Criteria, standards and assessment in statistical education

Helen MacGillivray
Queensland University of Technology, School of Mathematical Sciences
Brisbane Q4000, Australia
h.macgillivray@qut.edu.au

1. Introduction

The core roles of assessment in learning are reflected in the variety and extent of the literature on assessment, in both genres and areas. The areas range over general higher education, school education, educational research, and discipline-specific literature, while the genres range over qualitative and quantitative research, case studies, reports and overviews, and individual views. Recent emphasis in the tertiary sector on criteria-based assessment has highlighted the difficulties for discipline-based tertiary teachers in absorbing, analysing and reconciling the various nearly disjoint bodies of literature on the topic. This is particularly so in disciplines such as statistics, and at the introductory level with large and highly diverse cohorts in backgrounds and capabilities.

This paper is based on part of a research project whose objectives were to analyse, compare and reconcile: discipline-specific work on assessment in statistics and mathematics; general work on criteria and standards-referenced assessment; and information from the development and implementation of criteria and standards-referenced assessment in a senior school system. After summarising the key results of the analysis, the paper outlines possible strategies for criteria and standards in introductory tertiary statistics.

2. Criteria and standards referencing in higher education assessment

The term criteria-referenced assessment (CRA) is frequently used in the literature to cover both CRA and standards-referenced assessment. A criterion is a property, characteristic or attribute, from the Greek word *kriterion* – a means for judging. In the context of assessment, a standard is a level of attainment, or a degree of quality viewed as a prescribed object of endeavour, from the Roman *estendre* – to extend. Applying these to an assessment task, we can say that the criteria are the properties or components of that task within the overall learning and assessment package, and that the standards look at the quality of the work in those components.

The general higher education literature on assessment emphasizes that the main objective should be an integrated package of learning and assessment, structured for student development, and based on the rich expertise, knowledge and experience of teachers. Students also need information on the roles of the components of assessment, formative and summative, within the overall integrated learning and assessment package. *Students study more effectively when they know what they are working towards…Students are anxious to … compare their performance against others*. (James, McInnis, and Devlin, 2002).

Ongoing feedback to students is essential to learning, but it must be feedback that helps students identify their strengths and weaknesses and provides practical, efficient and effective guidance on how students can progress. Effective feedback can consist of a mixture of whole group guidance, with specific comments for individuals. This point is particularly important in the use of criteria and standards, as over-focus on verbal descriptors of criteria and standards can lead to feedback to students just in the words of the verbal descriptors. Such feedback is clearly of little value in helping students know themselves and how to progress their own learning.

Assessment that helps students to develop understanding of themselves as individual learners is clearly of prime importance. Such assessment also facilitates active engagement by students, and helps them understand and value the criteria, standards and methods of assessment and learning (Angelo, 1999).

Over a number of years, staff at Oxford-Brooks University, UK, have worked on a project to develop, implement and evaluate systems of criteria and standards based on verbal descriptors and team meetings in a Business Faculty. The results surprised and disturbed the project team who observe that (O'Donovan, Price and Rust, 2004) *a single-minded focus on explicit articulation* [of standards] *will inevitably fall short of providing … meaningful knowledge of standards and criteria*, and that *over-reliance on explicit* [descriptors] *is as naïve as* [past] *over-reliance on tacit knowledge.* The recommendation from this extensive study is advocacy of *a transfer of knowledge of assessment criteria and standards that encompasses a spectrum of tacit and explicit processes*.

Tacit knowledge or tacit expertise is also referred to as connoisseur's knowledge in the literature. Sadler (1998) describes the tacit expertise that teachers can bring to assessment, and that arises from a combination of deep understanding of the discipline, the context and the cohort, together with expertise in the construct and implementation of learning and teaching strategies specifically designed for that discipline, in that context with that cohort.

Much of the more analytical, research-based literature on assessment refers to the extensive work of Sadler, whose most recent papers (2002, 2003) analyse grading policies that purport to be criteria-based, and whose benchmark paper of 1987 provides a deeply analytical, coherent and cogent basis for consideration of assessment in higher education. Amongst many highly pertinent points, a key point one is *In norm-referenced and criteria-referenced assessment it is the total 'score' which is used for ranking or reporting a level of achievement…in standards-referenced assessment it is the* ***configuration or pattern of performance***. Sadler (1987) advocates a combination of verbal descriptors (to draw attention to salient criteria at different points) and exemplars (such as marked past student work, representative assessment tasks and model solutions), noting that such a combination is unlikely to substitute for tacit knowledge. It is the ongoing striving to achieve this combination that helps to elicit the tacit expertise.

General verbal descriptors without context, framework, exemplars or tacit transfer can act as "warm fuzzies" – harmless unless used at the point of assessment, where they can be dangerous. For example, the following descriptor of the top "standard" in a conventional elementary statistics subject is dangerously misleading for students and the community: *Demonstrates an excellent mastery of statistical concepts; provides excellent explanations to theory-based questions.*

3. School and tertiary: contrasts, similarities and best practice in mathematics/statistics.

The state of Queensland has a well-established and highly regarded school-based, centrally-moderated assessment system for the last two years of high school (See http://www.qsa.qld.edu.au for details). The Queensland senior school syllabuses and their criteria and standards are subject-specific, written by discipline-based committees with no cross-referencing or comparability analysis of criteria across disciplines. Reporting against the criteria and the standards occurs during the two years, with exit standards based on the overall student portfolio of work. Moderation is also subject-specific, occuring at regional and then state levels and evaluating a specified range of student portfolios and the associated assessment package. Comparability queries across subjects are answered through analysis of student results, not through criteria. The challenge is the development, implementation and support of a state-wide system, with consistency in teacher understanding, student opportunity, student development and assessment within each subject across a great diversity of students and schools. The contrast with single semester university subjects or units is clear; no matter how large the tertiary class, assessment is per subject, per university.

However, there are significant parallels with the tertiary context in the good assessment packages in senior mathematics (which includes statistics modules). These demonstrate teacher depth of understanding of the discipline and of its learning; cover the syllabus and build on/link with the formative assessment in learning experiences; and provide opportunity and discriminators across all capabilities. Some best practices in communicating criteria and standards to students at senior school level include giving descriptors of required skills or criteria and standards particular to each assessment task, **when** the task is given out, and using a combination of component marks, comments and criteria for feedback.

Good learning and assessment packages in introductory tertiary mathematics/statistics subjects are integrated, balanced, developmental, purposeful packages with well-structured facilitation of student learning across the cohort diversity. Such packages possess an inbuilt configuration or pattern of performance for standards referencing as required by Sadler (1987). The configuration comes about through a combination of the construct of formative and summative assessment (aligned with student learning across the spectrum appropriate for the purpose and cohort), and the construct of timing, types and weights of assessment tasks. The criteria or characteristics of each component of assessment would be identified as part of the package. Within such a package, provision of standards for each component of assessment and marks or other form of grades combined across the components of assessment according to the known and pre-given weightings, satisfies the **configuration of performance** criteria of standards-referenced assessment. If verbal descriptors of overall objectives or criteria for the whole subject or unit are required, these can be written through a weighted union of the criteria of components of assessment.

For the student information and feedback aspects of good assessment practices, the following observations come from students, including those in mathematics/statistics. They ask for criteria and exemplars for projects and reports. They do not ask for criteria for exams or "ordinary" homework-type assignments, but they do value exemplars of these, including information on relationships of previous exams to the current learning situation. A "ticks in boxes" method is regarded as very poor feedback, and a combination of grades over components of assessment without an explicit weighting schema is also condemned by students.

4. An application to a first year data analysis learning and assessment package

The theme is the development of basic statistical data concepts and tools, and using them within data investigations from start to finish. The structure, examples and learning experiences are built around the planning, obtaining, handling, presenting, understanding, modelling, analysing, and interpreting of significant real datasets in meaningful contexts.

A learning and assessment package for a particular first year unit with approximately 600 students annually, includes

- laboratory practicals analysing real datasets from past student projects using a statistical computer package (formative)
- worksheets with full solutions (formative)
- brief fortnightly practical quizzes, with the best 5 out of 6 contributing 10%
- a whole semester group project in planning, collecting, analysing & reporting in a data investigation in a context of their choice: weight 20%
- a mid-semester test of similar types to the fortnightly quizzes, contributing 10%
- an end of semester exam covering the whole unit with a mixture of single and short comment responses using computer output and real data contexts: weight 60% (or 50% if optional essay improves overall result)
- an optional essay on how statistics revolutionised science in the 20th century, contributing 10% if this improves the overall result.

The assessment package meets the eight goals of Gal and Garfield (1998) in assessment in statistical data analysis. These goals can be classified as assessing the separate phases of statistical data analysis (the tools and building blocks of procedures, concepts and skills) and the synthesis (choosing, using, interpreting and discussing in whole data investigations). The construct of the package integrates a configuration of performance across these eight goals without over-assessing. Hence for best practice in standards-referenced assessment, as discussed above, the test/exam components need information on skills and knowledge required, possibly via exemplars, and marks for questions, and the project and optional essay need exemplars and descriptors of standards against criteria to be provided with the task information.

For details of the own-choice project, its impact on learning and teaching, and examples from over 1200 students projects, see MacGillivray (1998, 2002a, 2002b). Currently, information for

students includes guidance and description of the three criteria of the project, feedback on proposals and progress, access to model projects, past projects and datasets through practical exercises. After final marking, each group receives a written report with comments and marks for each of the three criteria; such feedback meets students' desirable feedback criteria as well as meeting standards-referencing requirements. For example, the second criteria of the project is *Processing data; understanding of data, variables and issues to be explored; exploring, presenting, commenting on data*. A possible schema for standards descriptors for grading on a scale of 0-6 is given in Table 1.

Table 1. Proposed schema for standards descriptors for second criteria of project

Mark	Description of standard
5-6	Correct identification of variables, types and observational subjects. Demonstrated understanding of nature of data and issues to be explored. Correct and judicious selection and use of graphs, plots, tables, with choices presenting most features of the data
3-4	Attempt at identification of variables and subjects. Some understanding of nature of data, with some issues identified. A mixture of correct/incorrect, judicious/non-judicious selection and use of graphs
1-2	Little understanding of data, types of variables and issues. Little attempt at exploring and commenting with very limited or incorrect presentations

REFERENCES

Angelo, T. (1999) Doing assessment as if learning matters most. *Bulletin of the American Association for Higher Education.* http://aahebulletin.com/public/archive/angelomay99

Gal, I., & Garfield, J. (1998) Aligning Assessment with Instructional Goals and Visions. *Proc. 5th Inter Conf on Teaching Statistics* 773-779, International Statistical Institute

James, R., McInnis, C., Devlin, M. (2002) Assessing learning in Australian universities. Melbourne: The University of Melbourne Centre for the Study of Higher Education

MacGillivray, H.L. (1998) Developing and synthesizing statistical skills for real situations through student projects, *Proc. 5th Inter Conf on Teaching Statistics* 1149-1155, International Statistical Institute.

MacGillivray, H.L. (2002a) Lessons from engineering student projects in statistics, *Proc. Australasian Engineering Education Conference* (2002a) 225-230, The Institution of Engineers, Australia.

MacGillivray, H.L. (2002b) One thousand projects, *MSOR Connections* 2(1), 9-13.

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

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

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5, 77-84.

Sadler, D.R. (2002). Ah! … So That's 'Quality'. In P.Schwartz & G.Webb (Eds). *Assessment: Case studies, experience and practice from higher education*. London: Kogan Page.

Sadler, D. R. (2003). How criteria-based grading misses the point. Keynote address to the Effective Teaching and Learning Conference, November 2003, Griffith University, Australia.

RÉSUMÉ

La variété et le degré de la littérature sur l'évaluation de l'éducation reflet ses buts centrales. Récemment, aux écoles supérieures et les universités, on a mis l'accent sur l'évaluation basé sur les critères. Cet a mis en lumière les difficultés pour les professeurs à absorber, analyser et réconcilier ces sections de la littérature variables. Ces difficultés sont évident dans les cours comme les statistiques, surtout au niveau préliminaire avec les étudiants qui viennent des formations divers avec les capacités différents. On rapport sur une partie de la recherche d'un projet pour analyser, comparer et réconcilier ; le travail générale sur l'évaluation avec les standards et critère, l'information du développement et implémentation de tel système dans un système de lycées, et le travail d'évaluation dans les disciplines spécifiques de statistiques et mathématiques. Spécifiquement, on examine les implications pour les analyses statistiques élémentaires à l'université et le développement de la réflexion statistique.