

Bond University
Research Repository



Assessment thresholds for academic staff: constructive alignment and differentiation of standards

Simper, Natalie

Published in:
Assessment and Evaluation in Higher Education

DOI:
[10.1080/02602938.2020.1718600](https://doi.org/10.1080/02602938.2020.1718600)

Licence:
Other

[Link to output in Bond University research repository.](#)

Recommended citation(APA):
Simper, N. (2020). Assessment thresholds for academic staff: constructive alignment and differentiation of standards. *Assessment and Evaluation in Higher Education*, 45(7), 1016-1030.
<https://doi.org/10.1080/02602938.2020.1718600>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

Assessment Thresholds for Academic Staff: Constructive alignment and differentiation of standards.

Natalie Simper

Faculty of Engineering, Monash University, Victoria, Australia

natalie.simper@queensu.ca

This qualitative study utilized episodic narrative interviews to investigate assessment thresholds involved in the development of assessment literacy. The goal of the study was to inform efforts toward quality improvements in higher education. Thirty-five academic staff from universities in Australia, Canada and Sweden shared stories of significant changes they made to their assessment practice. Thematic analysis found troublesome aspects to include instructor expectations, lack of consistency, differentiation of performance, student expectations, time constraints/ workload, logistical/ technological constraints, and assessment policy. A belief in meaningful learning, embracing constraints and the desire to meet the needs of students, added to other enablers for assessment change, such as resources, support and strategic use of technology. Findings suggested assessment thresholds of *constructive alignment* and *differentiation of standards*. Reflection, collaboration, and professional development were found to support the integration of assessment knowledge and build conceptual understanding of assessment thresholds. Authors recommend that higher education institutions provide academic staff with a foundation of conceptual understanding of these key areas to promote moves toward quality assessment practices.

Introduction

Over the past decade, a great deal of attention has been paid to assessment of student learning in higher education, led by guidance and regulatory publications from various agencies (González & Wagenaar, 2003; Harris, 2009; Barrie, Hughes, Crisp, & Bennison, 2011; Hénard & Roseveare, 2012; Jankowski, Hutchings, Ewell, Kinzie, & Kuh, 2013; Tremblay, 2013). Assessment has been positioned as central to efforts for quality improvements in higher education, yet academic support and development have been primarily focused on teaching practices (Steinert et al., 2006) and there has been limited evidence directly linking improved teaching with specific assessment practices (Baird, Andrich, Hopfenbeck, & Stobart, 2017). There is also debate about the effectiveness of different types of assessment on student learning (Falchikov & Boud, 1989; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012).

Research around assessment literacy suggest that academics require expertise beyond their disciplinary knowledge, incorporating specific knowledge and skills in education and assessment (DeLuca, LaPointe-McEwan, & Luhanga, 2016). That being, the language of assessment, the purpose and systems of assessment, ability to work with guidelines on standards, and the ability to illustrate the way standards are communicated (Abell & Siegel, 2011; Medland, 2019). This a tall order for academics in higher education. Meeting requirements for quality assurance have generated a culture of compliance when it comes to assessment, but not necessarily improved outcomes (Stensaker, 2003; Brown, 2004; Shah & Jarzabkowski, 2013). The use of assessment for accountability metrics has been known to undermine meaningful engagement in quality practices (Boud & Dochy, 2010; Rhodes & Finley, 2013; Biggs, 2014).

Assessment

Confusion can stem from the use of the word assessment, so defining terminology is an important step in facilitating discussion of quality assessment practices. It is commonly used as a label for the type of activity, such as high-stakes assessments, pre-assessments, formative assessments, or summative assessments (“Assessment Definition,” 2013). However, the practice of assessment “involves identifying appropriate standards and criteria and making judgements about quality“ (Boud, 2000, p. 151). The word evaluation is sometimes used in place of assessment, but more generally for determining teacher or program effectiveness (Popham, 1974). To confuse things further, assessment is also used interchangeably with the word assignment to describe a submitted product, as in, *students are penalized if they hand their assessment in late*.

Assessment literacy

The lack of clarity in terminology and limited understanding of assessment more generally has prompted research in the area of assessment literacy. Medland, (2019) suggested that key areas for assessment literacy were the ability to apply basic principles, provision of feedback and the use of constructive alignment. Constructive alignment is a is “an outcomes-based approach to teaching in which the learning outcomes that students are intended to achieve are defined before teaching takes place (Biggs, 2014, p. 5). Reimann (2018) described assessment literacy as having an understanding of:

- the language of assessment,
- the purpose of assessment
- the systems of assessment,

Further, the ability to:

- work with guidelines on standards, and
- illustrate the way standards are communicated.

The above knowledge and traits allude to thresholds in understanding or application when it comes to assessment. To date there has been no empirical research around threshold concepts for assessment.

Threshold Concepts

Threshold concepts have had increasing appeal since their conceptual origins (Hounsell et al., 2005). Described as *a portal* to new ways of thinking (Meyer & Land, 2003), but ambiguity in their properties have led authors to understand them in different and sometimes incompatible ways (Rowbottom, 2007). Threshold concepts revolve around the idea of liminality as a fluid state of understanding or being. The liminal space is where individuals work through stages of uncertainty, referred to as “receptivity, recognition and grieving, in which an established pattern of meaning is no longer tenable or valid for future practice” (Land, Cousin, Meyer, & Davies, 2005, p. xiii).

Characteristics of threshold concepts:

- Troublesome because they move from the familiar to unknown territory. Troublesome knowledge is characterized as inert; ritual; tacit; counter-intuitive; alien; conceptually

- difficult knowledge; or related to troublesome language. This ‘troublesomeness’ is what triggers or instigates the threshold concept ‘journey’ (J. Meyer & Land, 2003, 2006)
- Transformative: Involving a cognitive shift in understanding, or a change in ontological position (Land et al., 2005; Atherton, Hadfield, & Meyers, 2008; Irvine & Carmichael, 2009; Walker, 2013).
 - Integrative: The learner becomes aware of the inherent interrelatedness of threshold concepts as they move toward expert knowledge. (J. H. Meyer, 2010; J. H. F. Meyer & Timmermans, 2016).
 - Irreversible: A transformation in understanding makes it difficult to go back to a novice or naive view of concepts (Cousin, 2006; Davies, 2016).
 - Bounded: Arguably pertaining to a specific domain, but helping learners explore the edges of conceptual knowledge (Walker, 2013).
 - Discursive: A shift use of language provides a new way of talking about the concept (J. Meyer & Land, 2003; Carmichael, 2012)
 - Reconstitutive: Integration can also entail a shift or repositioning in the learner’s position (Land et al., 2005).

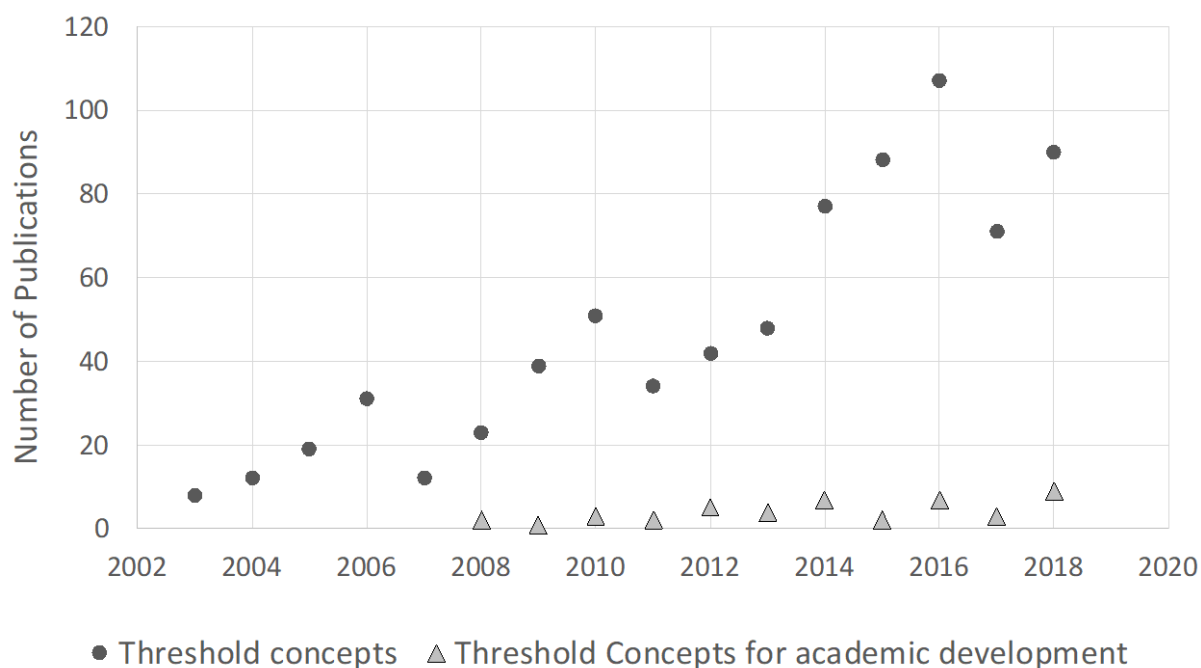
Although there is conceptual similarity, threshold concepts should not be confused with the frameworks for discipline-specific standards in Australian higher education, called Threshold Learning Outcomes (Tertiary Education Quality and Standards, 2015) or the benchmark statements in the U.K. (Quality Assurance Agency for Higher Education, 2018). These standards cover specific and general knowledge, skills, and application in the context of disciplines, but there has been no direct assertion that these are threshold concepts.

Methodologies for research on threshold concepts in academic development

A search for the term *threshold concepts* in the abstracts of publications in the last 10 years found 725 results encompassing 56 different disciplinary fields. Despite the breadth of investigation, there are inherent difficulties with empirical exploration of threshold concepts (Rowbottom, 2007; Barradell, 2013; Basgier & Simpson, 2019). For example, “what if a concept is troublesome and integrative but not transformative, is it still a threshold concept?” (Barradell, 2013, p. 266). Threshold concepts are themselves proposed to be a threshold concept for academic staff (Basgier & Simpson, 2019).

Threshold concepts have been primarily studied to develop discipline-specific understanding. Of the 725 threshold concept search results, only 11 of the publications related to *faculty or academic development*. Figure 1 plots the frequency of publications with *threshold concept* mentioned in the abstract, sourced from books, book chapters, journal articles, reports, or thesis/ dissertations. The sub-set of publications that include the *subject* of faculty or academic development are indicated with triangle markers.

Thresholds experienced by academic staff have been studied through interpretative phenomenology by investigating the lived experience using various methods such as “assessed dialogue” (Pilkington, 2019), case studies (Quinlan et al., 2013; Timmermans, 2014), critical incident interviews (Shinners-Kennedy & Fincher, 2013; Basgier & Simpson, 2019), semantic networks (Walker, 2013), semi-structured interviews (Quinlan et al., 2013; Wilcox & Leger, 2013; Rodger, Turpin, & O’Brien, 2015; Reimann, 2018), survey/ focus groups (King & Felten, 2012; Webb, 2015; Kilgour, Reynaud, Northcote, McLoughlin, & Gosselin, 2018), or reflective narrative inquiry (Carmichael, 2012; Timmermans, Bruni, Gorbet, & Moffatt, 2018).

Figure 1. Increasing frequency of publications featuring threshold concepts.

The *Critical incident* technique (Flanagan, 1954) stood out as a way of navigating the ambiguity surrounding the transformative nature of thresholds. It was suggested that thresholds can be qualified as ‘stand out’ experiences of profound learning, and that “acquisition of a threshold concept is described as an emotionally laden event that can have the effect of altering the learner’s view of themselves” (Shinners-Kennedy, 2016, p. 257). In the critical incident interview, experts are asked to provide the contextual background and circumstances surrounding a situation of critical significance to their role, with multiple perspectives contributing to a broad understanding.

The goal of the current study was to explore threshold concepts related to assessment in higher education, adding to the limited literature around thresholds for academic staff. The current study explored changes in assessment, seeking to uncover the point at which a symbolic threshold was crossed, with an irreversible change in knowledge, perception or behaviour.

Research questions

1. What troublesome areas triggered academic staff to change their assessment practices?
2. What thresholds were apparent in the conceptual understanding of assessment principles for academic staff?

Method

The critical incident technique (Flanagan, 1954) was adapted to create an episodic narrative interview protocol (Flick, 2000) for the purpose of uncovering the perspectives underlying assessment thresholds. The interview involved participant’s description of a significant change they made to assessment. It is theorized that episodic memory can (among other things) be accessed flexibly, expressed symbolically, and used as a basis for inferences (Terrace & Metcalfe, 2005).

Procedure

Interviews ran for approximately one hour, conducted in a private location of the participant's choosing. Ethical consent was obtained, and interviews were audio recorded to produce transcriptions. There were five introductory questions asking about participant's role, teaching experience and assessment at their institution. Following that, participants were asked to think about a significant time when they changed the way they assessed student learning. A list of ideas was provided for reflection, but the responses were not constrained to these: Moving to using rubrics/ quality frameworks; change in assignment or intended outcomes; including students in the assessment or peer assessment; change in assessment structure; moderation or training for consistent judgments. After sufficient time to reflect, participants were asked to explain the situation, providing as much detail as possible. Prompts included the context, details of the student group, size, year, and needs. They were also asked to explain the reasons behind the change, and the result or outcomes from the change.

Participants

In seeking rich accounts of the phenomena of thresholds for assessment in higher education, purposeful sampling was employed (Patton, 1990). Participants were recruited from medium-sized, doctorate granting, research intensive universities, one in Australia (n=12), one in Canada (n=12) and one in Sweden (n=11). The sample was delimited by academic teaching role, active involvement in assessment improvement activities or recognition of teaching excellence. To improve generalizability, participants were sampled from a range of learning areas. The sample comprised academic staff from Business (9%), Education (6%), Engineering (29%), Health Sciences (14%), Humanities (9%), Sciences (14%), and Social Sciences (20%). There was an array of seniority, Graduate teaching fellows accounted for 9% of the sample, Adjunct/ Sessional Lecturer, 9%, Assistant Professor/ Lecturer (continuing appointment), 17%, Associate Professor/ Senior Lecturer, 46%, and Head of Department 20%.

Analytical approach

A reflexive thematic analysis process was followed (Braun & Clarke, 2006), observing the centrality of researcher subjectivity in identifying data that related to classified forms to generate overarching patterns. The state of mind, or moment prior to change was of particular interest for the qualification of a 'stand out' experience. The seven characteristics of threshold concepts were reduced to four topics to encompass the overarching principles:

- Troublesome aspects for assessment;
- Transformation (irreversible cognitive or ontological shift);
- Integration and bounded knowledge;
- Specificity and the use of assessment language and repositioning.

Following a close reading of the transcripts, the data were grouped into the above areas, then themes were generated and relationships were drawn. Synthesis of the themes are represented in Figure 2.

Results

Participants spent differing amounts of time selecting an assessment change example that was meaningful for them. Some participants referred to changes in the type of activity used to evaluate student learning and others to changes in methods for determining the level of performance. Participants who had been teaching for longer generally spent longer sorting

Table 1. Troublesome aspects for assessment

| Themes | Number of cases | | | | Example quotes |
|--|-----------------|--------|--------|---------|--|
| | Site 1 | Site 2 | Site 3 | % Total | |
| A) Instructor expectations | 1 | 4 | 5 | 29% | <i>I wanted them to go outside their comfort zone, I wanted them to think about social responses, social impacts, political repercussions, or something (203). I'm finding that they're just not getting it. It would be better to work on an engineering project within their own community. That they can tackle, and take on board (308).</i> |
| B) Lack of consistency | 5 | 2 | 2 | 26% | <i>So, in order to make sure we're consistent, the marking guidelines need to be clear. Even with quite a detailed marking guide. I find that there's still ambiguity (306). We have various assessors, in various cities, and one assessor is having a bad day, that means a lot of students (can) fail. From a psychometric point of view, it doesn't make sense what we do (102).</i> |
| C) Differentiation between levels of performance | 4 | 2 | 2 | 23% | <i>...So, I asked myself what are we looking for, what defines 'good'? (101). (We had a) barrier to how we assess using the rubric, we decided that if we could put specific examples of a top mark response, what exceeds expectations looks something like this, the next category response looks something like this (212).</i> |
| D) Student expectations | 0 | 2 | 6 | 23% | <i>After the feedback was released we had so many requests for re-marks because the students couldn't understand why they got the mark they did, even though there was rubric that was very clear (212). (100% exams) It was just too much pressure for the students (202)</i> |
| E) Time constraints/workload | 2 | 3 | 1 | 17% | <i>There is this matter of time. It's not that much time, we don't have that much lecture time, or much teacher time either. It would be ideal to have a whole week just to sit and work with this task (108). You know, to be brutally honest, self-preservation was part of that. Last year there were one-hundred and thirty-six students (204).</i> |
| F) Logistical/technological constraints | 2 | 1 | 2 | 14% | <i>There were major logistic problems... you have to set up an arena with a lot of rooms and they have to be linked. (102). There are 75- 150 students so it's hard for them to really communicate with their peers directly (210).</i> |
| G) Assessment policy | 0 | 0 | 4 | 11% | <i>We're still struggling with the new assessment policy... we've got a long way to go, particularly around assessment (301).</i> |

Transformation

Transformative experiences resulted in profound learning leading to growth of understanding and changes in behavior or attitudes. There were transformational shifts related to the areas of assessment criteria, constructive alignment, constraints and student needs.

Assessment criteria.

Academic staff found it difficult to develop or apply assessment criteria to evaluate student learning, and recognized that they still had much to learn when it came to developing assessment criteria, “so I’m probably still in the midst of that” (203). Difficulties articulating assessment criteria often resulted in transformative experiences, “*the very first time that I developed a rubric I was aware that I am not capturing the wholistic nature of assessment here, I knew it. It turned out to be a huge learning experience*” (302). This learning experience happened when the participant first thought deeply about criteria for the intended learning. It was also transformative to work with peers to make consistent judgements of student performance. In the process of moderation common work samples were shared between markers and discussed to develop an understanding of the criteria. “*We did moderate across the group because they (multiple classes) ran at the same time, so it was really helpful*” (303).

Constructive alignment for meaningful learning.

Lack of constructive alignment can mean that assessment criteria can target skills of little importance to the underlying intention for learning. Such as this example, where the participant had been using an assessment rubric focused on things like numerical commands rather than concept attainment, “*it turned out she was failing on pointless details... things we really don’t find interesting for people to memorize*” (109). The shift in understanding happened because the participant engaged in a conversation with the student, and found that they were able to explain the concepts and how they were applied, yet because of the way the marks were allocated, the student had actually failed. There were multiple occasions when participants realized that their assessment rubric had been working against their intention for deep learning, such as, “*students were looking at those marks rather than looking at what needed to be done*” (304), or “*I wanted to focus on student learning, rather than ‘did they memorize the content the hard way’ ... like surface learning*” (106).

Embracing constraints.

Not all of the assessment change examples had positive outcomes. As educators we know that some of the most profound learning experiences happen when we make mistakes. “*I’d started to wonder what the value was of too much coursework, too many deliverables. When I tried to get them to do that in an assignment, it was a disaster. I did one year get them to do little mini-projects and that was a disaster too*” (203). Workload can be just as much a barrier for academic staff as it is for students. The following comment referring to being beyond capacity “*I said ‘this is nuts, I need help’*” (204). The recognition of time constraints in this case led to the first-time adoption of peer assessment practices.

The assessment policies at site 3 dictated a maximum of three assessments per course.

“We know it works against the principles of learning, which is timely, frequent feedback. If you’re only getting three pieces of assessment then you’re not getting that immediate reinforcement... so this is what we do to get around that...” (305).

By embracing this constraint, the participant re-framed the assessment as a journal that was handed in on continuous basis, with comments and interim marks provided back to the student at certain points during the term.

Meeting the needs of students.

Participants in the study spoke keenly about the desire to provide effective learning experiences for students. Some became aware of limitations within their practice and were motivated to adjust their practices. For example, *“I didn’t have the iteration, giving the time and letting the students re-write based on the feedback from those peer reviews and then hand it in”* (208). This was seen by the participant as a flaw and once they realized this they felt compelled to re-design the form of the assessment. Another participant re-designed their assessment specifically to de-escalate student anxiety, *“It wasn’t just about assessment. It worried them (students), because it’s inquiry and problem-based learning. There’s a whole lot of anxiety that goes with that”* (303). The resulting assessment scaffolded into manageable chunks for the students.

Integration and bounded knowledge

It was evident through the interviews that reflection on teaching and practice was an initial step toward integrating knowledge bounded within a discipline into the wider educational context. The process of reflection was seen to indicate receptiveness in the liminal space. From there, understanding was found to be facilitated through collaboration, professional development or leveraging their professional (disciplinary) background.

Collaboration.

Working with peers was a particularly integrative activity, for example:

We took their work and looked at it and thought, can we adapt it and make it fit us? The working group; bringing people together to sit and discuss our criteria. How we put words on what we look for when we evaluate thesis (101).

Participants typically built from the understanding of others, as indicated by comments such as *“I didn’t do it from scratch for sure, because I did it in consultation with two other faculty members which was really helpful. Together we developed the rubric. I feel like I wanted that, I needed that, chatting with them”* (211).

Professional development.

The majority of participants regularly participated in professional development activities related to learning and teaching. These activities ranged from seminars, workshops or short-courses on site or attending education-based conferences. Participants recounted cumulative experiences, and it was difficult for them to specifically pinpoint how activities influenced their teaching or assessment practice. Comments related to the benefits of professional development in building expertise. Such as *“these kinds of workshops... these are special times”* (202), or *“I learned when it came to pedagogical matters”* (107).

Educational research provided the springboard for adoption of new understanding or practices in assessment.

Last year, based on some input from a conference where somebody had talked about peer assessments... It's something that had not crossed my radar, but I went to this conference and somebody had mentioned it, and I thought 'oh, that sounds interesting' and so I did some more research; how does it work, what do you do. And I thought 'okay' (204).

This participant went on to redesign the assessments for a course to revolve around peer assessment to alleviate the time pressures while providing formative peer feedback. Some of the participants were actively engaged in scholarship of learning and teaching, and worked with others to present or publish in educational forums.

Leveraging professional background.

Comments suggested that disciplinary professional knowledge could provide a basis for the definition of learning goals and aid in recognizing quality performances. *"I know there's many different ways to solve a problem. You need discipline expertise to recognize a correct answer that isn't reflected by the marking rubric"* (310). Academics with professional backgrounds utilized their professional knowledge to inform assessment practice, such as, *"I'm from organization theory, so that's where my background management comes in. So, I had a quite clear picture of that... and what we consider to be a pass"* (103).

Specificity, assessment language and repositioning

Participants primarily demonstrated confident use of assessment specific language. Some recalled a time when their language around assessment had been lacking. For example, a participant described a situation where they were providing feedback to a student who was disappointed with their 'B' grade, and realized a shortfall:

I was kind of stuck and said to the student 'well it doesn't sparkle.' What is 'sparkle?' What does it mean? I remember being very excited about because it helped me define what 'sparkle' was. (Prior to that) I didn't have any of the language (202).

By utilizing the Ideas, Connections, Extensions (ICE) framework (Fostaty Young & Wilson, 2000), they were able to articulate the key components of learning involved in the activity.

Knowledge of terminology enabled participants to describe certain conventions or strategies for assessment, or recognize that others used terms with less specificity, such as: *"I wouldn't even call it a rubric, though X (peer) seemed to think of it as a rubric. It's not a marking rubric but a set of expectations for putting it together"* (204). While examining the quotes, it was important to recognize that some of the participants were non-native English speakers, thus language was sometimes impaired by translation. For example, *"what do you call them in English? Ah, yes examinations, we call them tenta"* (108).

Assessment thresholds

There were a range of conceptual gateways for assessment, but in coming to conclusions, cumulative evidence was sought for all the characteristics of threshold concepts. Figure 2 as

presented above, maps the evidence to define the concepts that could be conclusively described as threshold concepts.

Constructive alignment.

Prior to their conceptual understanding that assessment should be aligned with learning goals and activities, participants tended to focus on content rather than outcomes. The following comment was made when queried about their initial views on assessment, *"it really comes down to the textbook... usually it comes with a whole set of assessment material that is relevant to the topics they're working on"* (207). We can assume that the text book assessments were aligned with the content, but it was content driving the assessment. If principles of constructive alignment were adopted, the intended learning goal would have come first. Once this concept was internalization, it was readily apparent when an assessment fell short of its intended outcome:

The learning outcomes were all geared towards tolerance for ambiguity, decision making under those circumstances, and then using a multiple choice-based format where you've got to come up with the right answer, it was counter to the purpose of the learning outcome (201).

There were many comments referring directly or indirectly to constructive alignment, such as, *"It's very interesting because you have to 'design-up'. So, deciding what it was and then working backward and forward to make sure they (outcomes) lined up"* (303).

Differentiation of standards (and minimum competence).

Before this threshold was crossed, the tendency was for academics is to make on-balance judgements about performance based on their own expectations. Such as this participant who was an early-career academic, *"It's like you have to make an estimation. I think it's hard, too difficult to do explicit criteria, I kind of have this in my head of 'what is important?'"* (111). When an individual is receptive to the concept they adjust their practice, even if it moves away from the norm:

This was way before they started using assessment criteria, but I tried to distinguish, communicate to the students what they needed to perform to get a certain grade when it came to certain aspects I would look for, and so on (107).

The notion of differentiation of standards was something that was irreversible, as suggested by the comment *"So the move to the pass/fail system, was a 'lightbulb' or milestone moment for me personally"* (201). This quote identifying the liminal moment when minimum competency became an essential component of assessment.

Additional observations

Factors limiting change were listed as troublesome in the section above, but there were a number of factors discussed that facilitated or enabled changes in practice. These were not directly attributable to assessment thresholds, but relevant as enablers for change. These enablers were related to resources, support or technology. There were six cases (17%) where changes were enabled through the provision of financial resources or formal allocation of time. In five cases (14%) participants referred to direct help from a teaching support unit, such as assistance from an instructional designer or from a peer in realizing the intended change, and four (11%) of the changes were technology enabled.

Discussion

Many of the participants maintained the status quo and followed existing assessment practices when they began at their institutions. By casting back to a critical episode of assessment change, the pre-liminal space became apparent. Early in their careers many participants were not conscious of the need for transparent, objective assessment of student learning. Troublesome aspects were relatively easy to examine through the narratives, transformation was harder to pinpoint. Observing individuals with different levels of assessment capabilities made conceptual transformations easier to observe. Rowbottom, (2007) questions what makes the concept attainment 'significant', implying that there are degrees of transformation. Participants spoke about the purpose and rationale for their changes, the liminality observed through their action taken. Transformations generally aligned with troublesome areas; getting stuck with a problem prompted action such as going and researching an assessment technique, working collaboratively to develop performance criteria or processes for consistency in assessment. Transformational shifts in understanding occurred through engaged conversations with peers or students, or through consulting educational literature.

As shown in Figure 2, thresholds in assessment were evidenced by new understanding, belief or attituded. There was a close relationship between expectations for meaningful learning and the importance of constructive alignment. Participants found the development of performance criteria to be troublesome, especially when it came to articulating standards for minimum competence in requisite outcomes. Standards have implications for accreditation in professional programs and require a more complex understanding of the systems of assessment and being willing to working with guidelines on standards (Abell & Siegel, 2011). There has been strong opposition to performance standards, as they have also been intertwined with accountability (Deming & Figlio, 2016), and can be worrying when attached to things like performance-based funding (Brown, 2004; de Boer et al., 2015).

The quality agenda in higher education demands a level of accountability not previously experienced in the sector, but changes forced through implementation of policy led to push-back from some participants. Academics need assessment literacy; the skills and orientations as readiness for changes to policy, as well as practical support to overcome logistical or technological constraints. By focusing on key areas to build a foundation of understanding it may be easier for institutional leaders to communicate the purpose and intention of assessment and minimize push back. Direct support can involve expert advice, help drafting or providing feedback on documentation. It is equally importantly however that professional development focus on changing hearts and minds, thus changing institutional culture through building an understanding of the reasons behind and rationale for quality assessment.

Limitations

There are known difficulties in determining threshold concepts. Participants were asked to think back to an experience to uncover the thresholds in question. The episodic narrative method was chosen to avoid the pitfall of asking for speculation, but it is possible that there were gaps or embellishments when participants re-told their stories. It is also important to recognize the interpretations that were made to derive the themes. The author had worked in academic support and assessment for many years, informing the interpretations about the nature of the comments.

Academics fulfil dual roles, they facilitate learning and at the same time make judgements about achievement, like being both coach and referee. Myyry et al., (2019) suggests that this can be problematic and provoke both positive and negative emotions in academics. There were participants who spent much of the introductory section of the interviews venting about things that they were generally unhappy about. Such as not having enough time, specific challenges that they faced in their work and administrative expectations that they felt were unfair. The resulting data were not fully explored due to constraints of word limits here. Further research is needed for a full investigation of institutional culture as it relates to assessment.

Conclusions

In the current regulatory climate there is little doubt that academics need to understand how to effectively assess student learning. The current study utilized episodic narrative interviews at universities in Australia, Canada and Sweden, to explore thresholds in assessment knowledge and attitudes. Troublesome aspects prompted changes in assessment practices. These were observed when there was a mismatch between instructor expectations and student performance, when there was a lack of consistency in assessment and difficulties differentiating performance. There were limiting factors that were troublesome, such as how to deal with constraints in time, workload, logistics or technology. Student expectations and assessment policies were particularly problematic at one of the institutions. A belief in meaningful learning, embracing constraints and the desire to meet the needs of students promoted changes to assessment. Findings suggested assessment thresholds of *constructive alignment* and *differentiation of standards* as necessary in promote quality assessment practices. It is recommended that higher education institutions provide professional development in these key areas and recognize the benefit of collaborative processes and the role of educational research alongside disciplinary expertise for integrating knowledge toward threshold capabilities in assessment.

Acknowledgements

I was fortunate to meet with many academics from diverse backgrounds, who gave up their time and openly shared their stories with me. These people were some of the busiest people in their institutions. Thank you all. Thanks also to the anonymous reviewers who provided detailed comments and recommendations. Finally, I am very grateful for the guidance and feedback from Nicoleta Maynard and Katarina Mårtensson. This work would not have been possible without their generous support.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Natalie Simper has a senior-secondary teaching background with experience in curriculum development, large-scale assessment and evaluation. She managed research in teaching and learning outcomes assessment at Queen's University in Canada, with a focus on meta-cognition, the development of cognitive skills and teamwork. Natalie is currently managing curriculum development at Bond University while completing a PhD in Higher Education externally through Monash University.

References

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In *The professional knowledge base of science teaching* (pp. 205–221). Springer.
- Assessment Definition. (2013, May 15). Retrieved December 1, 2019, from The Glossary of Education Reform website: <https://www.edglossary.org/assessment/>
- Atherton, J., Hadfield, P., & Meyers, R. (2008, June). Threshold concepts in the wild. In *Threshold Concepts: From Theory to Practice Conference, Ontario, Canada*. Retrieved from www.doceo.co.uk/tools/Threshold_Concepts_Wild_expanded_70.pdf.
- Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317–350.
- Barradell, S. (2013). The identification of threshold concepts: A review of theoretical complexities and methodological challenges. *Higher Education; Dordrecht*, 65(2), 265–276.
- Barrie, S., Hughes, C., Crisp, G., & Bennison, A. (2011). *AAGLO Summary 1: The ALTC AAGLO project and the international standards agenda*.
- Basgier, C., & Simpson, A. (2019). Trouble and transformation in higher education: Identifying threshold concepts through faculty narratives about teaching writing. *Studies in Higher Education*, 1–13.
- Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, 1(1), 5–22.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167.
- Boud, D., & Dochy, F. (2010). *Assessment 2020. Seven propositions for assessment reform in higher education* (pp. 1–4).
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-Analytic Methodology and Inferences About the Efficacy of Formative Assessment. *Educational Measurement: Issues and Practice*, 31(4), 13–17.
- Brown, R. (2004). *Quality Assurance in Higher Education: The UK Experience Since 1992*.
- Carmichael, P. (2012). Tribes, territories and threshold concepts: Educational materialisms at work in higher education. *Educational Philosophy and Theory*, 44, 31–42.
- Cousin, G. (2006). *An introduction to threshold concepts*. Routledge.
- Davies, J. (2016). ‘Threshold Guardians.’ In R. Land, J. H. F. Meyer, & M. T. Flanagan (Eds.), *Threshold Concepts in Practice* (pp. 121–134).
- de Boer, H., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottmann, A., ... Vossensteyn, H. (2015). *Performance-based funding and performance agreements in fourteen higher education systems* (p. 164). Enshede, Netherlands: Center for Higher Education Policy Studies.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272.

- Deming, D. J., & Figlio, D. (2016). Accountability in US Education: Applying Lessons from K-12 Experience to Higher Education. *Journal of Economic Perspectives*, 30(3), 33–56.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395–430.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.
- Flick, U. (2000). Episodic interviewing. *Qualitative Researching with Text, Image and Sound*, 75–92.
- Fostaty Young, S., & Wilson, R. J. (2000). *Assessment and learning: The ICE approach*. Portage & Main Press.
- González, J., & Wagenaar, R. (2003). *Tuning educational structures in Europe*. Retrieved from http://eua.be/eua/jsp/en/upload/TUNING_Announcement_Closing_Conference.1084282515011.pdf
- Harris, K.-L. (2009). International trends in establishing the standards of academic achievement in higher education. *An Independent Report and Analysis*. Retrieved from http://melbourne-cshe.unimelb.edu.au/_data/assets/pdf_file/0005/1669343/Intl_trends_standards_Harris_Aug09.pdf
- Hénard, F., & Roseveare, D. (2012). *Fostering Quality Teaching in Higher Education: Policies and Practices* (p. 53) [An IMHE Guide for Higher Education Institutions]. Retrieved from Institutional Management in Higher Education website: <https://www.oecd.org/education/imhe/QT%20policies%20and%20practices.pdf>
- Hounsell, D., Entwistle, N., Anderson, C., Bromage, A., Day, K., Hounsell, J., ... Meyer, E. (2005). Enhancing teaching-learning environments in undergraduate courses. *Final Report to the Economic and Social Research Council on TLRP Project L, 139251099*.
- Irvine, N., & Carmichael, P. (2009). Threshold concepts: A point of focus for practitioner research. *Active Learning in Higher Education*, 10(2), 103–119.
- Jankowski, N., Hutchings, P., Ewell, P., Kinzie, J., & Kuh, G. (2013). The Degree Qualifications Profile: What it is and why we need it now. *Change: The Magazine of Higher Learning*, 45(6), 6–15.
- Kilgour, P., Reynaud, D., Northcote, M., McLoughlin, C., & Gosselin, K. P. (2018). Threshold concepts about online pedagogy for novice online teachers in higher education. *Higher Education Research & Development*, 0(0), 1–15.
- King, C., & Felten, P. (2012). Threshold Concepts in Educational Development: An Introduction. *The Journal of Faculty Development*, 26(3), 5–7.
- Land, R., Cousin, G., Meyer, J. H., & Davies, P. (2005). Threshold concepts and troublesome knowledge (3): Implications for course design and evaluation. *Improving Student Learning—Equality and Diversity*, Oxford: OCSLD. Retrieved from http://161.73.1.13/services/ocslid/isl/isl2004/abstracts/conceptual_papers/ISL04-pp53-64-Land-et-al.pdf
- Medland, E. (2019). “I’m an assessment illiterate”: Towards a shared discourse of assessment literacy for external examiners. *Assessment & Evaluation in Higher Education*, 44(4), 565–580.
- Meyer, J. H. (2010). *Helping our students: Learning, metalearning, and threshold concepts*. School of Policy Studies, Queen’s University, McGill-Queen’s University Press.
- Meyer, J. H. F., & Timmermans, J. A. (2016). Integrated Threshold Concept Knowledge. In R. Land, J. H. F. Meyer, & M. T. Flanagan (Eds.), *Threshold Concepts in Practice* (pp. 25–38).

- Meyer, J., & Land, R. (2003). *Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and*. Retrieved from <http://ecee.colorado.edu/ftep/ETLreport4.pdf>
- Meyer, J., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. Routledge.
- Myyry, L., Karaharju-Suvanto, T., Vesalainen, M., Virtala, A. M., Raekallio, M., Salminen, O., ... & Nevgi, A. (2020). Experienced academics' emotions related to assessment. *Assessment & Evaluation in Higher Education*, 45(1), 1-13.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods, 2nd ed.* Thousand Oaks, CA, US: Sage Publications, Inc.
- Pilkington, R. (2019). Investigating the use of 'professional dialogues' when assessing academic practice: Revealing learning, managing process, and enabling judgements. *International Journal for Academic Development*, 24(1), 47–60.
- Popham, W. J. (1974). *Evaluation in Education: Current Applications*. McCutchan Publishing.
- Quality Assurance Agency for Higher Education. (2018). Subject Benchmark Statements. Retrieved August 19, 2019, from <https://www.qaa.ac.uk/quality-code/subject-benchmark-statements>
- Quinlan, K. M., Male, S., Baillie, C., Stamboulis, A., Fill, J., & Jaffer, Z. (2013). Methodological challenges in researching threshold concepts: A comparative analysis of three projects. *Higher Education*, 66(5), 585–601.
- Reimann, N. (2018). Learning about assessment: The impact of two courses for higher education staff. *International Journal for Academic Development*, 23(2), 86–97.
- Rhodes, T. L., & Finley, A. P. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Washington, DC: Association of American Colleges and Universities.
- Rodger, S., Turpin, M., & O'Brien, M. (2015). Experiences of academic staff in using threshold concepts within a reformed curriculum. *Studies in Higher Education*, 40(4), 545–560.
- Rowbottom, D. P. (2007). Demystifying threshold concepts. *Journal of Philosophy of Education*, 41(2), 263–270.
- Shah, M., & Jarzabkowski, L. (2013). The Australian higher education quality assurance framework: From improvement-led to compliance-driven. *Perspectives: Policy and Practice in Higher Education*, 17(3), 96–106.
- Shinners-Kennedy, D. (2016). How Not to Identify Threshold Concepts. In R. Land, J. H. F. Meyer, & M. T. Flanagan (Eds.), *Threshold Concepts in Practice* (pp. 253–267).
- Shinners-Kennedy, D., & Fincher, S. A. (2013). Identifying threshold concepts: From dead end to a new direction. *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research*, 9–18. ACM.
- Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*, 28(6), 497–526.
- Stensaker, B. m. (2003). Trance, transparency and transformation: The impact of external quality monitoring on higher education. *Quality in Higher Education*, 9(2), 151–159.
- Terrace, H. S., & Metcalfe, J. (2005). *The missing link in cognition: Origins of self-reflective consciousness*. Oxford University Press.
- Tertiary Education Quality and Standards. (2015, January). Higher Education Standards Framework [Australian Government]. Retrieved August 19, 2019, from <https://www.legislation.gov.au/Details/F2015L01639>

- Timmermans, J. A. (2014). Identifying threshold concepts in the careers of educational developers. *International Journal for Academic Development*, 19(4), 305–317.
- Timmermans, J. A., Bruni, C., Gorbet, R., & Moffatt, B. (2018). The flourishing of care in a multidisciplinary Faculty Learning Community. *International Journal for Academic Development*, 23(4), 367–373.
- Tremblay, K. (2013). OECD Assessment of Higher Education Learning Outcomes (AHELO). In *Modeling and measuring competencies in higher education* (pp. 113–126). Retrieved from http://link.springer.com/chapter/10.1007/978-94-6091-867-4_8
- Walker, G. (2013). A cognitive approach to threshold concepts. *Higher Education*, 65(2), 247–263.
- Webb, A. S. (2015). *Threshold concepts in the scholarship of teaching and learning: A phenomenological study of educational leaders in a Canadian research-intensive university* (PhD Thesis, University of British Columbia). Retrieved from <https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0167152>
- Wilcox, S., & Leger, A. B. (2013). Crossing Thresholds: Identifying Conceptual Transitions in Postsecondary Teaching. *Canadian Journal for the Scholarship of Teaching and Learning*, 4(2), 7.