3+3+4 Symposia

Hosted by
香 港 大 學
THE UNIVERSITY OF HONG KONG

March 1, 2010

The 8th Symposium: Standards Based Assessment and Honours Classification

Keynote Speech

What does it mean to assure Academic Achievement Standards?

Speaker: **Professor D. Royce Sadler**
Professor of Higher Education, Griffith University

Sponsored by
UGC 大學教育資助委員會 University Grants Committee

Organized by Heads of Universities Committee

8th 334 Symposium on Standards Based Assessment and Honours Classification
01 March 2010, Rayson Huang Theatre, The University of Hong Kong
A Keynote presentation

What does it mean to assure academic achievement standards?

D Royce Sadler, Griffith Institute for Higher Education, Griffith University, Brisbane.

Abstract: *Quality assurance is a hot topic in higher education worldwide. It has its own distinctive concepts and terminology, and its own literature and journals. In practice, exactly what is to be 'assured' has not always been made clear, nor have the mechanisms by which the proposed processes lead to the desired result. In this presentation, the agenda is restricted to academic achievement standards, with the emphasis on achievement. This is compatible with a focus on learning outcomes. Typical outcomes-based approaches are not sufficient to assure achievement standards, and this presentation will provide an analysis of the nature of the problem and outline a way forward.*

INTERNATIONAL CONTEXT

- 'Assuring' academic standards is high on the hot-issue list for universities worldwide.
- Unless academic standards can be 'assured', a single university (or the system of which it is part) cannot substantiate the quality of its graduates. It is also hampered in its efforts to establish or maintain a national reputation and international competitiveness.
- In many systems there is little if any hard evidence about the academic standards that prevail. Furthermore, nobody seems quite sure what is meant by 'standards', or who should 'own' them.
- Internationally, schemes to assure standards (or, as is said in some jurisdictions, 'secure' them) have gained considerable momentum in recent years.
- In individual countries (e.g. UK, HK, AUS): policy and action is being spearheaded by quality assurance (QA) and other agencies.
    - AUS: AUQA, Bradley Review, ALTC, TEQSA.
    - Groupings of countries: In the OECD, the 'AHELO' project; 'Tuning/Bologna' project in the EU.
- Some approaches being tried:
    - The so-called 'outcomes' approach is already widespread. (More on that later in this talk.)
    - Some schemes (UK especially) rely on benchmarking and external examiners (or reviewers).
    - Being talked about are broad-spectrum tests of graduate attributes, which could be used for scaling students grades from different institutions. These are not common (yet).
- Overall: Most 'solutions' are partial; none are perfect.

A PERSONAL PERSPECTIVE

- Many universities, university systems and QA agencies are strongly in favour of assuring standards, but rarely spell out the desirable characteristics of the desired end result.
- In particular, there is ambiguity about the scope of 'academic standards'. Does it cover everything to do with, say, teaching and learning? Resource levels – including the number and qualifications of academic staff, staff-student ratios, learning spaces, libraries, flexibility in delivery, quality of teaching, student throughput, student satisfaction, … on and on? Student achievement and the grades that represent it?
- When I ask the question (as I do quite often): 'What precisely is the nature of the problem to be addressed and solved?' no crisp answer comes back. To some extent, we are in the peculiar situation of being favourably disposed towards an idea without necessarily being sure what the idea really is.
- I come back to the fundamental question: What does it mean to assure academic achievement standards?
- Here is my answer, in brief:

  > ***Assuring academic achievement standards*** *means being able to guarantee the integrity of the grades that are progressively entered on student transcripts, and the integrity of honours classifications. This implies that the symbolic representations of achievement levels can be relied on to accurately reflect the attainments of students in the course and degree which they complete. When standards are assured, grades and degree classifications are comparable across courses and degree programs within an institution; they hold their value over time; and are comparable with those awarded in other high-quality institutions nationally and internationally.*

- In the rest of this presentation, I confine my attention to academic achievement standards, not the other issues (resourcing, staffing, facilities, teaching quality etc.), although I am quick to admit they are critical to quality in academic teaching and learning.
- My focus is on student ***achievement***. It ***is*** heavily focused on ***end points or outcomes***, and the grades and classifications that are employed to symbolise different levels of achievement. Note that there is nothing in my statement above that is procedurally prescriptive; it is all about the ***goal of the exercise*** and the ***characteristics of the desired end point***.
- The process of assuring quality is open; generally it will involves four steps:
  - Identifying the 'object' or objects of interest (in this talk, achievement and its representation by grades and degree classification);
  - Identifying the dimensions of quality for that object (the desired properties of grades and classifications, listed in my definition above);
  - Deciding on what would constitute valid data on how the system performs on these dimensions; and
  - Evaluating the performance levels reached (especially whether they are satisfactory).

THE DOMINANT MODEL

- The most common model is based on explicitly specifying Intended Learning Outcomes (ILOs). This one example of a class of methods generically called 'Outcomes-based' (William Spady, originally, 1980s onward).
- It has gained a firm place in thinking and practice as part of the large-scale shift in philosophy from teacher-centred to learner-centred teaching.
- This places strong emphasis on the following:

- Explicit specification of intended learning outcomes.
- Organising teaching and learning activities to maximise the likelihood of bringing these outcomes about.
- Ensuring that what is assessed is in accord with the intended learning outcomes.
- Grading student work according to the extent to which the ILOs have been achieved, a process which usually involves explicit lists or statements of criteria and standards.

- The intended outcomes at the assessment stage can be expressed at different levels of specificity. Higher levels of generality can be progressively broken down into finer and finer detail until the level of the criteria and standards relating to a particular assessment task are reached.
- The higher levels of generality are often referred to as 'grade descriptors'. Despite several variants in format, the descriptions of desired outcomes basically set out to specify the minimum requirements for different course grades.
- The finest grain may be expressed as a 'rubric' or criteria-standards matrix, which is a cross tabulation of criteria (listed in one direction), with 'standards' or levels for each criterion set out in the other direction. Somewhat less formal and detailed versions may be called 'marking guides', 'scoring schedules' or 'marking criteria'.

Example 1: Grade Descriptors

- *(Mid level) – Music, Cambridge, Extended Essay*

  80 +*: The essay will display a degree of independent thought, a refined and critical approach to its sources and a wide knowledge of the relevant scholarly field. It will be meticulously presented and written in a clear and engaging manner.*

  …

  60-69*: The essay will be interesting and purposeful, though lacking some of the refinements and maybe also the originality of a first-class piece of work. There will be a good level of presentational accuracy and acumen.*

- Note the qualifiers and hedge words: degree, independent, refined, critical, wide, relevant, meticulously, clear, engaging, lacking some, good.
- There are not standardised terms; what satisfies a 'degree of independent thought' for one marker may not for another. What is 'interesting' to one marker may not be 'interesting' to another.
- At the highest level, sometimes discipline level, negotiation takes place on the actual wordings of grade descriptors (also called 'discipline standards') to reach consensus that suits the context. The explicit aim is to settle on suitable generalised wording which can be interpreted in a wide variety of contexts.
- This very generality makes it a major contributor to the argument as to why discipline standards will not, in the long term, deliver much in assuring achievement standards.

Example 2: Criteria-standards matrix (or rubric)

- Here is a single line from a rubric for a RESEARCH REPORT (from the University of Newcastle, Australia)

Criterion	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Research base including citations and referencing	♦ Little or no recognition of existing research ♦ Faulty referencing	♦ Some use of existing research to provide a base for project ♦ Appropriate referencing with some errors	♦ Good use of existing research to provide support base for project ♦ Appropriate referencing	♦ Effective use of existing research to provide good support base for project ♦ Good referencing	♦ Extremely effective use of existing research to provide a solid support base for project ♦ Excellent referencing

- This demonstrates the same basic problems as grade descriptors, located in the systematisation of levels: 'Little or no', 'some', 'good', 'effective', 'extremely effective'.
- Question: What do these words mean and imply for judgments? How 'solid' does something have to be to deserve that label? And what does solid mean anyway?
- As bare forms, explicit statements may appear to be able to address the requirements for grade-achievement commensurability, but this model does not normally have 'anchorage' in a shared concrete reality. People may agree fully on the verbal descriptors, whether referring to high-level outcomes or specific criteria and standards, but still differ in their judgments about actual student works. This is because these types of words are contextualised differently, and interpreted differently, by different people. *The meanings are always interpreted in context.* That may seem fine in principle (and is inevitable anyway), but it also has fundamental significance when it comes to assuring standards: it means that the *words cannot be trusted with the 'carriage' of standards across lecturers*.
- Fact: In using verbal descriptors of whatever form, the core problem always remains largely unsolved. As an approach to quality assurance of assessment and grades, therefore, the model is structurally weak. This is so no matter how many words we use, which words we use, or how much detail we go into. They are always open to individual interpretation.
- Key examples (exemplars) that illustrate what the words mean can be a great help, because we then have concrete 'objects' to which the words refer.
- Would that then solve the problem? It is a part solution but not a full one, because words which are intended to apply to a class of things will not necessarily fit the exemplars PERFECTLY. However, if words ARE made to fit particular exemplars perfectly, they will then not necessarily fit future cases that will clearly be judged to fall into the same class. There are a number of reasons for this. I provide just two here:
    - The pool of potential criteria for appraising complex works is usually considerably larger than the set that is specified or implied by standards descriptors, so the set criteria actually form a sample from the pool. Other criteria would normally be invoked if the occasion presented itself. To that extent, therefore, insisting on the use of a fixed subset is misleading, and does not represent common ways in which complex judgments are made in real life.
    - Humans are quite able to make judgments about quality or merit on the basis of comparability, which is not the same concept as identicality. We mostly do not insist on identical outcomes from students. (In many situations, too much makes us suspicious.) We want them to exercise originality and creativity. Given that, however, some things cannot be traded off, such the ability to communicate ideas in written responses to certain types of assessment tasks. But by and large, we are able to judge complex student responses holistically, and then provide valid explanations for our judgments. That is the normal order of events, not the other way round.

SUMMARY SO FAR

- A key driver of developments for this outcomes-based approach was two-fold: (a) getting a high degree of alignment among intended outcomes, teaching and learning processes, and what is assessed; and (b) focusing on what students can actually do at the end of a course rather than on teaching from the teacher's perspective. This is part of being student-centred rather than teacher-centred in teaching.
- Despite its popularity in higher education institutions in relation to quality assurance, adopting outcomes-based approaches as the formal institutional or system approach specifically to assuring academic achievement standards takes it into territory where it is bound not to succeed. Logically, such an extension is not justified. This is because of a fundamental flaw in the assumption that academic achievement standards can be adequately captured or expressed in verbal form.
- Do not academics make wide use of explicit text for conveying and communicating standards now? They do. And the reason this is apparently unproblematic is that the statements are, in most cases, applied so that the range covered by the descriptors or standards is made to 'fit' the range of performance in a group of learners. In other words, the words are contextualised to the class, and this makes the essentially of relativistic interpretation.
- What if the counter argument is put to the effect that this does not take place in reality because the markers 'know' what standards are referred to, and do not vary them to accommodate particular classes or years? If that were strictly true, it would mean that an individual marker could be perfectly consistent from occasion to occasion. That would still not address inter-marker or across-course consistency of standards, so whatever standards are applied function as personal standards, because they draw on personal tacit knowledge for interpretation. Words which do not function as independent anchors are intrinsically incapable of providing stable reference points, which are constitutive of true standards.
- Three final comments in passing before leaving this section.
    - If a university formally adopts grade descriptors as its policy but nevertheless retains reserve powers to review and if necessary modify distributions of course grades (by such methods as changing grade cut-offs or statistically rescaling marks), the whole point of the Descriptors approach is undermined. On the other hand, if it abandons the review process, there is no guarantee that the desired end result (which is assured achievement standards) will be achieved. This poses the real dilemma we face.
    - Apart from the problem of interpretation (which in my view is a fatal flaw), the Descriptors approach to standards makes a lot of logical sense, at least superficially. Industrial standards are generally communicated in words, so why should academic achievement standards be any different? Actually, there is a world of difference, an account of which will have to wait until another occasion. But there are two other worries that bother me. First, I have not yet seen a rigorous argument as to why the Descriptors approach should, in principle, be up to task. By what mechanism is it supposed to be able to assure standards across academics, and across courses? Second, if empirical evidence were forthcoming as to its effectiveness, I would feel more comfortable. But for empirical evidence to be possible, there would have to be a credible 'criterion variable'. (Skip this if you are not familiar with that term). The criterion variable should, logically, be the 'true' marks or grades awarded not by individual academic teachers but as agreed to be fair and just by accredited assessors acting in concert. That is the whole point of the exercise.
    - The grade descriptors approaches focus on improving consistency between aims, teaching and assessment; on matching student achievement with stated outcomes;

and on being explicit about 'standards' for the benefit of students. These operations normally take place wholly within individual courses. Although they may lead to improvements within courses, they do pay no attention to ensuring that *grades are reasonably comparable across courses, degree programs institutions or time*. Unless such broader comparability issues are dealt with adequately, the basic conditions for assuring academic achievement standards are not met. There is then some likelihood that separate steps would be taken by a higher authority (such as a faculty assessment panel or an external testing agency) to control grade distributions or scale results using omnibus tests.

THE ALTERNATIVE APPROACH

- Although achievement standards cannot be satisfactorily encapsulated or expressed in propositional (declarative, explicit) form, a great deal of human knowledge can. Scientific journal articles, encyclopedias and instructional manuals are full of it. For that to be possible, the knowledge to be conveyed must be of a certain type, and certain conditions must be satisfied. The literature on knowledge transfer deals to some extent with that problem.
- Some standards can also be expressed in words. I refer to the large number of industrial, pharmaceutical, safety and compliance standards.
- However, academic achievement standards are abstract notions, concepts or ideas that are 'held' by individuals or groups in what was termed 'tacit knowledge' by Michael Polanyi in his seminal work published in 1964. Such standards reside in our heads. We 'know' what high quality work looks like. We know how to ***recognise*** it when we see it. It is part of our knowledge as professionals.
- For the present, I will call these 'personal standards', to reflect the fact that we are not totally sure of the extent to which our standards match the standards held and applied by our colleagues. (If we were sure, neither we nor our institutions would ever be worried about the grades awarded by colleagues, the proportions of different grades awarded, or the shapes of different grade distributions.) Getting commonality of standards is the issue that has to be addressed if academic achievement standards are to be 'assured'.
- Although the standards we hold as tacit knowledge may seem to be loose and vague, this type of knowledge plays a prominent part in both our professional and personal lives. It represents a significant resource at our disposal, and is so significant that it serves as the basis for the alternative model. The challenge is to work out how to harness and share what we know tacitly, and be sensible and practical in the ways we do that. In particular, we need to establish consensus about how our knowledge should be applied so as to assure standards.
- At this point I risk a couple of assertions:
    - The actual practice of doing this is essentially simple, and well within our reach.
    - The most difficult thing may well be to suspend our attachment to some traditional aspects of the way we assess, mark and report achievement, and think through the issues as an intelligent community of scholars.
- The proposed model starts with primary data and builds systematically upwards. By 'primary data' I mean actual work (or works) produced by students in response to assessment tasks. These works provide first-order evidence of achievement, and represent actual outcomes, not *intended* outcomes. We also know that personal standards may differ from assessor to assessor, and so cannot function as broadly based reference points that are stable over time.
- The quality assurance goal is to achieve a high level of consistency on the standards which different academics apply. In many courses taught by several teachers, it is not uncommon

for this to be carried out by 'consensus moderation' on samples of student responses to a single assessment item. (Consensus moderation is a method of particular importance to the approach I and others are in the process of developing, so other approaches to moderation, such as arithmetical averaging of marks, are not discussed here).

- Consensus moderation provides a way in which members of a teaching team can be mutually calibrated. In practice, its scope of application has generally been limited to moderating the marks assigned to responses to a single assessment item. I now describe how a staged extension of that process can serve our purpose.
- The overall structure of the alternative model involves progressively scaling up consensus moderation until it encompasses the expectations and standards in our respective disciplines, fields and professions.
- Throughout, this concept of 'calibration' is central. To the extent that it can be accomplished, academics who are consistent and calibrated against the consensus standards would not necessarily need constant re-calibration. Furthermore, calibration could take place outside the context and time pressures of end-of-semester assessments and marking.
- The quality assurance processes that would be required after initial calibration would involve periodic check calibrations, and probably random checks on grading standards being applied between calibration exercises.

ADVANTAGES OF THIS APPROACH

- It is a data-based concrete environment in which to work, not the abstract environment of statements of explicit statements of outcomes, objectives, criteria and standards. The whole process is direct.
- As an academic community, and speaking generally:
  - We know how to engage in consensus moderation with a view of coming to a joint collaborative view on the 'standards' to be used for grading student responses to particular assessment tasks;
  - We regard this as appropriate action to guard against quite different 'standards' used by academics; and
  - We regard peer review as an appropriate academic activity, and practice it mostly in: (a) the mark-moderation context; (b) peer review of research grant applications; (c) peer review of manuscripts submitted to refereed journals for publication; and (d) collaborative and collegial work on academic policy development in committees. This by no means that the current methods of implementing peer review are perfect. They could certainly be improved.
  - There should be no objection in principle to extending the concept and the practice of peer review to assessment practice.
  - Other approaches to developing consistent achievement standards which are comparable across courses, programs and institutions could be tested by evaluating the extent to which the processes actually result in consistency and comparability. But how would comparability itself be identified? The alternative model tackles that issue directly – by capitalising on the ability of academics to make high-quality professional judgments based on primary evidence.

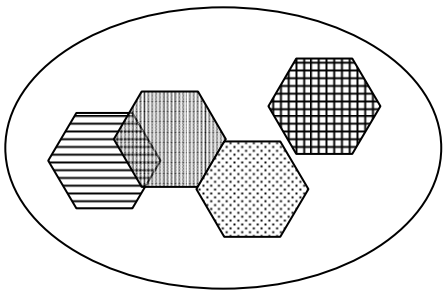
SCALING THIS APPROACH UP

- **LEVEL 1**: This is regular consensus moderation, as commonly practiced by members of a teaching team. The 'object' of scrutiny is the response to an assessment task. The working

data consist of the assessment task specifications, and a sample of student responses to that task. The four questions that need to be asked are:

- Has the task as specified been addressed?
- How well has the work been performed? Where would we rate it (say, on an arbitrary scale, such as 1 - 10)?
- How would we explain or account for our judgment? This necessarily leads to identification of salient criteria; whatever is needed to account for the judgment. The overall judgment comes first; the criteria follow. The accounting for the quality determination is to draw attention to the features or aspects that influence, count towards or detract from the quality of the response.
- What is the consensus mark or grade we feel confident in, all things considered?

- In implementing this Level, members of the teaching team actually come to agreement on two fronts, which they typically do not distinguish in their own minds, and do not need to. They come to a consensus on ***what constitutes quality.*** And they come to consensus on the points along the continuum of quality ***where that continuum should be divided*** for different marks or standards (the grade cut-offs). The cut points are essentially arbitrary, as everyone appreciates, but the object of the exercise is to use them consistently in making appraisal decisions.
- **LEVEL 2:** This retains the fundamental intellectual processes involved in the consensus moderation. The players again need to come to consensus on two fronts: (a) the nature of the 'object' of scrutiny, and (b) the location of the division points along a continuum, which allow for convenient representation of the level reached. Specifically, the 'object' is escalated from 'quality of a single student response' to '***level of achievement over the whole course***'. This is inferred from the evidence available from responses to all ***summative*** assessment tasks. The composition of the rationale for a grading decision changes as well. However, the objective of the exercise remains to reach consensus. Reaching consensus on every single grade is not necessarily required. What is necessary is that the assessors for the course scrutinise all the works completed for summative purposes by a sufficient sample of students for the assessors to become 'calibrated' to one another.
- **LEVEL 3:** This retains the fundamental intellectual processes, but takes another crucial step towards generalisation. The 'object' of scrutiny now consists of the achievements of students in different courses. It is suggested that consensus is reached by academics from a small number of cognate courses, with one course a little more distant (See Figure) to form a 'standards consensus panel'. In the figure, the course positions are intended to convey overlap, contiguity, and proximity. Again, the participants bring to discussion the actual student works that contribute to the summative grade, together with the annotations setting out the rationales for the grading judgments. ***This is the arena in which comparability across courses has to be ironed out.***



- In the event that the process produces grade distributions that appear to be potentially anomalous in some way, the relevant review board refers the matter back to the standards consensus panel for analysis. The panel would again work with concrete achievement

evidence (in the form of student responses, not the marks assigned) to arrive at a conclusion.

LOCKING DOWN ACHIEVEMENT STANDARDS

- This is a substantial topic in its own right, and there is not the time or space to go into detail here. So this is a brief introduction to the way forward.
- First, I reiterate an earlier point. The term 'standards' is used in many different ways for different purposes, including: curriculum 'standards', program 'standards' (for accreditation purposes); teaching 'standards' (for approaches and how they are implemented); and outcome 'standards'. All of these are legitimate areas of application, and all are worthy of attention. The focus here is strictly on *academic achievement standards* which, when applied to student works, lead to marks or scores for responses to assessment tasks, and to grades for achievement in courses. How can we nail these down?
- The initial temptation may be to try to formally define the 'standards' for present use, and keep these definitions for future use. I believe that this is technically impossible, despite the enthusiasm many academics and administrators have for this approach, and their faith in it. Part of the reason for this attachment to the sufficiency of word-based solutions may be that academics as a general rule create, trade with and thrive on 'text', by which broad term I include all textual representations of reality and symbolic forms of communication (including mathematical and other models).
- There may therefore be an uphill route to getting academics to even question the power of the written word to convey morph-resistant meanings (which is what we need for durable academic achievement standards).
- I now limit my attention to *course grades* and specifically do not include marking responses to individual assessment tasks. In practice, the integrity of course grades is where most of the international concern is. However, in the process of doing it (if it is done well), it will scoop up the Level 1 marking issue with it.
- We need some way of 'capturing' or 'arresting' agreed-upon standards so they can be unequivocally understood and retained for future use in grading decisions, for appeals against grading decisions, and for students to access at will. We required the standards to be fixed over time, and not to drift around even incrementally unless a deliberate decision needs to be made to intervene and reset them.
- We start with real, concrete student works or productions, but this time ones that have been graded. If the students do not either produce artefactual outputs or leave some other type of physical record, a separate record such as an audio or visual recording must be created. This may be necessary for musical, dramatic, clinical or similar performances.
- Each grade is to be accompanied by a clear statement as to why the student works in a course 'deserve', on the basis of their quality alone, the grade which has been awarded. It is necessary that both the grades awarded and the justifications for those grades have been scrutinised and agreed upon by a relevant panel. Ideally, there would be a perfect fit between the quality (and qualities) of each work, and the substantive content of the explanatory statement. Note, the judgment comes first; the explanation comes afterwards.
- In constructing the rationale, it is necessary to employ criteria. A justification cannot be created at all without using them. They relevant criteria are, by definition and by function, the properties that matter to the judgment.
- A further point is easy to state, critical to get right, but contrary to much modern practice. We have to invoke criteria *as they are needed.* There is no need to cite or comment on properties that are not instrumental to an appraisal decision. This implies that many (but not necessarily all) criteria can safely remain latent until the need arises for them to be used in particular appraisals. At that point, they need to be made manifest, or explicit. I deliberately

said ‘but not necessarily all’ because some criteria may always matter. But the full set of criteria is not necessarily required in every situation. That cuts across the practice of using fixed sets of criteria for all appraisals. (How would it be possible, before a judgment, to decide whether a particular latent criterion is going to matter?)

- A collected sample of agreed-upon and annotated artefacts (or records) is needed for at least some key grade levels. They form the reference framework for refreshing the calibration of human assessors as and when needed. That is why we have to lock them down.
- There is more to it than this, but that will have to do for now.

WHAT ABOUT THE STUDENTS?

- Here is a hint of how we can discharge our obligations to students.
- It is usually claimed that students need two key types of information if they are to get greatest benefit from assessment events in order to improve their learning. These two types of information are assumed to be indispensable, so that if students are to capitalise on the assessment of their works, we as academic teachers must supply them. The two claims are:
    - Claim 1: Students need to know, at the start of their course, the criteria and standards by which they are going to be assessed.
    - Claim 2: Students need to be given detailed feedback as to the strengths and weaknesses of their attempts.
- These appear laudable, even so obvious that they hardly require justification. The strong expectation if that they will be effective. However, one of the frustrations of teaching is that despite the time and care we put into supplying both of these, the effect on performance is often negligible.
- These two information types actually reflect a particular model of teaching, assessing, and learning from assessment. Again, an alternative is not only possible but also preferable.
- If we step back from current ways of doing things, we can go behind the scene. Yes, we do want students to learn, and to do so effectively and efficiently. How would standards help?
- Now set aside the two conditions of (a) criteria and standards specified in advance, and (b) prompt and comprehensive feedback on performance and ask a deeper question: What do students need to know in order to produce works of consistently high quality? My answer is contained in this compound proposition:
- In order to produce work of consistently high quality, students need to know:
    - What high quality works look like;
    - How to recognize them when they see them;
    - What ‘quality’ consists of in the context, how to account for a particular judgment of quality;
    - How their own work rates on the quality continuum; and
    - How to monitor and control the quality of their own work during its production.
- This knowledge does not come about by being told about it. The two Claims above are limited to ***telling***, which possesses limited utility in teaching students about how to produce high quality work.
- The knowledge students require comes about through immersion in a decision space that is similar to that inhabited by the teacher, through making and accounting for qualitative judgments, and knowing when and how to make changes.

- It comes about later through an intimate understanding of macro-level and micro-level determinants of quality and knowing how to shape the work as a whole through employing small-scale tactics.
- It comes about as students learn not only the technical side of self-assessment but also how to detach themselves from their own productions, whether these are under development or completed. They need to be able to make arms-length judgments about their quality of their emerging or completed work instead of seeing it as an extension of their person.
- This can become a powerful teaching approach, as I have proved myself in my own teaching with positive reactions from students.

CONCLUDING COMMENT

- The direction proposed is a low-risk enterprise, primarily because it:
  - Focuses on the desired end point (academic standards) directly;
  - Works from and retains contact with primary evidence (student work);
  - Employs processes we already accept in principle (moderation, peer review and their extensions);
  - Is easy to adapt to be accessible for students;
  - Retains ownership of both the processes and the standards by the academic staff within the University; and
  - Is likely to satisfy accreditation and quality assurance agencies.

Primary reference

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*. 34, 807-826.

Further reading

Sadler, D. R. (Forthcoming). Beyond feedback: Developing student capability in complex appraisal. Assessment and Evaluation in Higher Education.

Sadler, D. R. (On-line 2009). Fidelity as a precondition for integrity in grading academic achievement. Assessment and Evaluation in Higher Education. URL: http://dx.doi.org/10.1080/02602930902977756 [DOI: 10.1080/02602930902977756]

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (ed.) Assessment, learning and judgement in higher education. Dordrecht: Springer. (Ch. 4, 45-63).

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading in higher education. Assessment and Evaluation in Higher Education, 34, 159-179.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. Assessment and Evaluation in Higher Education, 30, 175-194.

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

(For details of other articles, please see my Griffith University web page.)

This document represents work in progress, and contains substantial original material that has not yet been published. The paper is essentially the Presenter's talk prompts, with some parts elaborated more than others. It does not pretend to be a fully polished paper, and the ideas in it are undergoing continuous development and trial. Permission is granted for the document to be circulated among Hong Kong universities and university authorities for purposes of discussion, provided this notice is included in its entirety. No part of the document in its present form may be cited in publications, or uploaded to any web site that is accessible to the public or to external subscribers.

For the latest information, contact the author r.sadler@griffith.edu.au