Remaining relevant: Assessment practices in undergraduate education

Katie Dunworth
Curtin University of Technology, Perth, Australia.
K.Dunworth@curtin.edu.au

Abstract: This paper identifies and analyses three broad, connected issues that are impacting on assessment practices in higher education today. The first issue relates to our desire to introduce alternative, more 'authentic' forms of assessment task. These tasks appear to have become particularly desirable as a consequence of the second issue described in this paper, the increased focus in higher education on assessing graduate competencies as well as knowledge. By their very nature, these constructs invite a criterion-based rather than norm-referenced approach to assessment, which leads to the third issue raised in this paper: the need for criteria that adequately and cogently describe those competencies and the requisite standard of attainment. The paper concludes by suggesting areas in which future research may assist us to create a more internally consistent set of processes that will be relevant for students and reflect more accurately our theoretical positions.

Keywords: assessment, undergraduate education

Introduction

In today's world of vast knowledge reserves and instant global communication we have come to expect a lot from our university graduates. To function in an increasingly complex milieu, it has been argued that graduates need to have the discipline knowledge that prepares them for employment, have communication, technical and language skills (Tynjälä, 1998), and be autonomous learners (Boud, 1990), who can think critically (Orsmond, Merry & Reiling, 1997), and solve problems and formulate questions (Segers & Dochy, 2001). In short, graduates should be able to demonstrate a range of cognitive, metacognitive and social competencies (Dochy & McDowell, 1997) in addition to having a firm grasp of their discipline area.

While the *tabula rasa* approach to the dissemination of knowledge has given way to educational theories that acknowledge students as active participants in constructing their own learning, it is also widely accepted that assessment in particular, at least as it is currently utilised, is a vital tool for directing and even driving student learning (Birenbaum, 1997; Brown & Knight, 1994; Dochy & McDowell, 1997; Hanrahan & Isaacs, 2001; Maclellan, 2004; O'Donovan, Price & Rust, 2004; Shepard, 2000; Struyven, Dochy & Janssens, 200; Swanson, Norman & Linn, 1995; Williams, 2005). The term 'assessment' can, however, be ambiguous, because this single expression is

used to describe both the provision of formative feedback of and for learning and the imposition of summative decisions on student achievement. We desire to nurture and support student learning through formative assessment, but at the same time employers, governments and the community demand that we produce externally comprehensible scales of achievement so that judgements can be made about the relative capabilities of large numbers of individual graduates. While one form of assessment undoubtedly influences the other, and while they may even be undertaken simultaneously, they involve separate actions on the part of the academic and are perceived differently by students. We need to be clear about the distinction, because it is graded assignments rather than the many other learning activities in which undergraduates engage which signal to students 'the kind of intellectual work which is valued and thereby influence the way students behave as learners' (Maclellan, 2004:20). In addition, we need to recognise that an emphasis on certification is likely to lead to, or at least encourage, an instrumental approach to learning.

It is often argued that assessed activities should be performance-based and task oriented, and include 'alternative' assessment types such as, inter alia, journals, portfolios, projects, and simulations (see, for example, Wiggins, 1998). Through such tasks, it has been asserted, we engage students at an affective level, we assess competencies authentically, and we foster deep learning. At the same time we tend to condemn 'traditional' forms of assessment, such as examinations comprising timed essays and multiple choice questions, as engendering an outdated approach to learning based on reproduction of texts and memorisation (Boud, 1990; Sambell, McDowell & Brown, 1997). Although there is evidence that some of these more recent forms of assessment are valued by students as more appropriate mechanisms for assessing their knowledge and skills (Klenowski, Askew & Carnell, 2006; Orsmond, Merry & Reiling, 1997; Sambell & McDowell, 1998), we need to be confident that our educational practices are built on solid foundations, and that they are justifiable and internally coherent. In this regard, this paper identifies three issues as areas where further research and debate are needed: the form of alternative assessments, particularly in relation to the notion of 'authenticity'; the focus on competencies, skills and attributes as well as content knowledge; and the use of assessment criteria.

Notions of authenticity

The central characteristic of alternative forms of assessment is their greater 'authenticity' (e.g. Dochy & McDowell, 1997). Not only do such activities have the face validity that traditional tasks lack, in that 'authentic' tasks apparently emulate those which are undertaken in the outside world, but they also appear by their very nature to have construct validity, since they incorporate both subject knowledge and the exercise of those skills or competencies which are considered valuable by both potential employers and the wider community. Yet, as Terwilliger (1997) has comprehensively argued, 'authenticity' is a construct that has not been subjected to scrutiny, is open to interpretation, and presupposes that other forms of assessment are 'inauthentic' and therefore by implication invalid. With its connotative associations with what is real, genuine and natural, the term itself has the power to lull us into assuming that authenticity is synonymous with construct validity, which it is not.

Even the premise that 'authentic' tasks are closer to real life tasks can be called into question. What aspect of real life are they to emulate? Sambell, McDowell & Brown (1997) suggest that, for students in their study, 'the assessment method's novelty lay in the lecturer's attempt to produce an activity which would simulate a real life (often vocational) context' (p.361). Even if we do narrowly restrict ourselves to assessed tasks that imitate those carried out in the course of employment, it is not immediately apparent what they might be, particularly considering the wide choice of careers open to graduates from all disciplines. Ultimately, in the absence of a database compiled by practitioners in the field, how do we determine that, say, a journal or an oral presentation is more authentic a task for a certain profession than the construction of a piece of written discourse intended to persuade (i.e. an essay)?

Furthermore, while the skills may be portable, all tasks are situated in a particular context, and we cannot ignore the fact that the classroom situation is artificial and the tasks contrived. For most educators many 'realistic' activities are not only impractical, but by virtue of being engineered lose the very authenticity they seek and may give rise to construct irrelevant variance in the interpretation of the outcomes. In short, in discussing the types of assessment task that are most appropriate for use in higher education, we need to take care that we are not swayed by intuitions about their face validity and apparent superiority at an affective level at the expense of ensuring that the constructs we purport to measure are those that we do, in fact, measure.

Competencies, skills and attributes

While the literature cited above strongly indicates that graduates should exhibit certain competencies, skills or attributes in addition to their discipline-specific knowledge, it is less clear how we are to determine what these should be. In the UK, the Dearing Committee's 1997 inquiry into higher education identified communication, numeracy and information technology skills as well as cognitive skills and learning how to learn. These are fuzzy superordinate terms, open to interpretation and potentially inclusive of an enormous range of sub-skills. In Australia, the 2000 Commonwealth Government funded report into employer satisfaction with graduate skills has been widely cited as indicating the types of skills and attributes that graduates require. Leaving aside the issue of whether an exclusively employment-focused investigation should be used in this way, the results themselves were problematic. Of 25 (researcher rather than respondent constructed) skills or attributes listed, the five that were most highly rated overall in survey responses were creativity and flair, enthusiasm, capacity for independent and critical thinking, personal presentation and grooming and problem-solving skills. These differed from those skills and attributes identified through qualitative means, which were academic achievement, literacy, numeracy, computing skills, time management skills, communication and interpersonal skills, teamwork and problem-solving skills and comprehension of business processes. The difference in the findings of the two data collection procedures in the same research project demonstrates how slight is the evidence to support the privileging of one skill or competency over another.

Those we identify as desirable by default exclude and marginalise others. To take just one example, traditional assessments have been criticized for their reliance on memory and recall rather than on understanding. Quite apart from the fact that

memory and understanding are complementary rather than dichotomous attributes that incorporate several different characteristics (Entwistle & Entwistle, 2003), the role of memory in almost every profession, as well as our everyday lives, is crucial. From pilots to soldiers the ability to recall key information under pressure of time is a crucial asset, and it could be argued that any form of assessment that through backwash helps students develop this capacity, such as a timed examination, is one that we should endorse.

In addition, competencies and skills are rarely binary constructs that graduates either possess or do not. They are more appropriately identified within a context-related continuum of performance, on which all of us are variously proficient at different stages of our lives and on different occasions, which is why they require descriptive criteria if they are not, ultimately, to become norm referenced. Unfortunately, as will be seen below, this is not a straightforward process, and may have the opposite effect of that intended. Rather than fostering individual development, we may end up with something that 'becomes increasingly bland since the temptation is to include all interests and then reduce them to what is most easily measurable' (Ecclestone, 1999:36). Alternatively, we may seek to be as comprehensive as possible but in the process circumscribe learning and produce in students instead 'uncritical conformity to external injunctions' (Ecclestone, 1999:43).

Even if we are confident that we have identified the competencies and skills which are of most value, it is difficult to assert that one particular assessment task rather than another will be the most likely to lead to the development of those competencies or skills, unless we have a large database of comparative studies from which to draw. Without comparative studies, how do we know that a portfolio assignment leads to, for example, more improved levels of critical thinking than might a discursive essay? This is not to deny that portfolio assignments are useful tools that can have many benefits (Klenowski, Askew & Carnell, 2006), but to suggest that we need to engage in more research in this area before we come to conclusions about those tasks which we should be prioritising.

Finally, as James Gee has pointed out, summative assessment is unjust unless those being assessed have had the same opportunity to learn at an experiential level within the relevant semiotic domain (Gee, 2003). If we decide, for example, that a desirable outcome might be 'the ability to communicate orally in public' and decide to assess that ability summatively through an oral presentation, then the assessment is unfair unless we have actively provided all our students with the opportunity to learn how to do it. Otherwise we entrench disadvantage (Leathwood, 2005), which is again precisely the opposite of our intentions in opening up the assessment process. The provision of such opportunities can be seen as one function of formative assessment, and given that form and content are inextricably linked in terms of ascertaining construct validity (e.g. Birenbaum, 1997), then the more we make use of alternative assessment tasks that have had little opportunity to become normative, perhaps the more we may need to make way for this formative function in our curricula.

Assessment criteria

The goal of inducting new students into their chosen discipline still features in most descriptions of the function of higher education, but it no longer occupies the centre stage it once held, when the university academic was the gatekeeper, an acknowledged expert with an insight into, and ability to judge, the requisite standards of achievement (O'Donovan, Price & Rust, 2004). The changing cultural context and the broadening goals of higher education have brought with them pressure for grading processes that are more accountable, transparent, meaningful, reliable and comprehensible. A major way in which this has been attempted is through the use of explicit assessment criteria.

These have brought with them their own problems. If we wish to measure competencies and skills as well as knowledge, we need criteria that indicate that they are to be measured; in the process ensuring that we do not incorporate so many criteria that we render the assessment process unmanageable. We need to be careful that our descriptors, while attempting to be comprehensive, do not collapse into meaninglessness under the weight of their own verbosity (Ecclestone, 1999; O'Donovan, Price & Rust, 2004), nor in our search for openness become linguistically ever more indeterminate as we seek to be explicit. In addition, as knowledge in higher education incorporates much that is tacit, not easily atomised, as well as that which is explicit, perhaps we should acknowledge that stated criteria are insufficient instruments to permit full dissemination of the discipline's requirements.

Finally, in the interests of reliability, criteria should refer to standards that are understood by all those involved in the measurement process. Unless all staff (or students, where peer assessment is summative) are inducted into a shared understanding of assessment standards, 'different markers [will continue to] award widely varying marks to the same work' (Price, 2005:215). The use of criteria alone does not achieve this aim, since we are limited by the language that is available to describe them. The most abstract a term is, the more open it is to multiple interpretations by staff and students alike. Studies have shown that this applies not only to assessment criteria (see e.g. O'Donovan, Price & Rust, 2004), but also to the rubric of assessment tasks (Chanock, 2000) and the language of feedback. While it may have been ever thus, there is a danger that we may be deluded into automatically equating explicit criteria with reliability and transparency, when this is not the case.

Conclusion

The intention of this paper has been to examine some of the inconsistencies that currently exist with regard to assessment practices in higher education, with a view towards drawing us closer to an internal coherence by acknowledging the dichotomies and suggesting areas where future research will be fruitful. We now have a wide range of assessment options at our disposal. If we select those that most reflect our goals, permit the most valid and reliable interpretations of learning, engage our students and delimit rather than constrain ways of understanding, then higher education will remain relevant to future generations of undergraduates.

References

Dunworth, Catherine (2006) Remaining relevant: assessment practices in undergraduate education, in *HERDSA 2006 International Conference*, University of Western Australia, Perth, 10-12 July 2006.

ACNielsen Research Services (2000). *Employer satisfaction with graduate skills*. Commonwealth of Australia.

Birenbaum, M. (1997). Assessment preferences and their relationship to learning strategies and orientations. *Higher Education*, 33, 71-84

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15(1), 101-111.

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

Chanock, K. (2000). Comments on essays: do students understand what tutors write? *Teaching in Higher Education*, 5(1), 95-105.

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23(4), 279-298.

Ecclestone, K. (1999). Empowering or ensnaring?: the implications of outcome-based assessment in higher education. *Higher Education Quarterly*, 53(1), 29-48.

Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: the interplay of memorising and understanding, and the development of knowledge objects. *Higher Education Research and Development*, 22(1), 19-41.

Gee, J. P. (2003). Opportunity to learn: a language-based perspective on assessment. *Assessment in Education*, 10(1), 27-46.

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development*, 20(1), 53-70.

Klenowski, V., Askew S., & Carnell, E. (2006). Portfolios for learning, assessment and professional development in higher education. *Assessment and Evaluation in Higher Education*, 31(3), 267-286.

Leathwood, C. (2005). Assessment policy and practice in higher education: purpose, standards and equity. *Assessment and Evaluation in Higher Education*, 30(3), 307-324.

Maclellan, E. (2004). Authenticity in assessment tasks: a heuristic exploration of academics' perceptions. *Higher Education Research and Development*, 23(1), 19-33.

NCIHE (1997). *Higher education in the learning society: a summary report*. London, HMSO.

O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, 9(3), 325-335.

Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education*, 22(4), 357-369.

Price, M. (2005). Assessment standards: the role of communities of practice and the scholarship of assessment. *Assessment and Evaluation in Higher Education*, 30(3), 215-230.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, 23(4), 391-402).

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371.

Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in Higher Education*, 26(3), 327-343.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment and Evaluation in Higher Education*, 30(4), 325-341.

Swanson, D. B., Norman, G. R., & Linn, R. L. (1995). Performance-based assessment: lessons from the health professions. *Educational Researcher*, 24(5), 5-11+35.

Terwilliger, J. (1997). Semantics, psychometrics and assessment reform: a close look at "authentic" assessments. *Educational Researcher*, 26(8), 24-27.

Tynjälä, P. (1998). Traditional studying for examination versus constructivist learning tasks: do learning outcomes differ? *Studies in Higher Education* 23(2), 173-189.

Wiggins, G. (1998). *Educative Assessment*. San Franciso, Jossey-Bass.

Williams, K. (2005). Lecturer and first year student (mis)understandings of assessment task verbs: 'mind the gap'. *Teaching in Higher Education*, 10(2), 157-173.