaching in Higher Education, 17*(1), 1–12.

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science, 22*(6), 430–436.

Theising, K., Wu, K., & Sheehan, A. H. (2014). Impact of peer assessment on student pharmacists' behaviors and self-confidence. *Currents in Pharmacy Teaching & Learning, 6*, 10–14.

White, C. J. (2013). Higher education emotions: A scale development exercise. *Higher Education Research and Development, 32*(2), 287–299.
Wood, A. M., Froh, J. J., & Geraghty, A. W. A. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review, 30*(7), 890–905.
Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback process. *Teaching in Higher Education, 18*(3), 285–297.

Chapter 12
Conditions and Effects of Feedback Viewed Through the Lens of the Interactive Tutoring Feedback Model

Susanne Narciss

Abstract Feedback is an essential component of assessment for learning processes. Recent feedback frameworks and reviews consider the learner as an active constructor of knowledge and thus emphasize the formative function of feedback. This chapter analyzes the conditions and effects of formative feedback in (higher) education on the basis of the interactive tutoring feedback model (ITF) (Narciss S, Informatives tutorielles Feedback. Entwicklungs- und Evaluations-prinzipien auf der Basis instruktionspsychologischer Erkenntnisse, Waxmann, Münster, 2006; Narciss S, Feedback strategies for interactive learning tasks. In: Spector JM, Merrill MD, van Merrienboer JJG, Driscoll MP (eds) Handbook of research on educational communications and technology, 3rd edn. Lawrence Erlbaum Associates, Mahwah, pp 125–144, 2008; Narciss S, Digital Educ Rev 23:7–26. Retrieved from http://greav.ub.edu/der, 2013). The ITF-model conceptualizes formative tutoring feedback as a multidimensional instructional activity that aims at contributing to the regulation of a learning process in order to help learners acquire or improve the competencies needed to master learning tasks. It integrates findings from systems theory with recommendations of prior research on interactive instruction and elaborated feedback, on task analyses, on error analyses, and on tutoring techniques. Based on this multidimensional view, interactive feedback strategies in (higher) education should be designed in ways to empower students as self-regulated and productive lifelong learners. This chapter describes the ITF-model and outlines conditions affecting feedback efficiency. Furthermore, it illustrates how the components and assumptions of the ITF-model may be linked to formative feedback-design principles. Finally, implications of the ITF-model with regard to scaling up assessment for learning are discussed.

S. Narciss (✉)
Technical University of Dresden, Dresden, Germany
e-mail: susanne.narciss@tu-dresden.de

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_12

Introduction

Feedback is considered a core component of assessment processes and one of the most powerful factors influencing learning in various instructional contexts including higher education (HE) (e.g., Evans, 2013; Hattie & Timperley, 2007; Hattie & Gan, 2011). In instructional contexts the term *feedback* refers to all post-response information which informs the learner on his/her actual state of learning or performance in order to regulate the further process of learning in the direction of the learning standards strived for (e.g., Narciss, 2008; Shute, 2008). This notion of feedback can be traced back to early cybernetic views of feedback (e.g., Wiener, 1954) and emphasizes that a core aim of feedback in instructional contexts is to reduce gaps between current and desired states of learning (see also Ramaprasad, 1983; Sadler, 1989; Hattie, 2009).

Feedback can be provided by various external sources of information (i.e., teachers, peers, parents, computer-based systems) in a large variety of ways and by internal sources of information (i.e., information perceivable by the learner while task processing). Modern information technologies increase the range of feedback strategies that can be implemented in instructional contexts. Furthermore, they can be used to scale up the use of formative feedback strategies (e.g., Whitelock, 2015) and make it possible to provide learners not only with outcome feedback types but also with tutoring feedback strategies (Narciss, 2004, 2008, 2012, 2013): *Tutoring feedback strategies* combine formative elaborated feedback with tutoring and mastery learning strategies. They provide formative evaluative feedback components that help the learners to become aware of any discrepancies that exist between their desired and their current state of competencies. Additionally, they provide (access to) elaborated feedback components (e.g., hints, explanations, attribute-isolation examples) that are aimed at supporting learners in acquiring the competencies necessary for mastering the learning tasks. In doing so, *tutorial feedback strategies* offer (the access to) strategically useful information for task completion, without providing immediately the correct solution, and prompt the learner to apply this information to a next attempt in accomplishing the learning task. Furthermore, after successful task completion, they provide confirmatory positive feedback components (Narciss & Huth, 2006 describe in detail such a tutorial feedback strategy).

To provide instructional designers and researches with a framework for designing and investigating the conditions and effects of such tutoring feedback strategies, I have developed the **i**nteractive **t**utoring **f**eedback model (ITF-model; Narciss, 2006, 2008, 2013). The ITF-model conceptualizes formative tutoring feedback as a multidimensional instructional activity that aims at contributing to the regulation of a learning process in order to help learners acquire or improve the competencies needed to master the learning tasks in a given instructional context. It has been originally developed to provide a basis for the design and evaluation of *interactive tutoring feedback strategies* for computer-based learning tasks in (intelligent) tutoring systems (for a summary of ITF-studies see Narciss, 2013; Narciss et al.,

2014). Yet, the conceptual and empirical work on the ITF-model has also served as a basis for deriving prescriptive principles for designing formative feedback strategies (e.g., Narciss, 2006, 2012). Interestingly, these principles are congruent with the principles suggested by several authors for feedback in HE contexts (e.g., Boud & Molloy, 2013; Carless, Salter, Yang, & Lam, 2011; Evans, 2013; Nicol & Macfarlane-Dick, 2006; Shute, 2008). This chapter firstly describes the ITF-model. Secondly, it uses the model to outline conditions affecting feedback efficiency. Thirdly, it illustrates how the components and assumptions of the ITF-model may be linked to the feedback-design principles. Finally, I discuss implications and limitation of the ITF-model with regard to scaling up assessment for learning.

The Interactive Tutoring Feedback Model (ITF-Model)

The ITF-model is a multidimensional framework that has been developed on the basis of a synthesis of several lines of research (for further details, see Narciss, 2006, 2008, 2013). It suggests that a core function of feedback in instructional contexts is to provide learners with formative or tutoring information on their current state of learning in order to support them in regulating their learning process successfully. Putting the formative or tutoring function into the focus of interest, the ITF-model considers feedback as a multidimensional instructional activity that aims at contributing to the regulation of a learning process in such a way that learners acquire the knowledge and competencies needed to master learning tasks (Narciss, 2013). This notion of feedback situates feedback as a core assessment for learning strategy (see Carless, Chaps. 1 and 8, this volume).

Furthermore, the ITF-model attracts attention to three sets of conditions that have to be taken into account for the design and evaluation of feedback strategies: firstly, the quality of the feedback strategy (e.g., scope, nature, and structure of the information provided and form of presentation); secondly, individual learning conditions (e.g., prior knowledge or level of competencies, metacognitive strategies, motivational dispositions and strategies); and thirdly, situational conditions of the instructional setting (e.g., instructional goals, learning content and tasks).

Rooted in the cybernetic paradigm from systems theory, the ITF-model emphasizes that feedback has to be viewed as one of several basic components of a generic feedback loop, not as an isolated element of instruction. A generic feedback loop consists of a *controlled process* and a *controller.* The *controlled process* has to be described with regard to the variables which are relevant for controlling the system (=*control variable(s)*), the diagnostic components necessary for assessing the current state of the control variables (= *sensor)*, and a component executing control actions if necessary (= *control actuator*). For regulating the controlled process, the *controller* needs clearly defined *reference values or standards* for the relevant control variables, as well as the *feedback* on the current value of the control variables, in order to compare the current value to the reference value and generate a

control action if the comparison reveals a gap between the actual and the reference values. The cybernetic description of the core components of a feedback loop might at first glance appear very technical and not applicable to instructional contexts. Yet, it makes evident that feedback can only contribute effectively to control a system or a process in the direction of the standards strived for, if it is mindfully integrated into a feedback loop.

In contrast to technical systems, however, the learner (i.e., feedback receiver) has to be considered an active participant in the assessment and feedback process. Thus, the ITF-model differentiates two interacting feedback loops for all instructional contexts with an external feedback source (see Fig. 12.1): the learner's feedback loop and the feedback loop of the external feedback source (e.g., teacher, peer, parent, tutoring system). Due to this differentiation, the ITF-model has also been referred to as the interactive-two-feedback-loops model (ITFL-model, Narciss, 2008).

The two feedback loops are intertwined and interact as discussed in what follows (cf. Narciss, 2013). The starting point for both, the internal and the external feedback loop, is the *controlled process* which in an instructional context consists of the acquisition of competencies necessary to master the demands associated with learning tasks (e.g., a written assignment). Based on multidimensional notions of competencies (e.g., Weinert, 2001) as well as models of self-regulated learning (e.g., Butler & Winne, 1995; Boekarts, 1996; Winne & Hadwin, 1998), the ITF-model assumes that the acquisition of competencies has to be controlled on the cognitive, motivational, and metacognitive level. Determining the *control variables* in a learning process with regard to these control levels requires the identification and specification of the competencies which are relevant for mastering the learning tasks. For a complex task in HE such as essay writing, this includes breaking down the complex task into subtasks (e.g., drafting an outline of argumentation, formulating arguments, writing an introduction, language and style) and to identify for each subtask behavioral descriptors that may serve as criteria for measuring to what extent the subtask has been accomplished (e.g., logical correctness of line of argumentation).

However, the identification of the control variables is only the first component for describing an effective feedback loop. As outlined above, several further components with regard to the internal and external feedback loops are needed for a successful control: First, for each of the controlled variables, a *standard or reference level* has to be determined (e.g., line of argumentation follows the principles of scientific inquiry). The ITF-model assumes that in the learner's loop, the learner's *subjective representation of competencies and the related standards* serve as a basis for the learner to set an *internal reference value* with regard to the standards for example in terms of a *desired level of competencies* (i.e., mastering the task requirements of a written assignment on competency level A). In the external loop, the *external representation of standards, competencies, and task requirements* serves as a basis to determine *external reference values and standards* (e.g., an academic essay written by a bachelor student should fulfil scientific inquiry standards).

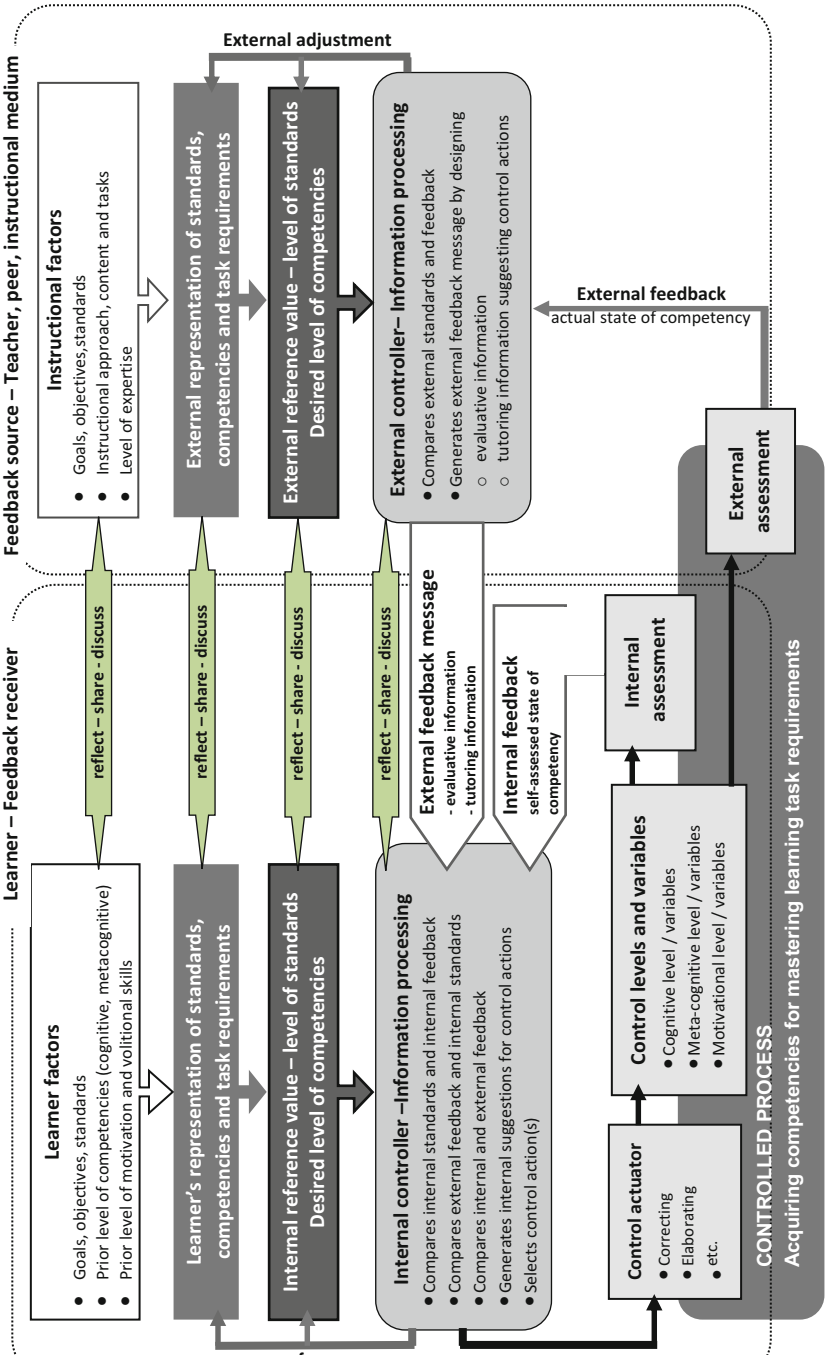

Fig. 12.1 The interactive tutoring feedback model (ITF-model; translated and adapted from Narciss, 2006, p. 70)

Secondly, a core component of generic feedback loops is the assessment of the current state of the relevant control variables. Hence, the ITF-model assumes an assessment component for both, the internal and external feedback loops. More specifically, it assumes that the assessment of the current state of the control variables can be done by the learner (*internal assessment* – e.g., self-assessing if the line of argumentation follows the rules of scientific inquiry*)* and also by the external feedback source (*external assessment*) and that these assessments result in an *internal* and an *external feedback of the actual state of the learner's competency*.

Thirdly, the ITF-model uses the cybernetic terms *internal and external controller* to emphasize that the information provided by the internal and the external feedback can only be used to regulate the process of competence acquisition if in both loops the feedback information is processed mindfully and if the result of this information processing is used to generate the control action(s): More specifically, students and external feedback sources (e.g., a teacher) have to compare the current state of competencies to the respective desired level of competencies in order to evaluate to what extent the desired level of competence has been achieved (e.g., the feedback reveals that the line of argumentation does meet only some of the scientific inquiry principles; however, the desired level was set at meeting all principles). Furthermore, the result of the comparison has to be used to select adequate control action(s).

Fourthly, based on this comparison and selection, the *external controller* has to generate an *external feedback message.* This external feedback message may simply consist of appraising or confirming that the desired competency state has been achieved, if there is no discrepancy between the desired and the current competency state. However, if a gap between the current and the desired level of competencies has been detected, this external feedback message may provide evaluative information revealing this gap, as well as tutoring information (i.e., all kinds of suggestions for control actions that help closing the gap). It is important to emphasize here that there is a very large variety of feedback components (for an overview of a content-related classification of feedback components, see Narciss, 2008, 2012) that can be combined in manifold ways into feedback messages and strategies.

Fifthly, the learner's *internal controller* has to process the external feedback message along with the internal feedback in order to generate and select control actions. More concretely, the learner has to compare (a) the internal feedback (i.e., the internally assessed current state of competencies) with the desired level of competencies, (b) the external feedback message with the desired level of competencies, and (c) the internal feedback with the external feedback message. Based on the results of these processes, the learner has to generate *internal control action(s)* (e.g., error correction strategies; revision activities) that may help to proceed in the direction of the desired level of competency.

Finally, learners have to select and transmit the *internal control action(s)* to the controlled process, where these control action(s) have to be implemented. If learners have selected and implemented the adequate control action(s), their competencies should improve in the direction of the desired state of competency.

Conditions for Efficient Feedback Strategies

The ITF-model suggests that the efficient regulation of competence acquisition with external feedback may be affected by at least three groups of factors: firstly, factors related to the requirements of learning tasks and the competencies necessary to meet these requirements; secondly, factors related to the internal loop that are individual learner factors; and thirdly, factors related to the external loop (the external feedback source).

Requirements of Learning Tasks: Conditions of Instructional Context

As described above, the controlled process for both feedback loops is the process of acquiring the competencies necessary to master learning tasks in a given instructional context. Depending on the requirements of the learning tasks and the instructional objectives, this process can be more or less complex. For both, the learner and the teacher (i.e., external feedback source), identifying and describing precisely the competencies necessary for meeting the task requirement of complex tasks (e.g., writing an academic essay) is more difficult than doing so for simple tasks (e.g., responding to factual questions).

Yet, in order to regulate a system successfully, it is crucial to describe its controlled process carefully and precisely, since a precise description of competencies provides the basis for (a) identifying the relevant control variables and their respective standards, (b) assessing the current state of these control variables, and (c) generating control actions. In other words, if a precise description of competencies is missing, the student and also an external feedback source (e.g., teacher, peer, instructional medium) cannot know which competencies are needed to master task requirements and thus can neither reliably assess them nor generate feedback or actions for improving these competencies.

Internal Loop Factors: Conditions of the Learner – Feedback Receiver

The ITF-model reveals that several learner factors may promote or constrain how well learners will be able to improve their competencies in the direction of the desired standards if provided with formative feedback messages by an external feedback source. Thus, the ITF-model suggests taking the following learner factors into account when designing and investigating formative feedback strategies (cf. Narciss, 2008, 2013):

- *Learner's representation of standards, competencies, and task requirements*: An important prerequisite for mastering the requirements of learning tasks successfully is to understand and represent the task requirements, the related standards, and competencies. Whether learners are able to understand and represent task requirements adequately and precisely depends on the complexity of these requirements but also on individual factors such as learner's prior level of competencies (i.e., knowledge, metacognitive knowledge and strategies, as well as motivation and volitional skills).
- *Learner's self-assessment skills:* Learners have to monitor and assess their process of learning or competence acquisition in order to generate their internal feedback. This internal feedback can be compared to the learner's desired level of competence and to the external feedback. It may thus serve, at least to some extent, as a frame of reference for the external feedback. Hence, the learner's monitoring or self-assessment skills have to be considered an important factor for the processing of the external feedback.
- *Learner's skills and strategies in information processing:* Learners have to process the internal and external feedback and compare them to their desired level of competencies and generate control actions. Thus, a further learner factor identified by several reviews and meta-analyses of feedback research is learner's mindful processing and integrating information from several sources (e.g., Mory, 2004; Narciss, 2008; Shute, 2008).
- *Learner's will and skills in overcoming errors and obstacles:* As shown in studies on feedback seeking, even the most sophisticated feedback is useless, if learners do not attend to it (e.g., Aleven, Stahl, Schworm, Fischer, & Wallace, 2003) or are not willing to invest time and effort in feedback uptake. Besides the will, students need also the skills necessary for accomplishing the requirements related to feedback uptake.

External Loop Factors: Conditions of the External Feedback Source

The ITF-model reveals also several factors of the external loop that determine if the external feedback will provide valuable information to the learner's loop. According to Narciss (2008, 2013), the following external loop factors may promote or constrain the efficiency of the interplay of the two feedback loops:

- *Quality of the external representation of standards and competencies:* Identifying and describing precisely the competencies necessary for mastering learning task requirements is an important starting point for mostly all processes in the external loop. Based on this description, the external feedback source has to generate an external representation of these competencies and requirements. Furthermore, it has to determine the external level of standards. As in the internal loop, the quality of the representation of task requirements, competencies, and standards

depends, on the one hand, on the complexity of the task requirements. On the other hand, it is determined by characteristics of the external feedback source (e.g., how well the teacher or peer has understood the task requirements and identified the necessary competencies and their standards).

- *Quality of the external assessment:* The accurate and reliable assessment of the learner's current level of competence is a second important factor for generating the external feedback. This includes identifying appropriate indicators for measuring different levels of competencies as well as selecting or developing measurement instruments with regard to these criteria. In case of a human external feedback source, the level of assessment competency has to be considered an important factor influencing how well the assessment meets diagnostic quality criteria.
- *Quality of the external data processing:* Meaningful comparisons between the assessed level of competence and the desired level of competence are only possible if the external feedback source processes the information at hand adequately.
- *Quality of the design and communication of the external feedback:* The quality of the feedback message(s) generated by the external feedback source and how well it is adapted to instructional goals and learner's level of competence is a core factor influencing the power of a feedback strategy (Narciss, 2008; Narciss et al., 2014). If a discrepancy between the desired and the current level of competence is detected, the external feedback source may for example simply elicit this discrepancy for a learner with a high level of competence. In the case of a learner with a low level of competence, however, it should also provide suggestions for control actions that may help to reduce this discrepancy. Generating control actions requires that the external feedback source is able (a) to understand why learners did not master the task requirements at the desired level, (b) to identify which external feedback information may be provided to these learners in order to help them acquire the lacking competences or improve their week competences, and (c) to communicate this external feedback to the learners in such a way that learners will attend to it and process it mindfully.

Implications for Feedback Strategies in HE

As mentioned above, the ITF-model has been originally developed to provide a basis for the design and evaluation of *interactive tutoring feedback strategies* for computer-based learning systems (e.g., Narciss, 2013; Narciss et al., 2014). However, as illustrated by Fig. 12.1 and Table 12.1, it provides also a valuable theoretical framework for deriving principles of how to design feedback strategies for other instructional contexts.

The overview in Table 12.1 describes how existent prescriptive principles for designing feedback strategies are linked to the components and assumptions of the ITF-model (for overviews on these principles see, e.g., Boud & Molloy, 2013;

Table 12.1 Feedback-design principles, their links to the ITF-model, and examples of practice

Feedback-design principle	Rationale: link to ITF-model	Examples of good practices
Distinction of an internal and external feedback loop		
Feedback strategies should be interactive rather than just focusing on transmitting external feedback	The external and internal feedback loops are intertwined and aim at controlling and regulating the same process of task completion or competence acquisition	Ask-tell-ask-act strategy (e.g., French, Colbert, Pien, Dannefer, & Taylor, 2015)
		GROW-strategy (e.g., Whitmore, 2010)
		Interactive tutoring feedback strategy (Narciss, 2008, 2013)
Controlled process – control levels and variables – standards and reference values for control variables		
Identify malleable and measurable variables that may serve as reliable and valid indicators of how well the process of task completion or competence acquisition is currently running	Reliable and valid control variables are a core component of feedback loops. If irrelevant variables are measured and their values are fed back to a system (e.g., a learner), this information will be useless if not harmful	Competence modelling (e.g., Campion et al., 2011)
Select or specify the criteria and standards of high-quality task processing and completion	Feedback can only serve to close a gap between the standards strived for and a current state of learning if there are (clear) criteria and/or standards of how to accomplish task requirements on a high level of performance	Common core standards
		Assessment rubrics for written assignments
		Competency definitions or models
Reflect, share, and discuss the criteria and standards of high-quality task processing and completion within instructional context (e.g., teacher-to-student, peer-to-peer, student-to-teacher)	Students' and teachers' representations of successful task completion serve as a basis for determining the reference value of the control process. Discrepancies between these representations diminish or even eliminate the impact of external feedback	Provide exemplars of diverse quality to make students actively engage in reflecting, sharing, and discussing high vs. low levels of quality criteria
		Provide competency matrices with criteria for various levels, and make students reflect and discuss them
Assessment instruments and processes: internal and external		
Select or develop means/devices (e.g., assessment rubrics for written assignments) for assessing the current state of competency	The quality of assessment instruments and strategies is essential to measure the current state of competency in a reliable and valid way. The reliable and valid measurement of the current state of the control variable is a core prerequisite for generating feedback	Tabular presentation of assessment rubrics with Likert-like response options
		Competency matrices eliciting behavioral descriptors for various competence levels

(continued)

Table 12.1 (continued)

Feedback-design principle	Rationale: link to ITF-model	Examples of good practices
Reflect, share, and discuss the assessment instruments	Transparency with regard to assessment instruments and criteria is crucial for mindful feedback processing	Provide worked examples of diverse quality to clarify if and how the assessment instrument can be applied to measure various levels of competency
		Provide occasions for peer-assessment and/or peer-feedback
Provide occasions for generating internal feedback before offering external feedback	The internal feedback loop is essential with regard to empowering students as self-regulated lifelong learners. Thus, the generation of adequate internal feedback (i.e., accurate self-assessment) needs to be trained. Offering immediately external feedback may hinder students in self-assessing their task processing and in actively developing a grasp of criteria/standards	Prompt self-assessment with regard to the relevant assessment standards and make self-assessment overt
		Provide self-assessment work sheets (e.g., competency matrices, rubric tables) to help students to identify which of the criteria they have met as well as those they have not (yet) met
External assessment	The current state of the control variables needs to be assessed using the selected assessment instruments in order to generate external feedback	Use assessment work sheets (e.g., competency matrices, rubric tables) to identify the current level of competency
Generate and provide external feedback		
Generate external feedback message(s)	The generation of a competency-oriented, concrete, and actionable external feedback message includes	Apply W3-strategy to identify gaps and select or specify control actions that may contribute to close the gaps
	(a) the comparison of the current state of competency to the standards	What worked well?
		What did not work well?
	(b) based on the results of this comparison, the selection and specification of the evaluative as well as tutoring information pointing to the control action(s) necessary to achieve a higher level of competency	What can be done to ...?

(continued)

Table 12.1 (continued)

Feedback-design principle	Rationale: link to ITF-model	Examples of good practices
Provide external feedback that is information on the externally assessed level of performance	The external feedback is an important source of (a) calibrating the internal feedback and (b) deriving control actions	Provide a tabular work sheet detailing standards/criteria as well as external feedback with regard to the criteria (same structure as the internal feedback work sheet)
Internal processing of feedback and generating of control action		
Scaffold students in using the external feedback mindfully for (a) detecting gaps that have to be filled in order to meet the required standards and (b) deriving control actions to close these gaps	Students need to compare the external feedback with their standards strived for and with their internal feedback that is their self-assessed level of performance/competencies in order to identify to what extent their performance meets the required standards. Based on these comparisons, they have to identify which control action(s) would be adequate to close the gap	Provide a tabular work sheet detailing standards/criteria and internal as well as external feedback with regard to the criteria
		Provide guiding questions for comparing standards with external and internal feedback
		Prompt students to make a feedback action plan
		Provide access to tutoring information (e.g., hints; guiding questions, worked examples) that may be used to identify control actions
Reflect, share, and discuss potential control actions for closing gaps between the current level of performance and the standards	Adequate control actions (i.e., corrective actions) are a necessary condition for moving the controlled process in the direction of the standards strived for	Ask students to share and discuss their feedback action plan
		Ask for a revision letter detailing what corrective actions to take in order to close the gaps
		Collaborative revision of writing assignments based on feedback
		Collaborative formulation of concrete suggestions for corrective action
Control action – cyclical processes of acting – assessing – feedback – information processing – controlling		
Offer the occasion for applying the selected corrective actions	Applying the control actions in a further attempt with the task is a necessary condition to reveal if the control action (a) has been correctly selected and implemented and (b) will contribute to close the detected gap.	Provide occasions for revision, another attempt of task completion etc.

(continued)

Table 12.1 (continued)

Feedback-design principle	Rationale: link to ITF-model	Examples of good practices
Prompt student to self-assess and generate internal feedback	Each feedback cycle should start with triggering the internal feedback loop	Provide self-assessment work sheets (e.g., competency matrices, rubric tables) to help students to identify which of the criteria they have met now, what has been improved, and what might still need improvement
Provide external feedback together with tutoring information (i.e., hints, guiding questions, explanations, analogies) to help students select and apply adequate corrective actions if they failed to do so without assistance	As in the first cycle, the external feedback is an important source of (a) calibrating the internal feedback and (b) deriving control actions	Elicit progress and emphasize successful attainments of high-quality standards Emphasize What worked well now What has been improved What can be done to further improve ...
Offer occasions for engaging in corrective actions which apply the tutoring feedback information	As in the first cycle, applying the control actions in a further attempt with the task is crucial to reveal if it will contribute to close the detected gap	Provide occasions for revision, another attempt of task completion etc.

Carless et al., 2011; Narciss, 2012; Nicol &Macfarlane-Dick, 2006; Shute, 2008; Evans, 2013). Moreover, in systematically analyzing the ITF-components with regard to their implications for feedback in HE contexts, it became apparent that for complex tasks such as academic writing, it might be necessary to include interactive "reflect-share-discuss" steps or phases in order to design an effective feedback strategy (see green arrows in Fig. 12.1). Furthermore, the analysis revealed that several specifications and differentiations for some of the existent principles need to be considered (e.g., before criteria/standards can be communicated, control variables that are relevant to meet the standards have to be described in terms of behavioral descriptors).

Implications for Scaling Up Interactive Feedback Strategies in HE

The ITF-model reveals that the effects of an interactive feedback strategy can be manifold depending on the various conditions and factors of the instructional context (e.g., complexity of knowledge domain and learning tasks), the feedback receiver (i.e., learner), the feedback source (e.g., teacher, peer, computer-based

system), as well as the properties of the feedback strategy. In view of this complexity, researchers and practitioners are faced with several challenges when designing and evaluating the effects of interactive feedback strategies (see also Narciss, 2006, 2013). In the following I address those challenges that I consider most relevant for scaling up the use of interactive feedback strategies in HE.

High Complexity of Study Tasks and Competencies Students in HE have to acquire highly complex competencies and by doing so work on highly complex study tasks (e.g., solving engineering tasks, understanding scientific texts, interpreting philosophical texts, diagnosing diseases). As mentioned above, a precise description of competencies and their related standards are crucial conditions for students and teachers to reliably assess current levels of competencies as well as generating feedback and actions that may help to improve the competencies in the direction of the standards. However, for very complex tasks, it might be difficult to describe the competencies and standards with enough precision and detail, and this may be a limiting factor for scaling up the use of interactive feedback strategies for such tasks. A promising way to overcome this limitation might be to analyze best-practice exemplars in order to extract criteria of high quality and then ask students to compare and discuss their own solutions and the worked example with regard to these features (see also Carless, Chaps. 1 and 8, this volume). Denise Whitelock and her colleagues have shown that this approach works also when implemented with modern information technology (e.g., Whitelock, Field, Pulman, Richardson, & Van Labeke, 2014).

Heterogeneity of Learners' Levels of Competency (i.e., Cognitive, Metacognitive, and Motivational Dispositions) Learners' heterogeneous levels of competency are challenging teachers in all fields of education, since they raise the issue of how to tailor instructional means (such as feedback strategies) to the various competency levels. With regard to interactive feedback strategies, learner factors may not only constrain how well students uptake the feedback information but also how well they will be able to assess their own work and generate adequate control actions. Hence, the role of students' levels of competency as well as students' feedback perceptions and/or the ways they are generating and processing feedback warrant investigation in order to gain a differentiated picture of the individual conditions influencing the efficiency of various types of feedback. Such a differentiated picture is necessary for scaling up the depth of implementing interactive feedback strategies.

Characteristics of the Feedback Source (e.g., Teacher, Peer) Feedback sources may differ in their characteristics, including their levels of domain-specific knowledge, diagnostic as well as communication competencies, and motivation in engaging in formative assessment and feedback processes. As mentioned above, these differences may affect the quality of assessment as well as feedback generation and communication. To me three lines of research seem to be of particular interest for scaling up interactive feedback strategies: firstly, research identifying characteristics of the feedback source that have an impact on feedback efficiency (e.g., Brown, Harris, & Harnett, 2012; Strijbos, Narciss, & Duennebier, 2010); secondly, training

studies developing and evaluating training tasks on how to apply feedback strategies effectively (e.g., Voerman, Meijer, Korthagen, & Simons, 2015); and thirdly, studies on factors supporting teachers in developing and applying effective teaching strategies such as interactive feedback strategies (e.g., Kraft & Papay, 2014).

Multiple Design Varieties Feedback messages and strategies, in particular interactive formative feedback strategies, are multidimensional. Accordingly, they can be designed in many different ways. Hence, teachers and learners are provided with choices and flexibility when implementing interactive feedback strategies. Yet, the design varieties make the comparison of evaluation studies challenging, since in many studies in the field of HE the feedback interventions are not described in enough detail (Evans, 2013).

Multiple Effect Levels and Functions External feedback messages and/or interactive feedback strategies may have effects on several levels of the process of competence acquisition. So far feedback in HE contexts has been evaluated with regard to learner reactions (e.g., feedback perceptions), learner behavior (e.g., feedback uptake), or various learning outcomes. However, Evans' review revealed that motivational and metacognitive effects need further investigation (Evans, 2013). Using the ITF-model, further evaluation issues are worth to be raised, including the investigation of effects of interactive feedback strategies on (a) learner's representation of task requirements, (b) their self-assessment accuracy, (c) their understanding of external feedback messages, or (d) their skills in generating control actions, for example, in terms of feedback action plans (see, e.g., Rowe, Chap. 11 and Pitt, this volume, Chap. 10).

Summarizing Conclusions

In developing the ITF-model, I synthesized insights from experimental feedback research and cybernetic views on the core components and processes of a feedback loop to derive specific and differentiated assumptions on the interplay of the internal and external feedback loops in instructional contexts (e.g., the feedback loops of students and teachers in HE). The ITF-model attracts attention to the many components and processes that have to be taken into account in order to align the internal and external feedback loops in a way that they serve the main purpose of assessment for learning – empowering students as self-regulated and productive lifelong learners. So far the ITF-model has served as a heuristic framework for designing and investigating technology-enhanced interactive feedback strategies (Narciss, 2013). Yet, as illustrated by this chapter, the differentiated view of the ITF-model on the interplay of the internal and external feedback loops reveals also links between the state of the art of experimental feedback research and so-called constructivist approaches of feedback practices in HE. By doing so, it provides a theoretical basis for refining existent feedback strategies. Furthermore, it reveals

where and how teachers and students have choices in implementing and combining components and steps of interactive feedback strategies. Finally, in my courses on learning and instruction with teacher education and psychology students, the ITF-model has proven to be a fruitful basis for making teachers and students reflect on principles of effective interactive feedback strategies.

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Psychology, 62*, 148–156.

Boekarts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist, 1*, 100–112.

Boud, D., & Molloy, E. (Eds.). (2013). *Feedback in higher and professional education: Understanding it and doing it well*. London: Routledge.

Brown, G. T., Harris, L. R., & Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education, 28*(7), 968–978.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281.

Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology, 64*(1), 225–262.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120.

French, J. C., Colbert, C. Y., Pien, L. C., Dannefer, E. F., & Taylor, C. A. (2015). Targeted feedback in the milestones era: Utilization of the ask-tell-ask feedback model to promote reflection and self-assessment. *Journal of Surgical Education, 72*(6), 274–279.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Hattie, J. A. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hattie, J. A., & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis, 36*(4), 476–500.

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 745–783). Mahwah, NJ: Lawrence Erlbaum Associates.

Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology, 51*(3), 214–228.

Narciss, S. (2006). *Informatives tutorielles feedback. Entwicklungs- und Evaluations-prinzipien auf der Basis instruktionspsychologischer Erkenntnisse*. Münster, Germany: Waxmann.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). Mahwah, NJ: Lawrence Erlbaum Associates.

Narciss, S. (2012). Feedback strategies. In N. Seel (Ed.), Encyclopedia of the learning sciences, (Vol. F(6)) (pp. 1289–1293). New York: Springer Science & Business Media, LLC.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review, 23*, 7–26 Retrieved from http://greav.ub.edu/der.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training on written subtraction. *Learning and Instruction, 16*, 310–322.

Narciss, S., Schnaubert, L., Andres, E., Eichelmann, A., Goguadze, G., & Sosnovsky, S. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56–76.

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*, 199–218.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4–13.

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

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Strijbos, J. W., Narciss, S., & Duennebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291–303.

Voerman, L., Meijer, P. C., Korthagen, F., & Simons, R. J. (2015). Promoting effective teacher-feedback: From theory to practice through a multiple component trajectory for professional development. *Teachers and Teaching, 21*(8), 990–1009.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies*. Seattle, WA: Hofgrefe & Huber.

Whitelock, D. (2015). Maximising student success with automatic formative feedback for both teachers and students. In *Computer assisted assessment. Research into E-Assessment* (pp. 142–148). Cham, Switzerland: Springer International Publishing.

Whitelock, D., Field, D., Pulman, S., Richardson, J. T., & Van Labeke, N. (2014, March). *Designing and testing visual representations of draft essays for higher education students*. Paper presented at 2nd international workshop on discourse-centric learning analytics, 4th conference on Learning Analytics and Knowledge 2014 (LAK2014), Indianapolis, IN.

Whitmore, J. (2010). *Coaching for performance: Growing human potential and purpose: The principles and practice of coaching and leadership*. London: Nicholas Brealey Publishing.

Wiener, N. (1954). *The human use of human beings: Cybernetics and society*. Oxford, UK: Houghton Mifflin.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.

Part IV
Using Technology to Facilitate Assessment for Learning

Chapter 13
Technology-Enhanced Assessment Feedback

Claire Moscrop and Chris Beaumont

Abstract The rapid development of technology provides both challenges and opportunities for educators. Opportunities because there are new ways of interacting with students and achieving scalability and challenges to ensure the technology is constructively aligned with principles of good practice. To determine whether technology-enhanced feedback is valuable, it is important to evaluate it against established good practice. One framework which effectively integrates good practice principles into a process model is the Dialogic Feedback Cycle (DFC) (Beaumont, O'Doherty, & Shannon, Stud Higher Educ 36(6):1–17, 2011) which considers the feedback process in three stages: preparatory guidance, in-task guidance and performance feedback. A key element of the model is timely formative dialogue with the student about their work to clarify good performance, model self-assessment and enhance motivation. This chapter identifies good practice in assessment feedback and then discusses various forms of technology-enhanced feedback at each stage of the Dialogic Feedback Cycle (DFC) together with their potential for scalability. These include approaches to build students' assessment literacy and engagement in large classes and a novel intelligent tutoring system which conducts extended dialogue with students to develop metacognitive skills.

Introduction

The use of technology in the learning and assessment processes for higher education students is now so common that it is taken for granted. Furthermore, new technology is constantly being developed, some of which may well be useful and some may not. In our view, it is important that technology is evaluated to determine whether it adds value by enhancing students' learning, by which we mean that its use must be constructively aligned with the learning objectives and process.

In this chapter, we focus on a critical part of the learning process, namely, assessment feedback. New technology can be used to deliver traditional feedback,

C. Moscrop (✉) • C. Beaumont
Edge Hill University, Ormskirk, Lancashire, UK
e-mail: Claire.Moscrop@edgehill.ac.uk

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_13

though some of the more exciting aspects are using technology in new ways that were not previously possible. In line with the aims of this book, we will evaluate technological innovations from the perspective of scalability and alignment with recognized good assessment for learning (AfL) practices.

In order to determine whether the technology-enhanced feedback is valuable, it is essential to establish meaningful criteria against which to judge it. These criteria must be derived from well-established principles of good practice encompassed within assessment for learning (AfL). The Dialogic Feedback Cycle (DFC) (Beaumont et al., 2011) is a good model to use because it effectively integrates these principles in a *process* model rather than simply a set of good practices. It considers feedback in three stages: *preparatory guidance, in-task guidance* and finally *performance feedback*. It therefore provides an analytical framework for evaluating where in the feedback process technology can be applied.

This chapter firstly outlines recognized good practice in assessment feedback and the overall features of the DFC to establish an evaluation framework. It subsequently discusses each stage of the DFC in sequence and evaluates various forms of technology-enhanced feedback against each stage, discussing how best to incorporate them. Established approaches such as technology enhanced peer assessment and more recent examples of audio and screencast feedback are discussed together with real-time approaches such as audience response systems.

Assessment Feedback in Higher Education

Assessment is the key activity which defines the curriculum in students' eyes (Ramsden, 2003) and has a major influence on their learning (Biggs, 2003). However, assessment alone is not sufficient for effective learning (Gibbs & Simpson, 2004), and it is generally accepted that constructive feedback is essential for improving performance (Shute, 2008). Indeed Laurillard (2002, p. 55) claims that 'action without feedback is completely unproductive for a learner'. Hattie and Timperley's (2007, p. 84) analysis shows that high-quality feedback is an extremely powerful influence on student achievement with an effect size of 0.79. They point out that 'the most effective forms of feedback provide cues or reinforcement to learners; in the form of video-, audio, or computer-assisted instructional feedback; and/or relate to goals'. In the higher education setting, Hounsell (2007) states that feedback can enhance learning in three significant ways: by accelerating learning, by optimizing the quality of what is learned and by raising individual and collective attainment.

Despite this central and vital role in teaching and learning, assessment feedback has also become a major concern of higher education institutions in recent years since it has consistently emerged as the least satisfactory aspect of student experience in the UK National Student Surveys (Nicol, 2013).

Research has yielded plenty of principles for good practice in assessment and feedback (Hattie & Timperley, 2007, Shute, 2008). For example, Nicol and

Macfarlane-Dick (2006) identify that good feedback practice helps clarify what good performance is (goals, criteria, expected standards), facilitates the development of self-assessment (reflection) in learning, delivers high-quality information to students about their learning, encourages teacher and peer dialogue, encourages positive motivational beliefs and provides opportunities to close the gap between current and desired performance. Gibbs and Simpson (2004) add that feedback is provided both often enough and in enough detail, that feedback is timely in that it is received by students whilst it still matters to them and in time for them to pay attention to further learning or receive further assistance, that feedback is appropriate to the purpose of the assignment and to its criteria for success and critically that feedback is received and acted upon by the student.

Whilst these are undeniably useful, they do not conceptualize feedback in any systemic way within the assessment process; it is left to the teacher to make sense of how to incorporate them in their particular context. A recent shift in the conceptualization of feedback is described by Molloy and Boud (2013) who argue that students should not be passive recipients of teachers' comments (transmission model), but they should be central to the feedback process, taking an active role. Such activity involves dialogue to enable students to explore, clarify and internalize assessment criteria and standards – a possible route to self-evaluative expertise and the holy grail of 'sustainable feedback' (Sadler, 2010). This dialogue can be with peers or teachers so long as it is focused on future improvement. Burke (2011) attacks the problem through a structured form-based approach to promote effective dialogue between the tutor and student around the feedback which has been provided on an assessed task, an approach which formalizes and enforces reflection and conversation with an expert in order to stimulate students' engagement with and action upon the guidance provided.

Theories and frameworks that connect these concepts in a systemic way are thin on the ground. One recent approach which models assessment feedback as a guidance system, known as the DFC (Fig. 13.1), is provided by Beaumont et al. (2011). This is particularly appropriate as it incorporates the recent ideas of formative dialogue, engagement with criteria and student activity. It meets Leese's (2010) call for structured activities and more academic support and models the feedback process as three stages, each of which is also shown as a cycle, emphasizing the importance of iterative dialogue, principally between teacher and student. The name acknowledges that this model enacts several of the principles cited above by emphasizing iterative and formative discussion at each of the stages as students work through an assessed coursework task.

It also addresses the criticism that courses are characteristically 'end loaded' with summative feedback that is often irrelevant or too late to be of practical use (Hounsell, 2007) since it emphasizes the guidance provided during the process of assessment.

Furthermore, the DFC is empirically based on a qualitative study of 176 students in six UK schools and further education colleges and three universities. It provides a distillation of the good practice in the systemic guidance that the students experienced in schools and colleges. It is worth emphasizing that there was a stark

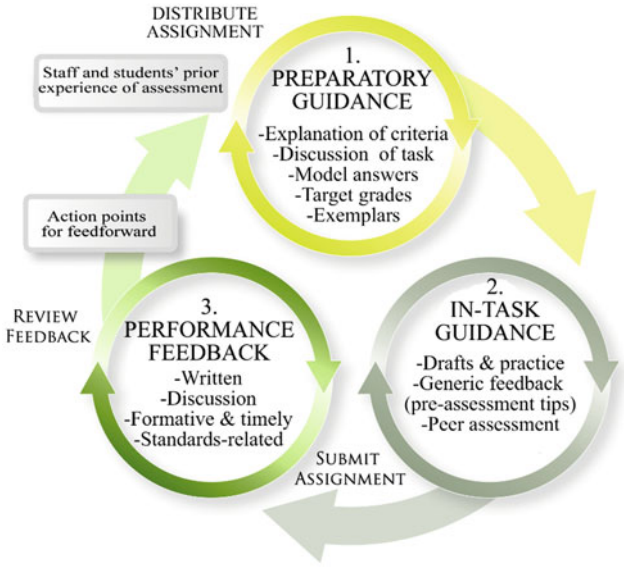

Fig. 13.1 Dialogic Feedback Cycle (DFC) (Beaumont et al., 2011)

difference in the perceived quality and quantity of feedback that students experienced in college or school compared with paucity in HE (Beaumont et al., 2011).

Beaumont, Canning and Moscrop (2016) tested the model by applying it to a first year university module. Their intervention revised the feedback methods employed in the module and included activities for students to develop self-assessment abilities (Sadler, 2010) by applying assessment criteria to exemplars. Students also received relatively rapid (1-week) feedback on a draft assignment and engaged in discussion at all stages, including 1:1 sessions on their summative work.

Statistically significant improvements ($p < 0.05$, effect size 0.33) were achieved in several important aspects. The focus on grading criteria yielded better understanding of the criteria and what was expected in the assignment, improving students' assessment literacy. Furthermore, 84 % of students agreed that feedback was timely enough to be useful, it helped their learning, and they used it to prepare for the next assignment. A pleasing result is that students were more willing to ask tutors for help and they had gained confidence to plan, research and apply knowledge in a more independent way. These were clear indicators of development in self-regulated learning abilities, and we concluded that adopting the DFC as a model was effective in those respects – the improved levels of confidence show a more successful transition to the 'alien world' of HE.

Additionally, the DFC mirrors Zimmerman and Schunk's (2008) model of self-regulated learning, which consists of three phases: forethought, performance and self-reflection. Zimmerman and Schunk claim that learners who engage in high-quality forethought are more effective at self-regulated learning. Forethought occurs at the preparatory guidance stage of the DFC, including goal setting and task

analysis; thus, in seeking to develop self-regulated, rather than dependent learners, it makes sense to focus more attention at the start of the DFC. Each stage is also represented as a cycle to emphasize the iterative dialogue that students often highlighted.

A criticism of this model is that it was derived from an analysis of face-to-face classroom practice and that a dialogic approach between tutor and student is a challenge as classes get larger. However, we would argue that the model itself is independent of delivery method and that it focuses on principles of feedback which are appropriate in any setting. Furthermore, we will show that scalability can be achieved by increasing the role of students in (peer) dialogue, which can be effective if students are suitably prepared.

Thus, we are confident that the DFC encompasses currently perceived good practice in assessment feedback, providing much more than a checklist, since it integrates activities into a system. We therefore consider it as a suitable framework against which to evaluate technology-enhanced learning (TEL) feedback. Having established the basis for evaluation, in the rest of this chapter, we discuss how these principles can be achieved in a scalable way using technology. Given the plethora of technology available, some practical examples will be given as a guide to what is appropriate to support each of the phases of the DFC when managing larger groups.

We structure what follows in the same sequence as the stages of the DFC, drawing on our own research and carefully selected examples from the literature to examine how they support feedback at each stage.

Preparatory Guidance Stage

In this initial phase of the cycle, activities are included to engage students with the grading criteria and gain a good understanding of the task and a plan for successfully completing it.

A useful approach within a classroom environment has been to present students with exemplars which they assess against given criteria. Whilst this has been shown to be an effective approach to enhance students' understanding of assessment requirements (Beaumont et al., 2016), the student to student or student to tutor dialogue required can be a time-consuming approach for larger cohort sizes.

One approach to this scalability issue is screen casting. Screen casting can be used to provide the tutor's perspective of assessing an exemplar against criteria with an aim to develop a shared cohort understanding of the criteria and the assessment process (Low & Soden, 2011; JISC, 2010). The benefits of this method are that it gives students the opportunity to revisit and reuse the resource at their point of need, during the writing of their own assessment, on and off campus.

Another innovative approach to creating a shared understanding of assessment through technology that can also be applied to large cohorts is the use of real-time sharing and analysis of assessment descriptors. Knight (2015) used real-time collaborative software (in this case, GoogleDocs) to present a given assessment in a lecture

to 160 students. Students were encouraged to read the assessment and associated criteria, discuss with peers and update and annotate the assessment document in real time. Knight reported that this real-time analysis gave an immediate appreciation of the areas requiring clarification, allowing him to present an updated assessment and marking criteria the following week. He also reported a greater student 'ownership' of the assessment.

The importance of presenting clear requirements at the point of need was emphasized in a study by Moscrop and Canning (2015), where they introduced the use of 'digital assessment guides' (DAG) to scaffold learners' understanding of assessment requirements. The DAG is an audio over PowerPoint iSpring file that talks the students through the assessment in a very structured way.

Black and Wiliam (1998) suggest that students must be supported in closing the gap between their knowledge and what is expected in an assessment task. Moscrop and Canning (2015) noted that many students took little away from initial assessment descriptions in class. It was apparent from the student comments that there was an issue with remembering the initial explanation of the assessment given by the tutor, with comments such as 'I knew as soon as I got home what you said would be gone'. Students often repeatedly asked for guidance already covered in other sessions and in written resources available on the VLE. The DAG was presented with the intention of reducing the students' extraneous cognitive load as it integrated the assessment guide and tutor description of the expected deliverables (Chandler & Sweller, 2009). The clear structure allowed students to skip to the pertinent sections at their point of need, meaning they could listen to the tutor's explanation of the requirements and assessment criteria for the particular section they were working on, revisiting it exactly when they needed it.

If we consider the initial explanation given to students in week 1 regarding assessment outcomes and expectations, it is often very difficult for them effectively to understand and apply that explanation as they do not yet have the requisite knowledge to build a clear mental picture of the requirements. For example, one of the modules DAG was introduced into was a third year project management module, with technical terms such as 'PRINCE2' (a project management methodology), 'Product Breakdown Structures' and 'Product Descriptions' all described as requirements for the assessment. Unless you are already familiar with project management, these terms are meaningless – and it is the same for students: it is not until they attend the appropriate seminar that they start to understand what the assignment means in detail.

Stewart (2012) notes that the role of the teacher is that of a guide, providing 'scaffolding' of learning to ensure that the student has the requisite knowledge, skills and support to negotiate a new piece of learning. The main benefit felt by the students using DAGs appears to be tied to the 'requisite knowledge' element of the above statement from Stewart (2012). Therefore, the DAG provides the scaffolding necessary to allow students to access these detailed descriptions at a point when they have the most understanding of the necessary concepts.

The result of this intervention was a doubling of the percentage of students achieving a mark above 70 %, increased student satisfaction with the assessment process and a reduced workload for tutors in terms of dealing with student questions by email and face to face (Moscrop & Canning, 2015). The benefit of this use of technology is that it is entirely scalable to any cohort size.

In-Task Guidance Stage

The *in-task guidance* phase of the DFC is about supporting students to develop their ideas, clarify misconceptions and analyse the problem and data to help them identify a suitable approach to completing the work. Depending on the task, it can also assist them with their academic writing. Typical activities at this stage comprise formative assessment via draft submissions and peer assessment. Informal dialogue with students during class is also beneficial.

The opportunity to submit drafts has increasingly been regarded as important – and almost seen as a 'right' by students in recent years – many students really value feedback which indicates that they are 'on the right track', though it is essential to set clear expectations since it is impractical to provide the same level of detail for drafts as the final summative submission (Beaumont et al., 2011).

Peer assessment can be problematic unless students are well prepared: Beaumont et al. (2011) report that whilst some students find it constructive and motivational, others did not engage deeply. Often, we have experienced student criticism such as *it is not my job.* Other complaints are that they don't have the expertise to assess or they don't trust the peer's assessment. These comments reveal two vital issues that need to be addressed if peer assessment is to be implemented effectively: firstly, students need to believe that it is part of their role and acknowledge the benefits that they receive by giving and receiving feedback. This is a significant culture shift, about the role of students in higher education. The second criticism is equally valid and requires effective and sustained training of students together with the very careful design of assessment to ensure students have the requisite knowledge to assess reliably.

Peer assessment, if implemented well, can both improve student learning and can be scaled up. The challenge is therefore to ensure students experience benefit and perceive value. For example, we have used a Wiki to promote peer feedback on draft sections of dissertations, which (after some initial resistance) successfully led to active critical comments being provided on peer scripts (Su & Beaumont, 2010).

However, other activities in this stage of the DFC (tutor marking of drafts and tutor-student dialogue) present difficulties since the workload for the tutor increases proportionately with the number of students. In response to this challenge, we have explored a number of approaches: intelligent tutoring systems (ITS), automated online assessment and the use of audience response systems.

Dialogic Feedback from an Intelligent Tutoring System

There are a huge variety of ITS which are typically special purpose computer-assisted learning systems that present learning material, tests and automated feedback. A good system is adaptive to a student's needs and often employs artificial intelligence techniques.

Beaumont et al. (2011) describe an ITS which was designed to provide extended (text) dialogue with students. The prototype that was developed helps students to analyse and clarify assessment tasks and plan reports. The ITS, known as a learning coach (Fig. 13.2), was designed to be employed in the early stages of a problem-based learning scenario. It was constructed using chatbot technology which provides dialogue with students in order to guide them through the initial stages of developing personal learning objectives. User testing with undergraduates suggests that the system prompted them to analyse the scenario in more detail (one of the foundational skills needed in a PBL task) and that the technology can be both usable and provide the adaptability required.

The coach can operate in one of two modes: a 'quick consultation' or a 'full consultation'. In a quick consultation, the coach provides answers to students' questions (including a set of FAQs). These questions were derived from the primary study of students' concerns, for example, 'What should I do next?' and 'When is XXX due?' relating to clarification of the task and scenario and some help on monitoring progress.

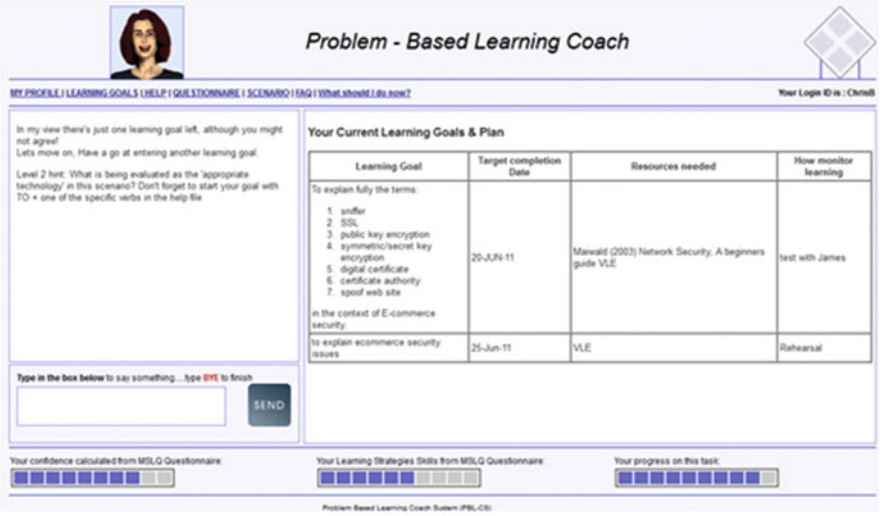

Fig. 13.2 Learning coach interface

In the 'full consultation' mode, the coach takes a more proactive approach by guiding students through the analysis of the scenario and prompting them to plan the structure of a report and create well-formed learning goals. Well-formed goals are defined as those which are specific enough to be achievable and include the outcome, context and criterion for achievement. The coach also prompts students to identify the learning resources needed to achieve the goal, set a completion date and state how they will monitor the completion of the goal. Learning goals and progress are stored in the learner model and can be retrieved and printed by the student.

Thus, the learning coach helped to scaffold some self-regulated learning (analysis and planning) activities. It also included a diagnostic questionnaire which is a self-report measure of SRL. It consisted of an abbreviated form (42 questions) of the Motivation Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1991) together with a help file to provide guidance on learning strategies and writing well-formed learning objectives. Students can view and explore their profile produced from their questionnaire results. In the study reported here, the self-efficacy and control of learning beliefs measures predict a 'confidence' value which determines the level of guidance provided (higher confidence, less specific guidance).

As is common in an ITS, the coach provides hints to assist students. The coach provides three levels of hint, with higher levels of guidance providing more detail. The level given depends on the students' initial confidence value and dynamically adapts depending on the quality of a student's answers throughout the dialogue. This aspect of adaptation was considered to be one of the most important design features in order to customize the guidance to students' individual needs.

Beaumont et al. (2011) reported that their results suggest that chatbot technology can be used to create a simple adaptive guidance system that assists students effectively to analyse a scenario and plan their learning objectives. In particular the tests from a small sample showed positive affective influence on all participants, and students self-reported that it helped them analyse the scenario:

> It made me think more ... not letting me miss anything out, ... making me find from the scenario what I need to put in.

> It made me see which terms I didn't fully understand, ... I wouldn't have looked at those at the start.

These are not comments we would expect from a surface or superficial approach and suggest that the technology has the potential for significant learning benefits for students by engaging them in relevant and extended dialogue with tasks.

The investment in time required to construct such a system is high, even with the assistance of chatbot systems. However, for large classes, this could be worthwhile, and since the systems provide a web interface, they are naturally scalable. The current system requires written communication, but as speech recognition is increasingly becoming a common interaction with computers, for example, Microsoft's Cortana, there are exciting possibilities for this approach.

Audience Response Systems

A further method of in-task guidance can be provided by audience response systems (ARS) and electronic voting systems (EVS). These are now common in higher education with a recent shift from hardware 'clicker' systems to the use of students' own smartphones or tablets.

Strawson (2013) highlights the fact that, even in small lectures, students are generally reluctant to ask questions, yet research shows that students' retention of information, deep learning and problem-solving skills increase when they have opportunities to ask questions (Harper, Etkina, & Lin, 2003; King, 1989, 1991).

Kay and LeSage (2009) suggest that the introduction of audience response systems effectively promotes student interaction in lectures and have three areas of benefit:

1. Classroom environmental benefits (improved attendance, attention, anonymity, participation and engagement)
2. Learning benefits (interaction, discussion, contingent teaching, learning performance and quality of teaching)
3. Assessment benefits (formative feedback)

Moscrop (2015) explored if these benefits were experienced with a new generation of smartphone ARS apps. She introduced a whole cohort of first year undergraduate computing students ($n = 160$) to the Socrative app – a free audience response application that students can download or access directly through the Socrative website (Socrative.com). Socrative was used during every lecture over 10 weeks of the first semester. The application was used to punctuate lectures by asking students to discuss and answer questions. The method of student reply to these questions was varied between open-ended answers, yes/no or true/false questions and a choice between a number of options (e.g. A–D). In order to further increase engagement, students were asked to discuss their answers in small groups. Selected answers were then shared with the students. The key outcomes of this study were:

- Students had a generally negative perception of 'traditional' lectures due to issues such as disengagement and distractions caused by other students. The quantitative whole cohort survey noted that 82 % preferred lectures using Socrative to traditional lectures.
- Students strongly linked the themes of 'engagement' and 'interactivity'. Students indicated clearly that the interactivity created by Socrative improved their lecture experience and engagement, with 30 instances of positive comments relating to the improved engagement theme in the qualitative survey data and 19 instances in the focus group data. Some examples were 'It adds an extra level of interaction which is harder to simulate in most lectures' and 'It is good to be interactive and get involved during lectures as it helps me learn'. This was corroborated by survey data with 73 % of students agreeing that they felt more engaged by the Socrative lectures (7 % disagreed) and 76 % agreeing that the fact that they may be asked questions using Socrative made them more attentive (5 % disagreed).

- Anonymity was also a key factor highlighted by students, with comments like 'It's anonymous, as well, so you haven't got that.....fear' and 'We all have questions but we're not going to ask those questions, you know, because you don't want to be labelled as being stupid'.
- Students noted that they also felt they had improved understanding and recall of the lecture content: 'I found it helped me retain information more effectively, by testing our knowledge immediately after learning it' (Moscrop, 2015).

These results supported the previously stated benefits of ARSs from Kay and Le Sage's (2009) review. It is likely that the cost of EVS/ARS will continue the shift towards BYOD for such interventions.

Performance Feedback Stage

The final stage of the DFC is usually delivered in both written and verbal forms, again providing opportunity for dialogue which focuses on explanation related to criteria and planning for future improvement.

In our research (Beaumont et al., 2011), we found that students express a strong desire to receive grades/marks together with personalized feedback comments. A further theme of reassurance and motivation also permeated the responses of students and teachers, demonstrating a strong, shared awareness of the power and impact on self-esteem that assessment and feedback can have. Feedback on drafts was often reported to be attended to by students and seen as critical; however, students mentioned action planning as a result of post-assessment feedback very rarely.

Whilst technology can assist geographical scalability for this stage of the DFC, it can be difficult to scale up to large numbers of students, since it is largely dependent on tutors. Speed is therefore a key concern when managing large cohort feedback. We have found that rubrics and standardized comments within tools such as Turnitin/Grademark (Turnitin, 2015) can help by enabling the tutor to create a bank of commonly written comments, whilst still allowing tailored comments if required. However, it is worth pointing out that there is a trade-off between standardized comments and personalized feedback.

An alternative – or supplementary – approach uses audio and video feedback. Studies over several years have shown that such feedback can *speed* delivery after sufficient practice and improve student satisfaction (Dixon, 2015; Rotheram, 2009; Stewart, 2008). A key benefit identified by Knauf (2016) was that students found it more *personal*, and some reported being able to assimilate it better. Tools such as Grademark now have integrated audio feedback, meaning staff can record and save comments efficiently when marking a script.

Another technology 'fix' for dealing with large automated essay scoring (AES) applications has been around since the 1970s and arguably reached commercial viability in the 1990s by being indistinguishable from human evaluators for short

essays with a specific focus, though the validity of their results is questioned for more complex tasks. Despite these concerns, EdX have introduced an AES application that will integrate within its MOOCs (Balfour, 2013).

This final step of the DFC is also about closing of the feedback loop or the 'feedforward' process. The continuous assessment example REAP project mentioned earlier demonstrates how students can benefit from continuous assessment in a module, using feedback throughout each stage to enhance the next assessment (Nicol, 2009). However, feedforward in a modular system can be problematic as students often perceive no links between modules, (Hepplestone, Holden, Irwin, Parkin, & Thorpe, 2011).

Guidance from the national union of students notes outstanding practice to be an assessment process that means 'Students' personal development takes account of all the feedback they have received throughout their course' (NUS, 2015, p. 1). Despite this exhortation of the value of feedforward, tools to enable students and tutors manage this process have been slow to develop, though the Moodle VLE now has a number of plugins that enable collection of feedback from across modules.

Conclusion

The importance of feedback to help students learn has been known for a long time, and principles of effective feedback are now well established. However, the processes of providing relevant feedback at an appropriate time and ensuring that students engage with it and act on it are more problematic. This chapter has presented a process model (the DFC) which is a framework for employing feedback as formative guidance throughout an assessment task. It responds to concerns raised by researchers that students neither understand what feedback is, not how to use it. A key component of this approach is dialogue – students engaging in discussion at all stages, so that they fully understand what is required and that they receive scaffolding guidance of how to use it.

In this chapter, we have provided examples of how the feedback process is assisted by technology including the digital assessment guides, intelligent tutoring systems and the use of apps such as Socrative for large lecture groups. Technology-enhanced feedback can assist with geographical scalability and in some cases with large-scale cohorts.

However, scalability in many of these processes is problematic so long as the tutor is seen as the focal point – providing the feedback and engaging each student in the dialogue. In our view, a more fundamental change is required, so that formative peer feedback becomes the norm. Not only does this address the scalability issue, but it promotes sound practices for learning based on social constructivism, where students develop concepts and meaning through social interaction (Uden & Beaumont, 2006). In order to achieve this, some fundamental changes are needed: Students need to believe that this is valuable – a change of attitude and culture for many – and they need to be trained to be able to provide effective feedback.

Such a radical approach also needs to be used carefully; tutors need to design assessment tasks in a way that students can provide valid formative feedback at suitable points and also design ways of checking the quality of the feedback so that students improve (feedback on feedback!). Winstone and Nash (2016) attack part of the problem by providing a toolkit of training materials to help students develop their assessment literacy which includes a guide to making the best use of feedback, portfolio/action planning tools and crucially a workshop to ensure students engage with it.

Our suggestion is also a major culture change for many tutors; indeed Gray and Ferrell (2013) identify that many projects experience staff resistance to change when implementing online feedback practices. However, whilst the implications of scalability are clearly a major challenge, they also provide a tremendous opportunity and driver to change the culture of higher education for the better.

References

Balfour, S. P. (2013). Assessing writing in MOOCS: Automated essay scoring and calibrated peer review. *Research & Practice in Assessment, 8*(1), 40–48.

Beaumont, C., Canning, S., & Moscrop, C. (2016). Easing the transition from school to HE: Scaffolding the development of self-regulated learning through a dialogic approach to feedback. *Journal of Further and Higher Education, 40*(3), 331–350.

Beaumont, C., Norton, L., & Tawfik, H. (2011). Intelligent tutoring systems: How well can they guide students in problem-based learning scenarios? In J. Davies, E. de Graaff, & A. Kolmos (Eds.), *PBL across the disciplines: Research into best practice*. Aalborg, Denmark: Aalborg Universitetsforlag.

Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? Studies in Higher Education, 36(6), 1–17.

Biggs, J. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham, UK: Society for Research into Higher Education and Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–78.

Burke, D. (2011). Now I've got the feedback, what do I do with it? Strategies for students to get more out of tutor feedback. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education, 6*(1), 43–60.

Chandler, P., & Sweller, J. (2009). Cognitive load theory and the format of instruction. *Journal of Cognition and Instruction, 8*(4), 293–332.

Dixon, S. (2015). The pastoral potential of audio feedback: A review of the literature. *Pastoral Care in Education, 33*(2), 96–104.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*(1), 3–31.

Gray, L, & Ferrell, G. (2013). *Changing assessment and feedback practice: How to approach large-scale change in assessment and feedback practice with the help of technology*. Retrieved from https://www.jisc.ac.uk/guides/changing-assessment-and-feedback-practice

Harper, K., Etkina, E., & Lin, Y. (2003). Encouraging and analysing student questions in a large physics course: Meaningful patterns for instructors. *Journal of Research in Science Teaching, 40*, 776–791.

Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

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.

Hounsell, D. (2007). Towards more sustainable feedback to students. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 101–113). London: Routledge.

Joint Information Systems Committee (JISC). (2010). *Effective assessment in a digital age*. Retrieved from http://www.webarchive.org.uk/wayback/archive/20140614115719/http://www.jisc.ac.uk/media/documents/programmes/elearning/digiassass_eada.pdf

Kay, R., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Journal of Computers & Education, 53*(3), 819–827.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology, 14*, 366–381.

King, A. (1991). Effects of training in strategic questioning on children's problem solving performance. *Journal of Educational Psychology, 83*, 307–317.

Knauf, H. (2016). Reading, listening and feeling: Audio feedback as a component of an inclusive learning culture at universities. *Assessment & Evaluation in Higher Education, 41*(3), 442–449.

Knight, C. (2015, June). *160 Students go into a room: Using real time collaborative software for assessment design and feedforward*. Paper presented at the SOLSTICE conference, Edge Hill University, UK.

Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies* (2nd ed.). London: Routledge.

Leese, M. (2010). Bridging the gap: Supporting student transition into higher education. *Journal of Further and Higher Education, 34*(2), 235–251.

Low, G., & Soden, B. (2011, May). *Generating complex formative feedback: Using blogs and screen casts to promote dialogue around content and skills*. Paper presented at University of York annual learning and teaching conference 2011, York, UK. Retrieved from https://www.york.ac.uk/media/staffhome/learningandteaching/documents/soden_low.pdf

Molloy, E., & Boud, D. (2013). Changing conceptions of feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 11–33). Abingdon, UK: Routledge.

Moscrop, C. (2015, July). *Improving student engagement in lectures through the use of free mobile applications.* Paper presented at Edulearn15: 7th international conference on education and new learning technologies, Barcelona, Spain. Retrieved from https://library.iated.org/view/MOSCROP2015IMP

Moscrop, C., & Canning, S. (2015, June). *iSpring: Digital assessment guides: Practical examples of use of this e-learning toolkit to enhance student experiences*. Paper presented at SOLSTICE conference, Edge Hill University.

National Union of Students (NUS). (2015). *Assessment and feedback benchmarking tool. NUS Connect*. Retrieved from http://www.nusconnect.org.uk/resources/assessment-and-feedback-benchmarking-tool

Nicol, D. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education, 34*(3), 335–352.

Nicol, D. (2013). Resituating feedback from the reactive to the proactive. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 34–49). Abingdon, UK: Routledge.

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.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: The University of Michigan.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London: Routledge.

Rotheram, B. (2009). *Sounds good: Quicker, better assessment using audio feedback.* Retrieved from http://www.webarchive.org.uk/wayback/archive/20140614213902/http://www.jisc.ac.uk/publications/reports/2009/soundsgoodfinalreport.aspx

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35*(5), 535–550.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Su, F., & Beaumont, C. (2010). Evaluating the use of a wiki for collaborative learning. *Innovations in Education and Teaching International, 47*(4), 417–431.

Stewart, M. (2012). Understanding learning: Theories and critique. In L. Hunt & D. Chalmers (Eds.), *University teaching in focus: A learning-centered approach* (pp. 3–20). Melbourne, Australia: ACER Press.

Stewart, W. (2008). *Audio supported enhanced learning*. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/usersandinnovation/aselplan.pdf

Strawson, H. (2013). *53 ways to deal with large classes*. Suffolk, UK: Professional and Higher Partnership Ltd..

Turnitin. (2015). *Online grading*. Retrieved from http://turnitin.com/en_us/what-we-offer/online-grading

Uden, L., & Beaumont, C. (2006). *Technology and PBL*. Hershey, PA: Information Science Publishing.

Winstone, N., & Nash, R. (2016). *A resource toolkit to support STEM students in making use of their assessment feedback*. Nottingham, UK: HEA STEM conference.

Zimmerman, B. J., & Schunk, D. H. (2008). Motivation, an essential dimension of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation & self-regulated learning: Theory, research & applications* (pp. 1–30). Mahwah, NJ: Erlbaum.

Chapter 14
How Does Technology Enable Scaling Up Assessment for Learning?

Phillip Dawson and Michael Henderson

Abstract This chapter brings recent critical thought from the field of educational technology to bear on the challenge of scaling up Assessment for Learning (AfL). Three different types of 'scaling up' are presented, illustrated through three different 'technology-enhanced' AfL approaches. Recent advances in providing feedback through audio, video and screencast technologies are used to explore 'doing more with less' as a form of scaling up. Technology enables providing more and richer feedback information while requiring less staff time – but it remains unclear if this results in better learning or just better student experience. Technology's ability to scale up our thinking from individual tasks up to programme level matters is explored through portfolios and curriculum mapping tools. Although these tools provide affordances for programmatic thinking, implementing these thoughts in the complex social environment of higher education presents its own challenges. Finally, scaling up AfL to serve large cohorts without linearly scaling up resources like teacher time is explored through Massive Open Online Courses (MOOCs). However, given the low completion rates in MOOCs, we question if access to AfL is the same as real AfL opportunity. The chapter concludes with implications for scaling up AfL that have been synthesized from these illustrative examples.

Introduction

There is much hype around the potential for technology to enhance assessment, including how it can enable the scaling up of Assessment for Learning (AfL). At the time of writing, technologies that enable new possibilities for assessment are prominent in two key hype barometers: the 'Hype Cycle for Education' (University of Minnesota, 2016) and the New Media Consortium's Horizon Report for Higher Education (Johnson et al., 2016). Technology can support scaling up of educational

P. Dawson (✉)
Deakin University, Burwood, VIC, Australia
e-mail: p.dawson@deakin.edu.au

M. Henderson
Faculty of Education, Monash University, Melbourne, VIC, Australia

D. Carless et al. (eds.), *Scaling up Assessment for Learning in Higher Education*,
The Enabling Power of Assessment 5, DOI 10.1007/978-981-10-3045-1_14

practice in a range of ways. It can allow us to do 'more with less', such as provide video feedback to students who traditionally received written comments, without requiring more staff time. Technology can enable us to scale up our thinking from single units or modules towards approaches like portfolios and curriculum maps that require thinking at the course or programme level. Technology can also enable near-infinite scaling up of AfL through approaches like Massive Open Online Courses (MOOCs).

Given all this potential, it is understandable that educational technology can evoke excitement, enthusiasm and even evangelism. 'Technological determinism' is the notion that technology in and of itself can change education (Oliver, 2011). It puts technology in the driver's seat with pedagogy its passenger. This is part of the 'positive project' of educational technology: an underlying belief that technology is a good thing for education (Selwyn, 2011). Phrases like 'technology-enhanced assessment for learning' carry with them the suggestion that technology can, will or does enhance education.

But technology does not always live up to the hype. In their extensive review of educational technology research, Tamim, Bernard, Borokhovski, Abrami and Schmid (2011) begin by retelling Thomas Edison's 1913 claims that the motion picture would very soon make books obsolete. This has been a recurring theme in education, in which emerging technologies are hailed with great fanfare but have at best only a modest impact on educational practice, supported by largely ungeneralizable research. An example is that of virtual worlds such as Second Life, which has been claimed to support experiential and situated learning including continuous cycles of feedback (e.g. see Dawley & Dede, 2014). However, while these studies argue that virtual worlds can improve feedback, they also confirm Warburton's (2009) observation that technology-enhanced AfL is unlikely to be a 'quick win' but rather a result of considerable risk mitigation and pedagogical strategic planning. The focus on the potentiality or 'state-of-the-art' uses of digital technologies in education has largely obscured the compromised and constrained 'realities' (Selwyn, 2010). Indeed, Laurillard (2008) has wryly observed, 'education is on the brink of being transformed through learning technologies; however, it has been on that brink for some decades now' (p.1). This arguably blinkered perspective, focusing on the 'potential' for technology to 'enhance' and 'provide opportunities', can also be also be found throughout the research literature relating technology-enhanced AfL (e.g. Gikandi, Morrow & Davis, 2011).

This chapter brings a critical perspective to the question of how technology may support the scaling up of Assessment for Learning in higher education. We make particular reference to the core AfL strategies from Carless, Chap. 1, this volume: productive assessment task design, effective feedback processes, developing student understanding of the nature of quality and students practising making judgments. Through synthesis of the literature on three sites of technology-supported AfL – feedback, programme-level portfolios and MOOCs – we explore issues of scale, context and unintended consequences. Although technology can allow us to do more AfL with less, across more curricula, and for more students, we have reason to be cautious and sceptical.

Technology-Enabled Assessment for Learning

The broad field of educational technology has been critiqued as being so obsessed with the 'state of the art' that it misses out on what actually happens in students' and teachers' lives, that is, the 'state of the actual' (Selwyn, 2010). What is the 'state of the actual' of technology in AfL? Despite the bright potential of assessment technologies, their wide-scale adoption has been much slower than technology advocates expected (Warburton, 2009). Even the research literature, which is dominated by intervention studies conducted by researchers (Stödberg, 2012), demonstrates slow progress. Stödberg's (2012) structured review of the literature indicates that the typical technology-supported assessment study is small scale, short term and focused on a multiple-choice intervention. When compared with 'state-of-the-art' approaches, like high-fidelity simulation or intelligent tutoring systems, it could be easy to get disheartened by the slow progress towards technology enablement of assessment for learning at scale.

However, assessment is notorious for being resistant to change. Assessment at universities is a complex system with many actors and a slew of policy and bureaucracy (Macdonald & Joughin, 2009). Within this context, educators and students are engaged in a range of (dis)trusting relationships (Carless, 2009), with risks and anxieties (Deeley & Bovill, 2016). Resistance to assessment change is powerful (Deneen & Boud, 2013), as is resistance to educational technology change (Blin & Munro, 2008). Technology-enabled AfL sits at the intersection of this resistance, so it is perhaps unsurprising that progress has been somewhat slow.

One area where large gains have been made has been the adoption of online submission and return of assignments. In an Australian context paper, submission of assignments is becoming uncommon. Although mundane, the online submission and return of assignments enables a range of assessment practices in a digital context. However, in and of itself, online submission does not enhance assessment or achieve AfL. As an example of the difference, online submission and return of assignments enables the use of feedback comment banks, ranging from low-tech copy-pasting from a list, to high-tech dedicated tools. This enables teachers to provide more information to students, but the degree to which that information enhances student learning remains reliant on the feedback technique of the marker, student attitudes towards feedback, student feedback literacy and the overarching feedback design of the assessment sequence. Further complications of online submission include double-handling of work (e.g. through printing, writing comments and then typing those comments later), slower on-screen marking, technology skill level and resistance (Tomas, Borg, & McNeil, 2015). Technology enables approaches, but logistics and staff experience can prevent enhancement. Technology enablement that is too difficult or unpleasant for staff faces great impediments in improving assessment (Bennett, Dawson, Bearman, Molloy & Boud, 2016).

Rather than technology-enhanced AfL, in this chapter, we adopt the relatively neutral term 'technology-enabled Assessment for Learning' focusing on technology

as not the enhancer or improver but as a tool that provides affordances which may enable assessment approaches. In the next three sections, we explore how technology can enable approaches that scale up AfL in three different ways.

Scaling Up Through Feedback Efficiency: Digital Modalities in Feedback Cycles

Feedback can powerfully influence student learning, and effective feedback processes are a core strategy of AfL. Large-scale meta-analysis of existing research in school education concludes that feedback has a substantial effect on learning in that sector (Hattie & Timperley, 2007). Compelling arguments have been made that these findings are transferrable to higher education and that they are consistent with the higher education-specific feedback literature (Hattie, 2009). Feedback even underpins the causal mechanisms of most of the top 10–20 factors that enhance student achievement (Hattie, 2009). Feedback is therefore a critical site for scaling up AfL. This section explores recent work on using technology to do more feedback with less time and resources.

Technology-enabled feedback approaches are discussed in depth (Moscrop and Beaumont, this volume). In brief, in recent years, there has been an increasing interest in providing feedback through different media in the AfL literature. These approaches typically replace written comments on student work with audio, video or screencast information (e.g. Henderson & Phillips, 2015). There are also bodies of research around efficient use of text-based feedback, such as writing effective feedback for online multiple-choice exams (Lefevre & Cox, 2016) or using feedback comment banks (Debuse & Lawley, 2016).

A key issue in any discussion of scaling up is that of sustainable workload. After an initial time investment in learning to provide feedback in this mode, the use of audio, video and screencast feedback has been reported to take less time while producing a greater volume of feedback information (Henderson & Phillips, 2015; Lunt & Curran, 2010) although some proportion of the extra feedback volume as measured by words may be due to the differences between spoken and written text (Laughton, 2013). Henderson & Phillips (2015) claim that the media affordance of communication efficiency (greater volume of words coupled with richness of media such as gesture and intonation) increased the number of issues that could be discussed, as well as the clarity (as reported by students), particularly in relation to complex issues such as drawing connections between current performance and what needs to be worked on in the future. However, they also point out that in their designs they also spent more time on relational and contextual issues, including recognizing and valuing the student's performance in context of personal circumstances. They claim that this both draws on and reinforces the pedagogical relationship between teacher and student thus facilitating student engagement with the substantive feedback comments.

These approaches scale up AfL by 'doing more with less', an approach to scaling AfL that we define as making improvements for an existing cohort of students while simultaneously reducing teacher time commitment or resourcing. As educators have limited time to implement improvements to assessment, this is a particularly appealing form of scaling up AfL. Doing more with less is not new, and although technology is the enabler here, pedagogy can also be its enabler, for example, Boud (1995) included a 'more with less' argument when justifying self-assessment.

However, with respect to technology-enabled feedback, what is 'more'? Measuring in terms of volume of comments betrays a 'telling' conception of feedback, one that focuses on information transmission from teacher to student. This is a popular conception of feedback, and it is embodied in national surveys of students (Carroll, 2014; Higher Education Funding Council for England, 2014), which asks if 'The staff put a lot of time into commenting on my work' (CEQ) or 'I have received detailed comments on my work' (NSS). However, over the past few decades, thinking about feedback has moved beyond a focus on information transmission and inputs towards a focus on change (Ramaprasad, 1983; Sadler, 1989). Comments on student work are only hopefully helpful information; they only enable feedback when they lead to change in learners. When this information does not lead to change, it is merely 'dangling data' (Sadler, 1989, p. 121). With that conception of feedback in mind, what might 'doing more with less' look like? Technology may allow us to provide more information, but this in and of itself does not generate more feedback.

For change to occur in learners in response to feedback information, the students first need to access the information. Against this hurdle, technology-enabled feedback appears to do more with less. For example, when compared with written paper-based feedback, students were ten times as likely to access audio feedback in one small-scale study by Lunt and Curran (2010). However, as noted by Fawcett and Oldfield (2016), the literature used to support conclusions around increased student access makes the outdated comparison of high-tech versus no-tech and often has research design problems. It is therefore unclear if these approaches result in increased access rates (the 'doing more' part) or just reduce teacher workload (the 'with less' part).

Student preferences and experiences of new feedback media may also suggest that these approaches enable 'doing more with less', in that they provide students with more of what they want. Existing research mostly supports the notion that students prefer these new approaches. This research uses a range of approaches, with emphasis on qualitative and nonexperimental quantitative designs. The one study employing an experimental design, which randomly assigned students to receive feedback information through different media, found no significant difference in the student experience between audio and written comments (Fawcett & Oldfield, 2016). Although they do not provide convincing detail or comparison of the structure or content designs of the text or audio interventions, their finding suggests that novelty effects or unfair comparisons may be partially at play when conclusions are drawn about student preference. Regardless, the suggestion that student preferences are necessarily a sign of improved learning is dubious; for an accessible dissection

of the 'myth' that 'the more they like it, the more they learn', see Clark (2010). While we value the student experience, on their own student preferences are not enough to say that new media enables doing more feedback with less.

It is also possible that a media switch may lead to improvements in the quality of feedback information. This may be verbal or nonverbal; audio feedback may provide prosody and intonation, or video feedback can capture facial expressions and gesture, which may lead to information value beyond words. Qualitative research suggests students can experience video feedback as more real, honest and authentic (Henderson & Phillips, 2015), lending support to this argument. However, there is also the possibility that this more embodied mode may lead to tentativeness, leniency or being less critical (King, McGugan & Bunyan, 2008). Although we would never advocate for being nasty, we would similarly never advocate for being confusingly nice either; approaches like the 'feedback sandwich' can hurt rather than help learning. Students come to expect the predictable 'mandated linguistic ritual' of the sandwich, with flattery largely ignored by learners seeking meaningful criticism in the middle of the sandwich (Molloy & Boud, 2014). For further critique of the sandwich, see also the chapter by Ajjawi and colleagues, this volume.

Evidence suggests that a media switch, on its own, is unlikely to improve feedback. There is a strong tradition of 'media comparison studies' in education, which compare learning outcomes for students across two or more different media conditions with the same instructional design (Russell, 2013). Although individual studies sometimes show better outcomes for one media condition, when the results of hundreds of these studies are pooled together, the overall result is one of no significant difference (Russell, 2013). This tells us that switching from lectures to videos, textbooks to audio books or face-to-face groups to online groups, but not changing our pedagogy, will most likely not lead to better learning (Russell, 2013). This of course assumes the technology functions well and everybody involved knows how to use it.

In educational technology research more broadly, there is compelling evidence that instructional designs, learning outcomes and assessments need to be tailored to suit the new media for educational technology to improve learning (Means, Toyama, Murphy, Bakia & Jones, 2010). The same may be true for feedback designs. Fortunately, the literature has much to say about how to improve feedback designs and proposes several useful models that lend themselves to technology enablement. For example, Boud and Molloy's Feedback Mark 2 model (Boud & Molloy, 2012), which involves an iterative dialogue between students, peers and teachers, could benefit from the logistical enablement provided by online peer review tools to enable peer feedback.

So, do technology-enabled feedback media switches let teachers 'do more with less'? The evidence is stronger for the 'less', than the 'more', depending on which 'more' you mean. If you want more 'feedback' (i.e. more change and improvement) then you should change the feedback design, not just the media.

Scaling Up AfL Thinking from Units/Modules to Programmes/Courses: Technology-Enabled Portfolios and Curriculum Mapping Tools

Assessment thinking has moved from a focus on the immediate needs of a single task or module-level learning outcome, towards a parallel focus on programme-level outcomes. For students and teachers, this can be a complicated task that is cognitively taxing. Where a student's assignment may have been a stand-alone artefact in the past, constructed for the immediate needs of the task and then discarded, it now is likely to be collected into a portfolio tool. Where the intended outcomes of a task were once stand-alone, they now form part of a mapped-out curriculum that can be viewed at a macro- or micro-level. The AfL strategy of productive assessment task design (Carless, this volume) therefore now operates beyond the immediate task at hand and forms part of an integrated and coherent suite.

Portfolios and curriculum mapping are obvious candidates for technology enablement even though they do not require any computerization at all. Both approaches require arduous administration, matching tasks up with unit outcomes, generic graduate learning outcomes with specific degree level outcomes and storing and reporting on large datasets. In a resource-strapped modern university, technology may make portfolios and curriculum mapping cognitively and logistically feasible, enabling the scaling up of AfL thinking from individual tasks up to degree programmes.

The implementations of these tools are large scale too. Portfolios are increasingly marketed and implemented as faculty-wide or institution-wide interventions with long-term agendas (Posey et al., 2015). It would be reasonable then to expect substantial evidence that portfolios enable scaling up of AfL thinking, at a mass scale. However, research investigations into eportfolios tend to focus on small-scale, short-term, self-report data on student preferences that are not particularly useful beyond an immediate context (Rhodes, Chen, Watson & Garrison, 2014). This represents a curious flipping of Selwyn's 'state of the art versus state of the actual': while the rest of the educational technology literature is obsessed with the bleeding edge, the research literature is currently trailing behind the state of practice in eportfolios. Although portfolios can possibly enable this level of thinking, there is little evidence that they commonly achieve this at scale.

Similarly, curriculum-mapping approaches are increasingly being undertaken through software as part of large-scale Assurance of Learning programmes. Like portfolios, these are large-scale endeavours and very common in some parts of higher education (Lawson et al., 2015). Assurance of Learning aims to ensure that on successful completion of a programme of study, all students have met the stated outcomes.

Lawson et al. (2015) report on interviews with leaders from 25 Australian business schools that conduct Assurance of Learning, finding that assessment,

workload and time burden were the most common challenges. We know much about the potential of eportfolios and curriculum mapping tools to improve education – but little of the challenges of large-scale implementations apart from war stories revealed in confidential interviews (e.g. Lawson et al., 2015). The drivers of adoption of these tools may also be not as bright as we would hope; behind the agentic and authentic rhetoric, there may be accreditation and accountability agendas. For example, Assurance of Learning in Australian business schools is driven by external accreditation, government audit and professional body requirements. Although these programmes may have ostensibly had formative goals, 'the actual practice was mostly use of summative assessment' (Lawson et al., 2015, p. 589).

Although tools like curriculum mapping and eportfolios may possibly enable the scaling up of AfL thinking, they do not make it automatically happen. Improving assessment towards AfL is challenging, and it is quite likely that large-scale portfolio and curriculum mapping implementations face similar challenges to other assessment interventions, e.g. resistance, complexity and assessment literacy (Deneen & Boud, 2013; Macdonald & Joughin, 2009; Price, Rust, O'Donovan, & Handley, 2012). Providing the technology tools and bureaucracy of scaled-up assessment thinking will not on its own scale up AfL; cynicism and 'box ticking' must be overcome (Lawson et al., 2015), which undermine AfL. Even if educator thinking about assessment is changed, changes to actual assessment practice are not guaranteed (Offerdahl & Tomanek, 2011).

As with educational technology in general, the social complexities surrounding eportfolios and curriculum mapping tools seem to be as influential as their technological affordances. Technology can reduce the paperwork burden, but addressing pedagogical and curricular challenges is more difficult. Technology can enable scaled-up AfL thinking, but it does not on its own scale up changes to AfL practice to achieve productive assessment task designs.

Scaling Up AfL to Serve an Infinite Student Body: MOOCs as AfL

Massive Open Online Courses (MOOCs) are a relatively new form of online course that provides free access to education for students around the world. MOOCs operate at a massive scale, with tens of thousands of students being enrolled in some courses (Hollands & Tirthali, 2014). In a typical MOOC, content is delivered through video lectures and readings, and students work through a variety of computer-based learning activities.

It is possible to characterize MOOCs as an AfL enterprise in that they are largely structured around meaningful tasks that provide (usually automated) feedback on performance and progress. We recognize that doing so is somewhat contentious, so in the following section, we systematically compare MOOCs against AfL as defined in this volume and then ask what the MOOC experience can tell us about scaling

up AfL in general. In addition to transmission teaching moments (video lectures, readings, etc.), MOOCs often include activities that in regular higher education would fall under the AfL banner, for example, formative quizzes, computer-based assignments that provide rapid feedback and online peer facilitated discussion.

MOOCs are largely built around tasks we class as assessment. Taking Joughin's definition, we view assessment as '[making] judgments 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' (2009, p. 16). The automated delivery environments employed by MOOCs are constantly making judgments about student work, ranging from unsophisticated evaluations of the correctness of their responses to multiple choice questions, to more complex analytics of student online activities aided by artificial intelligence. These are used to make inferences about student capability, leading to the award of certificates on successful completion of the course. MOOCs thus clearly involve assessment, even if a human assessor never sees the student's work.

MOOCs may involve assessment but do they meet this book's definition of AfL and employ AfL's key strategies? In the opening chapter, AfL is defined as follows:

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004, p. 10)

In mentioning the issue of priorities, Black et al. (2004) make a nod to the fact that assessment has a variety of purposes to serve; a single act of assessment may serve many purposes (Boud, 2000). Since MOOCs generate certificates of completion, it could be possible to conceive of MOOC assessment's primary purpose as certification. These certificates do, however, generally hold low status. In response, many MOOC providers offer certificates that can be used for credit in higher education institutions, but these are usually offered only on completion of an additional assessment conducted under more stringent conditions. That additional assessment has the explicit purpose of credentialing; it is clearly an Assessment of Learning event. The remainder of the assessment within an MOOC serves as preparation and guidance towards such an event.

Comparing MOOC assessment against the synthesis of main AfL strategies and processes at the commencement of this book shows the potential for strong alignment. AfL employs productive assessment task design, which is underscored conceptually by constructive alignment (Biggs, 1999). MOOCs can be highly modularized, with clear instructional goals for each section and sequences of learning activities intermingled with low-stakes assessment events that correspond to those goals. Students may be allowed to retake these low-stakes assessments until they are happy with their level of performance.

Assessment for Learning is also underpinned by effective feedback processes. As a resource-constrained teaching mode – resourcing is usually not proportional to the number of enrolled students – little of the feedback in MOOCs comes from human teachers. There is a rich body of research on computer-supplied feedback, and it supports the effectiveness of high-quality feedback on multiple-choice questions

(Lefevre & Cox, 2016); when this mode is used without any feedback at all, it can lead to learning untruths (Marsh, Roediger, Bjork & Bjork, 2007). In addition to often providing immediate feedback on multiple-choice questions, MOOCs also involve rich tasks with intrinsic feedback. For example, one author of this chapter is currently studying the 'R' statistical programming language in a MOOC, which features regular small assignments on which he receives detailed feedback information every few minutes. Peer feedback is also a common feature of MOOCs, which provides a clear tick in the AfL column. However, the motives for using peer feedback are typically driven by resourcing (e.g. Piech et al., 2013), and peer feedback often forms part of a summative peer assessment process; negative responses by students or educators to such an approach are unsurprising (Liu & Carless, 2006).

Students in many MOOCs have opportunities for practicing making judgments, another key strategy for AfL. MOOC research has focused intensely on a variety of judgments: self- and peer-assessment, self-determination, self-regulation and self-direction (Gasevic, Kovanovic, Joksimovic & Siemens, 2014). For better or worse, the free, open and impersonal nature of MOOCs may require a degree of self-regulation not common to traditional face-to-face or online courses.

MOOCs that involve peer feedback and peer assessment require students to make quality judgments. However, the degree to which MOOC students are capable of this, or supported to develop their skills with quality appraisal, has not been very well explored. Exemplars may be provided; however, there is no evidence that sophisticated pedagogies that utilize this as an opportunity to develop evaluative judgment are common. MOOCs do not systematically use the AfL strategy of developing student understanding of the nature of quality.

MOOCs thus are capable of conducting AfL and employing several key AfL strategies – and doing so in a way that scales without additional resourcing. This presents the obvious question: what can the AfL movement learn about scaling from the MOOC experience?

If we take scalability as a function of resources required to serve a particular student body, we see that different educational approaches scale differently. The audio/video feedback approaches described earlier scale almost linearly, which is to say that roughly twice the resources are required to provide feedback to twice the student body. MOOCs are required to scale nonlinearly: an MOOC with 10,000 students does not receive 10x the resourcing it would if it was taught to 1000 students. MOOCs inarguably scale in a nonlinear fashion; however, the degree to which they are AfL depends on the relative priority given to assessment's learning purposes and the strategies used to support student learning.

MOOCs that achieve AfL that scales nonlinearly do so through frontloading AfL resources into educational design, rather than marking or feedback time. AfL scholars already urge us to rethink our resource allocation towards where it best supports learning (Boud, 1995). The MOOC experiment suggests that when student numbers are huge and resources are not, we should invest heavily into design. MOOCs cost tens or hundreds of thousands of dollars to offer, the majority of which is invested in design and development (Hollands & Tirthali, 2014).

Where feedback scalability improved the educational experience for existing students, MOOC scalability relies on the assumption that 'something is better than nothing'. MOOCs provide educational access to students who previously lacked it. However, the low completion rates of MOOCs (ranging from 3 % to 15 % in Hollands & Tirthali, 2014) tell us that in scaling up access to AfL, the resource-constrained approaches taken may be simultaneously scaling down success. To borrow the mantra of the higher education student retention community, 'access without support is not opportunity' (Engstrom & Tinto, 2008). Access to AfL is not the same as a supported AfL opportunity. Although the MOOC experience has demonstrated AfL can scale infinitely, it may be doing so in a way that runs counter to the aims of the AfL community.

Conclusion

In this chapter, we have showcased three approaches to using technology to scale AfL. By switching the media used to deliver feedback, technology enables educators to improve the feedback experience for students without spending more time on marking. By supporting us to think bigger, technology can enable programme-level thinking about assessment with portfolios and curriculum mapping. By changing the relationship between student numbers and resourcing requirements, MOOCs allow AfL to be offered to tens of thousands of students at a time.

A key message in this chapter is that technology does not provide a simple 'out-of-the-box' solution to scaling up AfL. Black and Wiliam's (1998) landmark review study established the importance of context in AfL, and it appears that contextual influences are enduring. The layering of technology in any educational context inherently changes the practices involved, including the production, consumption and interaction of both educators and students. Moreover, the sheer complexity of education, not least the diversity of teachers and students, teaching and learning and policy and institutional culture, means that a successful design for technology enablement of AfL in one context is unlikely to automatically succeed in the same way in another context. As a consequence, it is imperative that AfL designers remain productively wary of technology innovations and the promises of potentiality that surround them. This critical perspective allows us to acknowledge the potential for technology enablement but affirms the need to critically redesign such approaches according to specific contexts and goals.

Indeed, a critical perspective of the three approaches presented in this chapter for using technology to scale up AfL has revealed three key issues that technology-enabled AfL designers should consider.

First, technology-enabled AfL interventions need to be guided by clear goals. Across the three examples explored in this chapter, the intended outcomes were not entirely clear. Were feedback media switches meant to improve learning or just student experience? Are portfolios and course mapping tools for improving learning or compliance? And are MOOCs meant to improve opportunities for AfL or just

access? Any technology-enabled AfL intervention in research or practice needs clear goals and a strategy to evaluate if these goals were met. In the absence of clear goals and evaluation plans, an outsider could reasonably suspect technological determinism: that this 'AfL' intervention is being driven by technology rather than pedagogy.

Secondly, technology-enabled AfL interventions should pay attention to relational and contextual matters. Assessment change is hard enough; when the additional conceptual shift towards AfL is added, it becomes more challenging, and when technology is introduced, the problem becomes more challenging still. Even high-quality portfolio or curriculum mapping tools can be defeated by 'box ticking' approaches by staff and students if they lack the time or support to fully engage. Technological affordances are just possibilities; changing technology without addressing underlying organizational matters is likely a doomed approach.

Thirdly, educators and institutions need to invest in improvements to assessment design. Key meta-analyses of educational technology (Means et al., 2010; Russell, 2013) concur that adding technology to an existing design and expecting improvements is a flawed approach. The feedback and MOOC examples show that investing in improved assessment designs (including feedback designs) is necessary to leverage the gains from technology.

Although we have taken a critical stance through this chapter, our conclusion is fairly positive and aligned with the principles set forth in the opening chapter of this book. Technology may support AfL in its quest for scalability; however, as with AfL in general, productive assessment task design and a concern for the people involved are crucial.

References

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2016). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology*. doi:10.1111/bjet.12439.

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research and Development, 18*(1), 57–75.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8–21.

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

Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. *Computers & Education, 50*(2), 475–490.

Boud, D. (1995). How can self assessment be implemented? In D. Boud (Ed.), *Enhancing learning through self-assessment* (pp. 177–188). New York: Routledge Falmer.

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

Boud, D., & Molloy, E. (2012). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712.

Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79–89.

Carroll, D. (2014). *Graduate course experience 2013: A report on the course experience perceptions of recent graduates*. Melbourne, Australia: Graduate Careers Australia.

Clark, R. C. (2010). *Evidence-based training methods: A guide for training professionals*. Alexandria, VA: American Society for Training & Development.

Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 723–734). New York: Springer.

Debuse, J. C. W., & Lawley, M. (2016). Benefits and drawbacks of computer-based assessment and feedback systems: Student and educator perspectives. *British Journal of Educational Technology, 47*(2), 294–301.

Deeley, S. J., & Bovill, C. (2016). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education.* doi:10.1080/02602938.2015.1126551.

Deneen, C., & Boud, D. (2013). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education, 39*(5), 577–591.

Engstrom, C., & Tinto, V. (2008). Access without support is not opportunity. *Change: The Magazine of Higher Learning, 40*(1), 46–50.

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.

Gasevic, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC research initiative. *The International Review of Research in Open and Distributed Learning, 15*(5), 134–176.

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.

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, New Zealand.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

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

Higher Education Funding Council for England. (2014). *UK review of the provision of information about higher education: National Student Survey results and trends analysis 2005–2013*. London: HEFCE.

Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and reality*. New York: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 higher education edition*. Austin, TX: The New Media Consortium.

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, The Netherlands: Springer.

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.

Laughton, D. J. (2013). Using audio feedback to enhance assessment practice – An evaluation of student and tutor perspectives. *Student Engagement and Experience Journal*.

Laurillard, D. (2008). *Digital technologies and their role in achieving our ambitions for education*. London: Institute of Education.

Lawson, R., Taylor, T., French, E., Fallshaw, E., Hall, C., Kinash, S., & Summers, J. (2015). Hunting and gathering: New imperatives in mapping and collecting student learning data to assure quality outcomes. *Higher Education Research and Development, 34*(3), 581–595.

Lefevre, D., & Cox, B. (2016). Feedback in technology-based instruction: Learner preferences. *British Journal of Educational Technology, 47*(2), 248–256.

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

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.

Macdonald, R., & Joughin, G. (2009). Changing assessment in higher education: A model in support of institution-wide improvement. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1–21). Dordrecht: Springer.

Marsh, E., Roediger, H., Bjork, R., & Bjork, E. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Report of U.S. Department of Education. Retrieved from https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Molloy, E., & Boud, D. (2014). Feedback models for learning, teaching and performance. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology*. New York: Springer Science+Business Media.

Offerdahl, E. G., & Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies. *Assessment & Evaluation in Higher Education, 36*(7), 781–795.

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.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). *Tuned models of peer assessment in MOOCs*. Paper presented at the 2013 Conference on Educational Data Mining, Memphis, TN.

Posey, L., Plack, M. M., Snyder, R., Dinneen, P. L., Feuer, M., & Wiss, A. (2015). Developing a pathway for an institution wide ePortfolio program. *International Journal of ePortfolio, 5*(1), 75–92.

Price, M., Rust, C., O'Donovan, B., & Handley, K. (2012). *Assessment literacy: The foundation for improving student learning*. Oxford, UK: Oxford Centre for Staff and Learning Development.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4–13.

Rhodes, T., Chen, H. L., Watson, C. E., & Garrison, W. (2014). Editorial: A call for more rigorous ePortfolio research. *International Journal of ePortfolio, 4*(1), 1–15.

Russell, T. (2013). *The no significant difference phenomenon*. Retrieved from http://www.nosignificantdifference.org/

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

Selwyn, N. (2010). Looking beyond learning: Notes towards the critical study of educational technology. *Journal of Computer Assisted Learning, 26*(1), 65–73.

Selwyn, N. (2011). In praise of pessimism—the need for negativity in educational technology. *British Journal of Educational Technology, 42*(5), 713–718.

Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education, 37*(5), 591–604.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research, 81*(1), 4–28.

Tomas, C., Borg, M., & McNeil, J. (2015). E-assessment: Institutional development strategies and the assessment life cycle. *British Journal of Educational Technology, 46*(3), 588–596.

University of Minnesota. (2016). *Hype Cycle for Education: Emerging Technologies in Higher Education*. Retrieved from http://hypecycle.umn.edu/page/what-hype-cycle

Warburton, B. (2009). Quick win or slow burn: Modelling UK HE CAA uptake. *Assessment & Evaluation in Higher Education, 34*(3), 257–272.