

Accepted version of Margaret Bearman, Phillip Dawson, Sue Bennett, Matt Hall, Elizabeth Molloy, David Boud and Gordon Joughin. How university teachers design assessments: a cross disciplinary study *Higher Education*. doi:10.1007/s10734-016-0027-7

How university teachers design assessments: a cross disciplinary study

Margaret Bearman, Phillip Dawson, Sue Bennett, Matt Hall, Elizabeth Molloy, David Boud and Gordon Joughin

This work was supported by the Office for Learning and Teaching. We do not have any financial interests or benefits from any direct application of this work.

Abstract

There are dissonances between educators' aspirations for assessment design and actual assessment implementation in higher education. Understanding how assessment is designed 'on the ground' can assist in resolving this tension. Thirty-three Australian university educators from a mix of disciplines and institutions were interviewed. A thematic analysis of the transcripts indicated that assessment design begins as a response to an *impetus for change*. The design process itself was shaped by *environmental influences*, which are the circumstances surrounding the assessment design, and *professional influences*, which are those factors that the educators themselves bring to the process. A range of activities or tasks were undertaken, including those which were *essential* to all assessment design; those more *selective* activities, which educators chose to optimise the assessment process in particular ways; and *meta-design* processes which educators used to dynamically respond to environmental influences. The qualitative description indicates the complex social nature of interwoven personal and environmental influences on assessment design and the value of an explicit and strategic ways of thinking within the constraints and affordances of a local environment. This suggests that focussing on relational forms of professional development that develop strategic approaches to assessment may be beneficial. The role of disciplinary approaches may be significant and remains an area for future research.

Keywords

Assessment; academic context; academic experiences; academic practice; teaching skills

Introduction

Assessment in higher education is a complex process. University educators develop assessment tasks within the nuances of individual, departmental and institutional circumstances, balancing multiple tensions across a diverse range of institutional and disciplinary factors (Price et al. 2011; Macdonald and Joughin 2009; Meyer et al. 2010). There is an abundance of literature which describes innovative and best-practice approaches to assessment, often providing advice with respect to particular activities. There is less consideration as to how individual educators translate these exemplars into local environments. This paper explores how educators design and implement assessment in what Schon (1983, p 42) would call the 'swampy lowlands' of academic practice, where problems and solutions are rarely based on a rational analysis of competing demands. Understanding how local circumstances influence assessment processes may assist both educators and institutions in developing more effective practices.

Assessment is critical to both student learning and student progression in higher education. It promotes learning as well as confirming attainment of standards. Assessment *design* and its associated processes is one of the most critical aspects of assessment practice (Carless 2015; Gibbs and Simpson 2004). Assessment design can be defined to include all processes that take place in order to form specific assessment tasks for a particular course or unit, including the selection and timing of tasks, development of rubrics and re-development of a task in response to student performance. It does not include the individual feedback provided to a student on a particular task, although we suggest that grading and feedback are strongly influenced by task design elements, such as rubrics or scheduling.

While assessment design is often the key to promoting student learning (Carless 2015; Gibbs and Simpson 2004), it is troublesome for many. In a recent phenomenographic study of conceptions of assessment in higher education, Postareff et al. (2012) suggest that the "most striking finding was that the assessment practices in the study context were mostly conventional and that most academics had difficulty describing the purpose of assessment". Price et al (2011) suggest that educators' 'assessment literacy' levels are low and that this has an impact upon assessment practice. Even if educators are familiar with current scholarly notions of good assessment practice they must then be able find ways of developing assessment tasks which take into account their circumstances and pedagogical beliefs and aspirations. Many studies report gaps between their educators' aims for their assessments and how they are actually implemented in practice (Quesada-Serra et al. 2014; Norton et al. 2005; Norton et al. 2013; Fletcher et al. 2012). For example, Norton et al (2013) note that of 586 educators undertaking postgraduate qualifications across UK institutions, only 12% of survey respondents reported that their teaching philosophy and assessment practices were aligned, with 57% of respondents reporting that they were 'uncertain'. Adding to the complexity, educators may believe their assessment philosophies and practices are in alignment, when in reality they may not be (Offerdahl and Tomanek 2011). In other words, the dual challenge is for educators to learn about assessment and to systematically integrate this understanding into their everyday work.

There is some analysis of assessment development from an institutional perspective, where educators are seen as part of a much broader system. Meyer et al (2010) propose relationships between: the assessment literature; professional regulatory bodies including program review processes; the institution and associated strategic plans and policies; teaching staff, including

expectations, research, professional development and codes of conduct; and students. Macdonald and Joughin (2009) echo these, suggesting additional influences on assessment practice: the module and program design; innovation and improvement within a program; departmental culture; institutional resources; institutional recognition; and the external broader higher education context. These frameworks are essentially conceptual rather than empirical. They provide an overview of the various components and relationships in an assessment system. They do not provide insight into what educators actually do, however this knowledge may be key in changing 'on the ground' assessment practices.

An in-depth qualitative exploration of assessment design may provide the necessary insights. To the best of our knowledge, there has not been this type of study focussed specifically on how educators' design assessment, although previous work has identified a range of influences on educators' assessment processes. These include: the tension between summative and formative purposes of assessment (Meyer et al. 2010); the constraints of administrative requirements (Meyer et al. 2010; Norton et al. 2005); the influence of the individual discipline on how assessment is practiced (Norton et al. 2005); and the 'backwash' effect – that is the impact on future design of the educator observing students engaging with the assessment task (Watkins et al. 2005). The diversity of these factors indicate the complexity of assessment design process, with the associated corollary that supporting educators' assessment design practices is both necessary and requires a nuanced approach. A deep qualitative understanding of how educators design assessment may also assist in developing these supports.

To summarise this diverse and brief exploration of the literature, for many individual educators, situated in a system of institutional regulations and relationships, there is a dissonance between what they would regard as ideal assessment design and what they actually do. Previous work indicates a complex terrain, which has not been fully explored. There is a need to understand how educators design assessments 'on the ground' in order to support them to achieve and expand their aspirations. Accordingly, the overall aim of this paper is to qualitatively describe the ways in which Australian university educators develop assessments. Our approach is interpretivist (Guba 1990); we seek to present the synthesised experiences of a few, to illuminate the generality. To this end, this paper investigates the following research question: how do university teachers design assessments in higher education?

This research was undertaken as a foundational study to underpin the development of resources for university educators. We believe that describing how academics enact assessment design within their local environments is critical to understanding and improving assessment in higher education. There are many practical implications; a deeper understanding of assessment practice allows educators to understand and manage their own circumstances as well as permitting institutions to optimise conditions which encourage innovative and exceptional assessment design.

Method

Focus of this study

This study focussed upon the educators responsible for assessment design. In Australia, these are generally unit or subject coordinators and are frequently, but not always, involved in grading and feedback.

Methodology

In describing educator experiences, perspectives and priorities, we undertook a qualitative description approach (Sandelowski 2010), whereby the data analysis remains close to the participants' report of their actions.

Data collection and sampling

We conducted 31 semi-structured interviews with 33 university educators (two interviews contained teaching pairs) from four institutions, which represent the diversity of Australian university traditions. We initially sought interviewees with a range of educational expertise, who were unit coordinators for either new units or units which had just had a major change to assessment. After our initial analysis, it was clear that assessment change was an on-going process and we extended our sample to include unit coordinators for large, stable, undergraduate units.

All authors conducted interviews which were audio-recorded and transcribed. Ethical approval was granted by Monash University Human Ethics Research Committee (CF12/2496 – 2012001350). We sampled from diverse disciplines, using a typology similar to Biglan's categories (Creswell and Roskens 1981). Table 1 and Table 2 indicate characteristics of the interviewees and their units/subjects respectively.

<Table 1 around here>

<Table 2 around here>

The interview schedule (see Box 1 for key questions) focussed the interviewee on assessment design work which occurred within the last 12 months. Interviewees were encouraged to bring unit guides, rubrics and other artefacts to the interview.

Analysis

Interviews were thematically analysed with a view to qualitatively describe how the educators themselves reported assessment design. The analysis techniques drew from grounded theory approaches (Kennedy and Lingard 2006; Strauss and Corbin 1990). In particular, we analysed concurrently with data collection, using an inductive approach that commenced with 'open coding' (Strauss and Corbin 1990), leading to higher order categorisation using 'constant comparison' techniques to finalise codes, themes and categories.

A subset of researchers (SB, PD, MB, EM and MH) primarily conducted the analysis. The analysis team read and independently 'open coded' twelve transcripts, which were selected to represent a range of perspectives. The researchers met and merged four individual coding sets, which had strong similarities, constantly referring to the original transcripts. The process was one of consensus; the analysis team compared each series of open codes for agreement and differences. While most of the analysis team were educators and educational researchers, and brought this understanding to bear

on the analysis, MH provided a perspective external to the higher education sector, and challenged practitioner assumptions. This initial work produced an in-depth coding framework consisting of categories, themes and subthemes and an initial descriptive model.

This initial coding framework was used to code all 31 interviews (PD and MH) using Dedoose (2012) software. Interview transcripts relating to each subtheme were extracted and re-analysed by at least one team member, to ensure the robustness of the analysis and its connection to the data. The team jointly reviewed extracts from each subtheme, reworking the initial coding framework and model, to produce the final qualitative description. Once again, there was a high degree of consensus. Minor adjustments to themes and subthemes were made by later email consensus. The results section presents themes **in bold** and subthemes *in italics*. All names are pseudonyms.

Results

Qualitative description overview

The thematic analysis indicates that assessment design or redesign process begins as a response to some kind of an **impetus for change**, which was then subject to two types of influences. These were: **environmental influences**, which are the circumstances surrounding the assessment design, and **professional influences**, which are those factors that the educators themselves bring to the process. Educators also explicitly and implicitly described a range of activities or tasks, which were required to actually implement the assessment design. These included those activities which were **essential** to all assessment design; those more **selective** activities, which educators chose to optimise the assessment process in particular ways; and **meta-design** processes which educators used to dynamically respond to environmental influences. Figure 1 presents the descriptive model in overview, with a graphical representation of the assessment design process and its subsequent impact upon the next iteration of the assessment. Table 3 provides a list of themes and subthemes. The rest of this section provides an extended fine-grained analysis, which explores each subtheme in detail, including illustrative quotes.

<Figure 1 around here>

<Table 3 around here>

Impetus for change

The theme **impetus for change** was identified as incorporating the phenomena which initiated the design or redesign of an assessment. Some were driven strongly by external circumstance, for example: *unit or program level changes or reviews*, which also might be in response to changing institutional strategic direction; *the unit coordinator being new to the role*; *an entirely new unit or subject*; *the need to save resources*, with educators keen to reduce the time required to implement or grade an assessment; or a desire to better *meet student needs*, which translated into trying to improve assessments and assessment processes. This was often in response to a previous iteration of assessment, which had itself already been a part of the design cycle (as represented by the grey dotted lines in Figure 1). For example:

‘What happened [before] I took over the subject [was that] there was a series of assessment tasks which I didn't see as being particularly meaningful. ... There wasn't any contact with

the tutor. [The students] would just do them, wallow in their own ignorance... and the standard of the work was pretty low. And it was quite difficult to mark... So my brief was to redesign the subject and look at more structured sessions.' Engineering ID31

The various impetuses for change were not necessarily mutually exclusive.

Professional influences

The educators themselves naturally were the most significant factor in how the assessments developed; these **professional influences** reflected the complex intersection of personal histories and professional identities.

Educators' *past experiences* of conducting assessment were strongly influential. Experiences of previous iterations of the assessment task – or similar assessment tasks possibly even at a different institution – were particularly important. Interviewees frequently noted that they implemented changes based on observation of students' performances and responses, with respect to previous assessments. This includes what Watkins et al. (2005) call 'backwash', with data stemming from previous iterations of the same assessment, but also included experiences of similar assessments from another unit, program or institution. Novices to assessment design were distinguished by focussing on past experiences of others' teaching, particularly role models. For example:

'There's one person. Yeah, yeah. He had a very good, strong teaching background that I've learned a lot from ...' Life sciences, ID19

Interviewees also drew from experiences of assessment which occurred to them as undergraduates, postgraduates or in the professional workplace. Some noted the influences from being an administrator or from formal professional development, but this was uncommon.

Beliefs about assessment were broad and varied, and often were co-constructed with other professional and environmental influences. Educators held beliefs about students and how they learnt; they held beliefs around types and modalities of assessment; and they described a range of conflicting beliefs about the purpose and form of grades. Beliefs about learning, often drawn from a disciplinary stance, influenced how educators thought about assessment. Two illustrations of a very diverse set of beliefs are:

'I hate multiple choice tests. They're my pet bugbear....' Life sciences ID2

'I also don't believe that all of that massive amount of testing for tiny amounts actually improves the learning of the students at all. I think it just makes them assessment-weary ...' Languages ID12

The *professional identity* of the participants influenced their assessment designs. Many did not identify as educators; on the other hand, some saw their role as primarily educational. Their particular discipline or vocation could strongly influence assessment design and implementation. As one interviewee noted:

'... because journalism being journalism, deadlines are vital. So we don't give extensions. Unless a meteorite's fallen on their house or they're carried off to hospital in an ambulance.' Journalism ID4

Educators' professional identities were closely intertwined with the *characteristics and circumstances* of the educator, which were also linked to their beliefs about assessment and the past experiences with assessment. These personal characteristics, such as caring for children or particular self-identified traits, affected their attitude towards teaching in general as well as assessment. One illustration of this is:

'... sometimes I wonder why I didn't leave [the assessments] exactly as they were. [But I] really enjoy taking [on] something that is challenging, something that is hard work, something that people hate ...' Allied Health ID6

Prior professional learning included both formal professional development programs ('university training') such as workshops and qualifications, as well as informal professional learning 'corridor talk', team teaching and networking at conferences. Sometimes colleagues who had undertaken formal training were referenced as particular influences. There was some mention of university educational support staff. A few interviewees reported the strong influence of scholarly ideas while most did not refer to the literature at all.

Influences from the environment

The theme **influences from the environment** was categorised into a range of subthemes, which represented the diverse and intertwined range of institutional, collegial and student-derived influences. All of these influences were necessarily mediated by the educators who were designing assessments, and so there was a strong interaction between the educators' own internal drivers and the ones which were external to them.

The most ubiquitous influences related to the *unit or program*. These included factors such as: unit documentation; how the unit was 'handed over' or inherited; size of the unit; mode of delivery; the unit's place in the program; the goals of the unit or program; desired student learning outcomes; and the complex influences of teamwork and negotiation. For example:

'...We have our degree that also not only runs vertically, but horizontally. ... Things aren't taught in isolation... So, if they hear it in mine, they're gonna see it again in Dan's that same year. A year later when they go into Dan's [subject], they're gonna hear it again... A lot of it's hallway talk about who's doing what in what subjects.' Education ID27

Educators saw *student learning* as a strong influence on assessment design. This included formal learning outcomes and objectives but was also strongly linked to broad ideas about the student's future trajectory outside the academy. An example of the latter is:

'...after you start to sample what [the general public] believe or what they know [about science], it's actually quite frightening.... So, there's opportunities there when [the students] actually educate everybody else.....' Life sciences ID3

There was a strong sense from many that assessments were successful when they noticed students had misconceptions corrected or when they had achieved higher order learning.

Educators were also influenced by *organisational requirements*. These included handbook entries, mapping of assessment to standards or frameworks and a range of approval processes. In considering this subtheme, it was difficult to disentangle formal requirements, recommendations,

norms and myths. The presence of institutional requirements generally promoted keeping the status quo rather than promoting change. Sometimes this led to frustration:

‘Policy restrictions stop us from being able to be responsive. If we see something in a unit that's not really working but [if] it's in the unit guide and it's in the handbook [it can't be changed]...’ Languages ID12

As is clear from the previous quotes, there were also influences relating to *organisational culture*. This worked at different layers, from the immediate department to the broader institution. It encompassed division of work and teaching roles, the spoken but unwritten ways of doing things, and the implicit and unspoken ways of doing things. One interviewee articulated:

‘So, I would say that our Head of Department has a certain philosophy, and a certain view of our program and certainly that filters down into all the individual units. And certainly I wouldn't make any dramatic changes without obviously consulting her.’ Allied Health ID18

Characteristics of *student cohorts* were a very strong influence on assessment design. Diverse cohorts were challenging; as were mismatches between standards and cohort capabilities. For example:

‘[The students are] given very specific questions and they're told how many lines to write for each question ... because they are first years and it's first semester.’ History ID11

Responses and interactions which provided data or feedback regarding a specific assessment also formed a significant part of the assessment design process. Formal and informal peer review strongly influenced assessment design. However, students were the most significant source of data either through formal interactions such as unit evaluations, or informal interactions such as observation of student behaviour or performance on an assessment task:

‘I think it's a case of thinking, "oh, they aren't doing too well in this.... They've written quite a bit, but it's not actually answering the question... Is that because the question was wrong or because we didn't make it very clear?’ So, we're making the questions clearer and we're getting better responses.’ Physical sciences ID23

There were *resource* influences, including time, money, technologies, scheduling and other logistical factors, people and suitable spaces. For example:

‘We would love to *not* have the online assignments and have written stuff for that, but you've got a thousand students and what do you do? ... And that's where our assessments do not meet our learning outcomes.’ Physical sciences ID23

Not everyone felt constrained and many found ways to approximate their desired assessment with the available resources. On occasion, assessments produced resources – such as videos for patients or wiki sites for other students – as well as consumed them.

External factors included those influences from the community, professional bodies and other universities.

‘Industry are telling us all the time that they need to improve their communication skills... then there's got to be some sort of assessment in that.’ Engineering ID31

Educator activities

Against this backdrop of influences, educators undertook the activities which led to design, development and implementation of their assessment. These are categorised into three themes: essential, selective and ‘meta’ design activities.

The **essential design activities** are those actions which must be undertaken in order to design and implement an assessment task. These are mostly self-evident but worth recording as they are part of the core functions of assessment design. Educators *developed assessment tasks*, determining the form as well as the content of the assessment task itself. Educators designed *marking/feedback processes*, such as rubrics, feedback timing and quantity, moderating marks, using peer or other feedback innovations, and the use of technology to support processes. *Aligning teaching and assessment activities* was a large concern. This included structuring assessments and assessment schedules, so that they linked to other teaching and assessment tasks, and optimised feedback, resources and student workload. For example one interviewee noted:

‘... They've looked at a general introduction to Aged Care. They've looked at individual aspect of older age.... So, it then makes sense to me that they [then] go and have a conversation with an older person....’ Allied Health ID10

Educators also provided *instructions to students*. In addition to documenting and communicating the tasks required of the students, educators also clarified expectations of the assessments and grading standards; provided exemplars; and communicated required skills such as ‘exam techniques’. In general, interviewees noted that time spent clarifying ‘at the start’ saved time later down the track.

In addition to continually repeating these essential activities, educators had a range of additional concerns which resulted in **selective design activities**. These concerns were individual and diverse. A subset of these activities was primarily concerned with the grade-bearing aspect of the assessments. Some educators actively took actions to *reduce cheating* through peer assessment, continual changes to assessment, and providing students with skills to avoid plagiarism. Some educators tried to *ensure student equity*. For example:

‘As of three years ago, upward of 10% to 15% of our cohort are international. And they tend to do very well on the written content but not so much on the oral part... So I think they would be disadvantaged in terms of doing the oral viva.’ Allied Health ID8

Educators also adapted their assessments to manage *degrees of difficulty*. This was an activity which included consideration of what the standards should be, what the students’ capabilities were, increasing complexity of tasks and matching different tasks to different students.

Another subset of these activities were primarily concerned with the learning aspect of assessment. Some educators actively *promoted student engagement*, with the assessment tasks. This included ‘intrinsic’ rewards such as making authentic and relevant tasks, peer assessment, rotating tasks, interactive activities such as role-play or using competition; and also ‘extrinsic rewards’ such as bonus marks, grades, compulsory assessments.

‘ ... the final assessment for the demonstration is built as a bit of a competition that also is part of the engagement. So the [software assessment] this year was a mine-detecting system. So we made a rough terrain, we buried some fake mines and really all they did was they had a big sort of piston that if you put too much weight on them – poof! ... And the cheers and the cat calls and the engagement.’ Engineering ID25

Some educators actively promoted a higher *degree of student agency*. This included allowing students to determine topics or tasks, or taking a research orientation, but it sometimes caused tensions with grading and comparing non-equivalent tasks.

Assessment tended not to be a singular process; it was iterative. The activities that educators undertook to manage their own design processes across iterations were categorised as **meta-design activities**. The first of these activities was *adapting against the constraints*; when faced with a conflict between their personal and environmental influences, some educators designed an assessment that could accommodate both solutions. A similar, but even more deliberate activity, took place when educators *strategized* regarding how to achieve change. Approaches included negotiating with colleagues and managing collegial input, keeping the head of department informed, mapping the assessment against frameworks and criteria, ‘selling’ the assessment to students, deliberately managing the paperwork to support change, ensuring the right teaching team was available, and changing assessment over several iterations of a unit rather than all at once. An example of this type of strategic approach:

‘I tried to get buy-in from the key stakeholders within the department...Typically, I would let the head of department know that I'm looking at changing this, and that I'll be sending some ideas their way shortly. So, just sow the seed of thought. I tend to always come up with a draft, or a few drafts of options ...rather than going to them and saying, "Help me come up with this.”’ Allied Health ID6

Discussion

The thirty-three interviewees described a complex picture of assessment design in higher education. The analysis suggests that educators develop assessments by drawing from both past experiences and personal and professional circumstances. Their current academic environment was a tangible factor in promoting or constraining assessment design processes. All educators conducted essential design activities, with variable emphasis on selective areas of interest. Some individuals undertook meta-design approaches to overcome challenges presented by their circumstances. It is worth noting that, in general, the relationships were not directly causal. In other words, an identified influence did not necessarily lead to a specific outcome or activity.

Overall, the data that we present aligns with, but varies from, previous conceptual work (Meyer et al. 2010; Macdonald and Joughin 2009). Almost all of the identified elements such as the influence of policies, regulatory bodies, departmental culture and institutional resources are present in the findings; the one exception is ‘codes of conduct’ which were not mentioned. Likewise, there are resonances and variations with previous empirical work on individual educators’ dissonances between assessment beliefs and practices (Postareff et al. 2012; Fletcher et al. 2012). However this qualitative, ‘on the ground’ perspective presents a more entangled view, with strongly interwoven

personal, organisational and environmental factors. While many educators experienced dissonances between what they wanted to achieve with assessment and what they actually implemented, many others did not: they found creative solutions when balancing competing demands. The following discussion both situates this study within the broader literature and presents key implications for approaching improving assessment processes and practices.

Past experiences and current contexts

This study provides a different perspective in that it does not purely focus on individual educators' conceptions of assessment (Postareff et al. 2012; Fletcher et al. 2012) nor provide a theoretical sociocultural Bourdiesian analysis of assessment practice (James 2014). It grounds an understanding of assessment practice in the design of specific assessment tasks. On the one hand, the findings support the importance of the individual educator, their beliefs and the choices that they make. On the other hand, the analysis also recognises that the educator shapes, and is shaped by, a broader sociocultural environment.

The relationship between the individual and the social extends throughout the findings. It is found within category *impetus*, which represented the 'spark' that initiated change at a particular point in time, often as part of an iterative process. The requirement to design or redesign assessment could be entirely driven by social circumstance; the most frequent is described in the subtheme of becoming a *new unit coordinator*. However, once educators had assumed the role of unit coordinator, they generally exercised control over many aspects of assessment redesign; this aligns with the findings of Bennett et al (2011), who suggest that Australian university educators have considerable flexibility in their design decisions. This subtheme highlights both the importance of the individual educator as well as the important status afforded the unit coordinator position.

When the *influences* category is considered as a whole, the mutually constructed and interlinked nature of the personal and contextual factors is striking. The previously experienced environmental influences became subsumed into the personal history of the educator as they developed assessment during their academic careers. In other words, today's assessment design experience becomes tomorrow's assessment design influence. This suggests that any approach to improving assessment design should not ignore the past contexts and experiences. In the same way we ask our students to draw from previous experiences to form current conceptions, it may be worth explicitly asking educators to compare and contrast their current trajectories with their past experiences of assessments as learners and as teachers.

The value and limitations of disciplinary traditions

The illustrative quotes generally convey the overall sense, most strongly represented in the subtheme *student learning*, that educators generally were very concerned with what students were thinking, learning and experiencing. This seemed to be often linked to educators' *professional identities*, in that they wanted to see students demonstrate thinking or writing or acting like historians or scientists or health professionals; what might be called a disciplinary approach to assessment. There was a sense that educators felt very strongly that this disciplinary approach would benefit the students' academic and life pathways.

Carless' (2015) learning-oriented assessment model proposes relationships between learning-oriented assessment tasks with a disciplinary focus, developing student evaluative judgement and student engagement with feedback. His model emphasises how learning oriented assessment is related to students developing 'evaluative judgement' – the capacity to identify and make complex judgements around markers of quality. A key insight from the present study, is that in promoting such models, it may be useful for 'on the ground' educators to focus on a disciplinary notion of 'evaluative judgement'. This is because 'evaluative judgement' is often seen as an abstract concept but how to promote thinking and writing like a historian is more immediately meaningful.

Overall, the theme **professional influences**, indicates the breadth of disciplinary variation, with interviewees describing exposure to a wide array of assessment designs through different courses, institutions, countries and colleagues. We regard this as positive, as the breadth and diversity of approaches is appropriate for a complex teaching process like assessment. However, it may also mean that educators may lack the capacity to see beyond their particular experience. We note that, with some exceptions, the language of assessment literature was absent. Educators often spoke passionately and thoughtfully about their assessment designs enhancing student learning in a very broad sense but without reference to literature or theory. James (2014) suggests that explicit consideration of 'doxa' – that which is so accepted it is not noticed (Bourdieu 1977) – is relevant to assessment practice and this study bears this out. Some participants focussed on disciplinary traditions such as exams or essays simply because they could not envisage an alternative. Likewise, many participants overlooked the role of formative assessment and focussed mostly on grade-bearing tasks. This issue, of expansion beyond the familiar, may be an area where organisational supports may be required to change assessment practice.

Local contexts: local solutions

The subthemes of *organisational requirements* and *organisational culture* findings suggest assessment is impacted by institutions, faculties and departments at every level. This could be positive – for example participants identified benefits, such as committees requiring peer review of proposed assessments – as well as concerns, such as challenges due to procedural restrictions on the speed of changing assessments. The role of the departmental leadership was particularly notable and appeared at its most powerful when the departments were small. The head of department seemed to determine the significance and merit of assessment innovation in several ways: firstly, through control of resources; secondly, through the 'status' afforded teaching in general and assessment in particular; and finally through their own beliefs as to what constituted good assessment. This role of local leadership in promoting assessment practice also underlines the social dimension of innovation.

Another way of thinking which may be useful for many educators was identified within the *meta-design* theme. As noted in the analysis, some interviewees recognised what was mutable or negotiable within their contexts generally, and what compromises could and couldn't be made surrounding their assessment design and implementation. What was perceived as challenging for some educators in one circumstance, was easily negotiated by others. There was a sense that these educators saw failure, or the risk of failure, as necessary. Some sought to minimize their workload while maximizing student learning. This type of inventive thinking is a dynamic skill. Educators have to be responsive to students and to colleagues, but at the same time, manage their scarce resources.

Assessment may be improved by prompting educators to actively think about how they implement their design within their particular local contexts. We have utilised some of the outputs of their study in a website intended to promote this type of thinking [REMOVED FOR REVIEW] but further work is needed on how these types of resources impact overall assessment design.

Implications for professional development

Many of the previously discussed issues – educators explicitly reflecting on past experiences of assessment; moving beyond the familiar assessment conventions; focussing on disciplinary examples of evaluative judgement; and asking educators to actively and strategically reconcile the natural tensions within assessment design – represent some solutions to the issue of translating assessment theory to assessment practice. This has real implications for professional development processes. The *prior professional learning* and *past experiences* subthemes also provide insights; they indicate the social nature of educators' learning. Educators emphasised the value of colleagues' experiences, corridor conversations and workshops but rarely mentioned reading the higher education literature. This suggests that improving assessment through professional development may not just be about individuals, but about relationships. This raises some key questions: how can we join individual learning to organisational learning? How can professional development in assessment be seen as more like a social privilege (such as a sabbatical) not an individual burden (such as mandatory training)? Do departments, faculties and institutions need to think about time within local face-to-face contexts to share, discuss and strategise about assessment? All of these questions may be more familiar to organisations when discussing researcher development rather than teacher development and suggests the relevance of James' (2014) Bourdieusian analysis of assessment in higher education, which proposes that assessment practice reflects the overall complex social structures inherent in academia. It may be that if we value assessment development in the way we value research development with its focus on teams, collegiate exchange and the value of strategic thinking, we may see some striking results.

Limitations

There are limitations to this study. The exclusively Australian nature of the sample may not translate to other nations; however like all qualitative studies the aim is for illumination rather than generalizability (Patton 1999). The data is self-report, and while this is appropriate for a study investigating educators' experiences, the qualitative description approach reflects the concerns and priorities of participants and details of issues of interest to assessment experts such as rubrics or feedback practices did not emerge. Different analytic lenses may provide different insights. Additional data would also be valuable. Observing educators in the design process would overcome some of the limitations of self-report. Likewise, exploring student experiences may complement the educator dataset. A close qualitative focus on a particular department may produce additional insights into the social nature of assessment development. These all form avenues for future research.

While the thematic analysis revealed examples of the influence of disciplinary thinking, it was not a focus of this research. As Norton et al (2013) suggests, the role of disciplinary differences in assessment requires further exploration. While this study supports Carless' (2015) proposition that disciplinary thinking may have a key role in learning-oriented assessments, future studies may reveal

more nuanced understandings about the intersection between disciplinary thinking and assessment design.

Conclusion

This study presents insights from Australian university educators on how and why they design assessment. It provides a thematic analysis of interview transcripts reporting broad categories of impetus, influences and activities. The analysis and discussion explores how themes and subthemes describe a complex, connected picture of the individual choices and social relationships which surround the construction of a piece of assessment. Educators were generally thoughtful and concerned about student learning, both in the short and longer term, and had some degree of control over the assessments within their units. Thinking about assessment and assessment design as both individual and social process has implications for organisational approaches to professional development. We suggest there are parallels to research development, which is high status, discipline-focussed and relational.

REMOVED FOR REVIEW

- Bennett, S., Thomas, L., Agostinho, S., Lockyer, L., Jones, J., & Harper, B. (2011). Understanding the design context for Australian university teachers: implications for the future of learning design. *Learning, Media and Technology*, 36(2), 151-167, doi:10.1080/17439884.2011.553622.
- Bourdieu, P. (1977). *Outline of a Theory of Practice* Cambridge: Cambridge University Press.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963-976, doi:10.1007/s10734-014-9816-z.
- Creswell, J. W., & Roskens, R. W. (1981). The Biglan Studies of Differences among Academic Areas. *Review of Higher Education*, 4(3), 1-16.
- . Dedoose, web application for managing, analyzing, and presenting qualitative and mixed method research data (2012). Los Angeles, CA: SocioCultural Research Consultants.
- Fletcher, R., Meyer, L., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and Students Conceptions of Assessment in Higher Education. *Higher Education*, 64(1), 119-133, doi:10.1007/s10734-011-9484-1.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, 1(1), 3-31.
- Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 17-30). Newbury Park: Sage.
- James, D. (2014). Investigating the curriculum through assessment practice in higher education: the value of a 'learning cultures' approach. *Higher Education*, 67(2), 155-169, doi:10.1007/s10734-013-9652-6.
- Kennedy, T. J. T., & Lingard, L. A. (2006). Making sense of grounded theory in medical education. *Medical Education*, 40(2), 101-108, doi:10.1111/j.1365-2929.2005.02378.x.
- 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): Springer Netherlands.
- Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., et al. (2010). An Investigation of Tertiary Assessment Policy and Practice: Alignment and Contradictions. *Higher Education Quarterly*, 64(3), 331-350, doi:10.1111/j.1468-2273.2010.00459.x.
- Norton, L., Norton, B., & Shannon, L. (2013). Revitalising assessment design: what is holding new lecturers back? *Higher Education*, 66(2), 233-251, doi:10.1007/s10734-012-9601-9.

- Norton, L., Richardson, T. E., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education, 50*(4), 537-571, doi:10.1007/s10734-004-6363-z.
- 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, doi:10.1080/02602938.2010.488794.
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research, 34*(5 Pt 2), 1189-1208.
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation, 38*(3-4), 84-92, doi:<http://dx.doi.org/10.1016/j.stueduc.2012.06.003>.
- Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: a critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education, 36*(4), 479-492, doi:10.1080/02602930903512883.
- Quesada-Serra, V., Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2014). What are we missing? Spanish lecturers' perceptions of their assessment practices. *Innovations in Education and Teaching International, 1-12*, doi:10.1080/14703297.2014.930353.
- Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in Nursing & Health, 33*(1), 77-84, doi:10.1002/nur.20362.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126): Basic books.
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*: Sage Publications, Inc.
- Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science, 33*(4), 283-309, doi:10.1007/s11251-005-3002-8.

Box 1: Key interview questions

- What led you to develop these assessment activities?
- What reasons did you have for altering it/maintaining it?
- What parts of developing these assessment activities do you feel ownership over, and what parts were given to you by someone else? Who else was involved?
- What formal documents were there to submit? How did this influence what you did?
- How typical is this [task] for what you normally do?
- Are there any other features in your course related to assessment or feedback, perhaps not grade-bearing?
- Can you summarise the key decision points in developing assessments for this unit?
- Under different circumstances, what would you have liked to have done?
- How has the way you design assessment changed or developed over the years?

Table 1: Interviewees – demographics (n=33)

Gender	22 females, 11 male
Graduate Certificate in Higher Education or other teaching qualification	18 had completed 9 had not completed 6 did not report
Level of employment	6 Professor or Associate Professor 25 Lecturer or Senior Lecturer 2 Assistant Lecturer or sessional

Table 2: Characteristics of units/subjects) (n = 31)

Disciplines	9 humanities professional degree subjects (education, journalism) 7 humanities general subjects (history, politics, geography, sociology & languages) 8 sciences professional degree subjects (allied health, engineering) 7 sciences general subjects (life sciences, physical sciences)
New unit	2/31 were newly developed units

Table 3: Overview of themes

<i>Categories</i>	<i>Impetus</i>	<i>Influences</i>		<i>Activities</i>		
Themes	Impetus for change	Professional	Environmental	Essential	Selective	'Meta'
<u>Subthemes</u>	Unit/program level changes/reviews New unit coordinator New unit/subject Saving resources Meet student needs	Past experiences Beliefs Professional identity Characteristics and circumstances Prior professional learning	Unit or program Student learning Organisational requirements Organisational culture Student cohorts Resources External factors Responses and interactions	Developing the assessment tasks Aligning teaching and assessment activities Designing marking/feedback processes Providing instructions to students	Reducing cheating Ensuring student equity Managing degrees of difficulty Promoting student engagement Managing degrees of student agency.	Adapting against the constraints Strategizing

Figure 1: An overview of the assessment design process