



REAP International Online Conference 2007 - Case Study: Addressing examination issues in the implementation of criterion-reference assessment

Clair Hughes

The University of Queensland

AUSTRALIA

clair.hughes@uq.edu.au

OVERVIEW

The intervention reported in this case study focuses on the application of criterion-referenced approaches to assessment by examination. The intervention was initiated by the introduction of an Electronic Course (Subject) Profile (ECP) that had been designed to provide institutional consistency in course/subject documentation and to encourage compliance with institutional teaching and learning policies. ECP support for criterion-referenced assessment (CRA), the approach that was specified in policy, included a prompt to upload “the assessment criteria and standards by which a student’s level of achievement will be judged for each piece of work” [emphasis added] (The University of Queensland, 2005). Where assessment tasks took the form of examination, difficulties and surprise in responding to this prompt suggested that examinations were commonly treated differently from other forms of assessment and that they had attained an implicit exemption from policy compliance.

This case investigates the feasibility of compliance with this institutional strategy for assessment from the perspectives of lecturers from two different discipline areas - Law and Occupational Therapy.

INFORMATION ABOUT THE CASE

Participants in this study were all from the University of Queensland, Australia. They included three lecturers from the School of Law, three from the School of Occupational Therapy and the educational consultant from the university academic development unit (ADU) with whom they collaborated. All the reported activities, including an evaluation of the project, occurred throughout the first semester in 2006. The case is an extension and expansion of an assessment project undertaken in the School of Law in 2005 (Hughes, Hinchy & Cappa, 2007).

Throughout the case the terms ‘criteria’ and ‘standards’ are used in accordance with Sadler’s (1987: 194) definitions:

Criterion: *a property or characteristic by which the quality of something may be judged. Specifying criteria nominates qualities of interest and utility but does not have anything to offer, or make any assumptions about, actual quality.*

Standard: *a definite level of achievement aspired to or attained. Standards are about definite levels of quality (or achievement, or performance).*



DESCRIPTION OF THE CASE

This case reports the efforts of lecturers to develop and apply criteria and standards for assessment by examination. All lecturers had previously developed their own ways of implementing the CRA policy with assessment tasks other than examinations, but expressed concerns that:

- the development of criteria and standards or assessment rubrics for examinations was likely to be an additional task to the existing practice of developing detailed marking guides
- the specificity required of rubrics that would serve as adequate marking guides would render them unsuitable for provision to students, particularly for use in preparation for unseen examinations
- marks were adequate feedback on examinations that consisted mainly or entirely of low-inference item types (eg multiple-choice questions) and so the effort of providing additional detail was not warranted.

The small group of lecturers participating in this activity included those who requested assistance from the educational consultant in complying with the institutional assessment policy and others who readily agreed to her request to provide add to her authentic experience as a basis for the provision of sound but realistic advice on this matter to the broader university community, many of whom shared the concerns expressed by case participants.

Case activities varied in duration and outcome from lecturer to lecturer but to some extent all collaborations involved:

- initial discussion between lecturer and the EC
- location/provision of resources by the EC
- drafting of criteria and standards or rubrics by the lecturer
- provision of feedback from the EC
- completion of the assessment rubric by the lecturer
- use of the rubric as the basis of assessment judgements by the lecturer (and tutors where applicable).

Initial discussions were undertaken with Law lecturers prior to the beginning of semester so the criteria and standards developed by two of them were able to be provided to students when classes commenced. Discussion with Occupational Therapy lecturers focussed on the revision of existing material and so these discussions occurred later in the semester. Initial discussions were generally lengthy. In some cases the criteria for the examination had already been identified, but even when this was the case, through the discussion process they were clarified and refined. Interpretations of CRA were also considered as the EC encouraged the development of verbal descriptions of standards in preference to some more commonly used alternatives identified by Sadler (2005) such as one-word descriptors or numerical ranges.

Two resources proved particularly helpful in the articulation of criteria and the development of verbal standards. The first was the revised version of Bloom's Taxonomy (Krathwohl, 2002) which was applied extensively to the analysis of the cognitive demands of examination items (including MCQs) as part of the criteria articulation process. The second was a Law Assessment Framework (LAF), a comprehensive set of generic criteria and



standards that had been developed for use in the School of Law in the previous year (Hughes and Cappa, in press). Though the standards described in this resource contained references specific to the law discipline, much of the material was sufficiently generic to be adapted by lecturers in Occupational Therapy.

Cycles of drafting and feedback were repeated until till the lecturer felt confident that the criteria and standards rubrics would serve as reliable and trustworthy assessment guides. Most lecturers developed assessment rubrics for short-answer (one or more paragraphs) examination types, though a pair of lecturers in Occupational Therapy investigated the feasibility of developing a rubric for a multiple-choice/very-short-answer examination paper.

Evaluative discussions were held within a few weeks of the end of semester to investigate the effectiveness of the rubrics in use and any limitations or problems that had arisen during the process.

RATIONALE IN TERMS OF EDUCATIONAL IDEAS

Why apply a criterion-referenced approach to assessment by examination?

There are sound reasons for focussing attention on the feasibility of applying criterion-referenced approaches to assessment by examination. Despite strong arguments for a shift to more authentic types of assessment (Falchikov, 2005), examinations are still a common experience for many university students. In the case study institution, for example, there are approximately 700 central examination papers set and taken by approximately 70,000 students at the end of each semester, a figure that does not include examinations conducted at the school level. Though this level of use of examinations is partly a result of the persistence of discipline or department traditions (Boud, 1995), anecdotal evidence suggests the influence in recent years of factors including increasing class sizes, efforts to minimise plagiarism, reduced marking time at the end of semester and the often explicit message that being research active is more likely to be rewarded than being assessment innovative. Nicol (2007) also suggests that increases in the use of MCQs can be attributed to an increased availability of automation for delivery, marking and the provision of feedback.

It is timely to question the practice of exempting such large numbers of students from the benefits claimed for criterion-referenced approaches to both formative and summative assessment. The clear articulation of criteria and standards is fundamental to the implementation of many of the formative assessment activities that exemplify the principles of good assessment practice that have emerged from the research literature in recent years. The first of the Eleven principles of good assessment design which underpin the REAP project is to Engage students actively in identifying or formulating criteria. Until a primary principle such as this is addressed, it is unlikely that lecturers will be able to address largely dependent principles relating to the provision of opportunities for self-assessment and reflection (Principle 2), useful feedback (Principle 3), dialogue around feedback provided (Principle 4) or the communication of high expectations (Principle 11). The development of a shared understanding of the criteria and standards to be used for summative assessment judgements is essential for ensuring their consistency, defensibility and transparency (Sadler, 1998). Additional support for a focus on criterion-referenced approaches to examinations is Elton's (1996) contention that favourable examination preparation can influence student motivation and commitment.

Though the lecturers in this case study had applied many of the principles of good assessment to tasks other than examinations, their initial step in moving towards a criterion-referenced approach to examinations was incremental. They therefore focused largely on the feasibility of the approach as applied to summative assessment. The evaluation findings however, suggested that their outcomes extended beyond their relatively modest intentions.



EVALUATION

The evaluation process conducted by the educational consultant incorporated ongoing informal observations, consideration of the outputs of the intervention and discussion of participant perceptions of their experience shortly after grades had been submitted at the end of the semester.

The major outputs of the study were the criteria and assessment rubrics that all Law lecturers and one of the Occupational Therapy lecturers developed as a basis for assessment in the examinations they conducted and their use as the basis for assessment judgements. Table 1, an extract from a set of criteria and standards developed for an Occupational Therapy examination consisting of a small number of short answer questions, is fairly representative of the rubrics produced by those using this form of examination.

Table 1: Extract from a criteria and standards rubric used to assess a short answer examination

Criteria	Unacceptable standard 1 – 12	Acceptable standard 13 - 17	Proficient Standard 18 – 21	Advanced Standard 22 - 25
Development and support of argument in relation to implications of policy and legislation for consumers of mental health services.	Makes statements about implications for consumers that are inaccurate and/or unsupported by evidence or examples	Presents superficial arguments around some obvious key implications for consumers. Few examples or elaborations used in developing arguments in relation to a range of implications for consumers.	Develops logical arguments in relation to key implications for consumers. Some relevant examples and elaboration used in support of arguments.	Insightful and cohesive arguments presented based on consideration of a range of alternatives and evidence. Arguments consistently elaborated and supported by relevant examples.
Understanding of the political, legislative, and ethical issues and factors impacting on mental health service delivery.	Does not identify or inaccurately identifies factors and issues impacting on service delivery.	Identifies obvious key issues and factors impacting on service delivery. Includes some inaccuracy or minor errors of omission	Accurately identifies most key issues and factors impacting on service delivery.	Accurately identifies a comprehensive range of relevant issues and factors impacting on service delivery; shows awareness of significant detail

After the initial discussion phase, two of the Occupational Therapy lecturers decided against the development of an assessment rubric because of the difficulties presented by the form of examination they used. As is common in the health sciences, their examination consisted of a number of pieces of stimulus material, usually clinical cases, followed by a combination of diverse items including MCQs, one-word and short answer questions. Though productive discussion succeeded in identifying the demands of many of these items using the revised version of Bloom's taxonomy, grouping of like items for the purposes of identifying inherent criteria became tedious and unwieldy. In addition, as assessment was done manually, the idea of grouping individual student responses to provide feedback on such demonstrations of learning as their ability to recall facts or to identify relevant clinical information and generate hypotheses was perceived as too daunting to pursue.

However, the four lecturers who did develop and apply assessment rubrics did report a range of generally positive outcomes for themselves and their students.



Student outcomes

Though data were not collected directly from students for the express purpose of evaluating the impact of this project, one of the law lecturers had noticed that ratings for several items on standard course and lecturer satisfaction surveys had improved. These were items that focussed on provision of feedback, an emphasis on thinking rather than memorising and clear understanding of assessment requirements. Though this form of student response could not be directly attributed to the provision of verbal standards, it is conjectured that the detailed rubrics made some contribution to these perceptions.

Three lecturers also noticed that students accepted their exam results more readily and were less inclined to contest or quibble over marks than they had been in the past.

Lecturer outcomes

Lecturers perceived summative assessment judgements based on clearly articulated criteria and standards as being:

- efficient - they allowed faster and more discriminating judgements on short essays and other short answer questions
- defensible - fewer students challenged marks and grades
- transparent - especially in clarifying the distinction between highest (7) and second highest (6) grades that could be awarded - an issue for both case study programs as a consequence of relatively high entry level requirements.

The development of assessment rubrics also influenced lecturer behaviour in several ways. The law lecturers had for the first time included examination criteria and standards in course/subject informational materials at the beginning of the semester. Two lecturers were able to substitute the rubrics for the marking guides that had been used for summative assessment judgements in the past while the other two used them in conjunction with the marking guides they had retained to provide the detailed guidance they felt was needed when using multiple markers. One of the law lecturers also reported a backwash effect on his teaching which had become more explicit and focused on the learning he wished to develop in his students.

Consultant outcomes

From the consultant's perspective, active participation in this and previous case studies developed her professional capabilities in several ways. Formulating verbal descriptions of standards across several disciplines has expanded her ability to describe student learning in meaningful language and has provided a store of personal experience that has enhanced her empathy for those undertaking this extremely difficult task, especially the inexperienced. The anecdotes and examples that are now readily available are useful in establishing her credibility with unfamiliar groups.

CONCLUSIONS

As this case study has demonstrated it is not only feasible to adopt a criterion referenced approach to summative assessment by examination but in several ways this innovation has proved more beneficial than the methods these lecturers had used in the past. The confidence gained through the successful experience may, in time, also encourage lecturers to use the rubrics developed as a basis for formative activities as well. The case study has also proved beneficial in clarifying the difficulties of implementing criterion referenced assessment with some extremely common forms of examination, especially those that incorporate multiple choice or hybrid combinations of items. Findings such as these have highlighted areas requiring ongoing investigation.



REFERENCES

- Boud, D. (1995). Assessment and learning - contradictory or complementary? In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 35-48). London: Kogan Page.
- Elton, L. (1996). Strategies to enhance student motivation; a conceptual analysis. *Studies in Higher Education*, 21(1), 57-68.
- Falchikov, N. (2005). *Improving Assessment Through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. London: Routledge Falmer.
- Hughes, C., & Cappa, C. (accepted for publication). Developing generic criteria and standards for assessment in law: Process and (by)products. *Assessment and Evaluation in Higher Education*
- Hughes, C., Hinchy, R., & Cappa, C. (2007). Applying a Criterion and Standards Approach to Assessment by Examination in Law. In G. O'Neill, S. Huntley-Moore & P. Race (Eds.), *Studies of Good Practices in Assessment of Student Learning in Higher Education*. Dublin: AISHE.
- Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.
- Miller, C. M. L., & Parlett, M. (1974). *Up to the mark: a study of the examination game*. London: Society for Research in to Higher Education.
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education*, 31(1), 53-64.
- The REAP Project. (2007). 11 Principles of good assessment design: Adapted from Nicol and Macfarlane-Dick (2006) and Gibbs and Simpson (2004). Retrieved 16 March, 2007, from <http://www.reap.ac.uk/resourcesPrinciples.html>
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28(2), 147-164.
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D. R. (1998). Letting students into the secret: Further steps in making criteria and standards work to improve learning. *Board of Senior Secondary School Studies* Retrieved 30 October, 2004, from http://www.qsa.qld.edu.au/yrs11_12/assessment/docs/criteria.pdf
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, 30(2), 175-194.
- The University of Queensland. (2005). Handbook of University Policies and Procedures: 3.30.1. Assessment Policy and Practices. Retrieved 8 March, 2007, from <http://www.uq.edu.au/hupp/index.html?page=25109&pid=25075>



This work has been made available as part of the REAP International Online Conference 29-31 May 2007 and is released under Creative Commons Attribution-Noncommercial-Share Alike 3.0 License. For acceptable use guidelines, see <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Please reference as:

Hughes, C. (2007). Case Study: Addressing Examination Issues in the Implementation of Criterion-reference Assessment. *From the REAP International Online Conference on Assessment Design for Learner Responsibility, 29th-31st May, 2007*. Available at <http://ewds.strath.ac.uk/REAP07>

Re-Engineering Assessment Practices in Scottish Higher Education (REAP) is funded by the Scottish Funding Council under its e-Learning Transformation initiative. Further information about REAP can be found at <http://www.reap.ac.uk>
