Standards and Assessment

Bruce Sutton,
The University of Sydney
November 2009
Version BS191109

Summary

Any university’s reputation is dependent, among other things, on the quality of its teaching and research. In other words, on what we commonly called its academic standards. The principal responsibility of the academic board is to “maintain the highest standards in teaching, scholarship and research” and it is clear why this is an important activity. As well, external stakeholders, particularly our Federal government, are increasing their demands on universities to be held accountable for their academic standards.

Assessment of coursework is a key element in establishing the academic standards achieved by students. This paper examines how assessment at this university may be conceptualised and implemented so that its academic standards can be defined, defended and deployed in ways that give all stakeholders a clear view of the processes and the standards themselves.

Standards are set and referred to in a context and, for a university, the roles of the university and its graduates in its society provide this context. The crucial civic role of universities and their unique contribution to knowledge economies are elaborated to show the real value of high academic standards. As well, the potentially more demanding question of what standards define a graduate with a pass grade is answered. There are other elements to our context that need to be considered. Government agencies such as AQF and TEQSA will inevitably set standards for award courses and disciplines. These will be most useful to us if they establish minima for the higher education sector and allow each university to pitch its standards distinctively above these minima. The standards each university sets for its disciplines will be strongly influenced by the university’s mission and the capability of its staff and students.

To some extent, these are familiar waters, even if we may need to be more explicit than in the past. A substantial change may be necessary in the way we conceptualise our courses, however, if we are to be able to demonstrate to all of our stakeholders that our standards, both in terms of expectations and student attainments, meet or exceed their definitions. Increasingly, the thresholds we must exceed are being cast in terms of learning outcomes that characterise the graduate in the discipline. To quote from the Quality Assurance Agency of the UK: “... a programme specification is not simply an
aggregation of module outcomes; it relates to the learning and attributes developed by the programme as a whole and which, in general, are typically in Higher Education more than the sum of the parts." Our curricula, typical of most Australian universities, are not designed in a way that assures student development towards this holistic goal, nor are our assessment practices generally appropriate to the task of assuring student achievement of the overall goal. Principles of program design and assessment patterns that may be effective are discussed.

Assessment practices will need to be effective, valid and reliable in contexts where divergent student responses to assigned tasks are to be expected. These tasks will generally become more open-ended, fuzzily-defined, integrative and sophisticated in their intellectual challenge as students progress through the course. As well, students need to be assessed solely against standards and criteria that are aligned with the learning outcomes. Two broad classes of assessment practice, criterion-referenced assessment and standards-referenced assessment are analysed, with discussion of which might best suit different assessment requirements. The thrust of the discussion is that, typically, standards-referenced assessment is most likely to provide the desired assessment environment. The final chapter offers a preliminary description of key issues in implementing standards-referenced assessment.
# Contents

Summary ............................................................................................................................................................ i

What do we mean by “academic standards”? ........................................................................................................ 1

Introduction ..................................................................................................................................................... 1

The functions of higher education .................................................................................................................. 2

The civic role of the university ......................................................................................................................... 3

The economic role of universities .................................................................................................................. 6

Standards for our University .......................................................................................................................... 8

Tangible academic standards .......................................................................................................................... 9

Externalities and equivalences .......................................................................................................................... 11

Threshold levels ............................................................................................................................................. 11

Professional accreditation ............................................................................................................................... 12

Graduate attributes ......................................................................................................................................... 13

Learning Outcomes ......................................................................................................................................... 15

Relationships ................................................................................................................................................... 16

Where should standards be specified? Where should assessment occur? ................................................. 19

Assessment and the aim of learning ............................................................................................................... 19

Curriculum design and assessment ............................................................................................................... 20

Towards holistic assessment .......................................................................................................................... 27

Criteria and standards .................................................................................................................................... 29

Criterion-referenced assessment .................................................................................................................... 30

Standards-referenced assessment .................................................................................................................. 32

Drawing it together ......................................................................................................................................... 35

Implementing assessment ............................................................................................................................... 38

Setting standards ............................................................................................................................................ 38

Reviewing standards ....................................................................................................................................... 39

Aggregating marks .......................................................................................................................................... 40

Scaling .............................................................................................................................................................. 41

Non-academic issues ......................................................................................................................................... 42

Forms of assessment ......................................................................................................................................... 42

Moving from norm-referenced assessment ..................................................................................................... 42
What do we mean by “academic standards”?

Introduction

The purpose of this paper is to develop the argument that standards referenced assessment provides a defensible system for maintaining and enhancing academic standards at this University, particularly in coursework teaching. That apparently simple task turns out to be difficult, not because of issues associated with standards referenced assessment, but with the more general notion of academic standards. If standards are to be our reference point, we should have a good idea of what we mean by academic standards. In the voluminous literature on academic quality assurance, this issue is deftly avoided in the main, although we see continual exhortations to maintain and enhance academic standards, without a precise definition of what academic standards are or why we should do it.

Royce Sadler\(^1\) offered, in the academic context, a definition of a standard: “A definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognized measure of what is adequate for some purpose, so established by authority, custom or consensus.” For me, this definition has a number of points that provide fruitful guides: the notion of attainment (as a result of some activity); a position in a spectrum of possibilities, an engagement with a purpose, and the authority of professional “custom or consensus”. It is to the last point we now turn: the question of the value of academic standards. Why should we, and our students, strive for standards any greater than those which prove an acceptable degree of professional competence?

To answer that question, I pose another: What, in modern society, is the role of the university and, particularly, the role of our university? The recently-released Bradley Review\(^2\) offers a contemporary insight.

“Education is at the core of any national agenda for social and economic change. Higher education with its twin functions of teaching and research will make a critical contribution to the nation’s capacity to adapt and to shape the nature of social and economic change.


Higher education is the site for the production and transmission of new knowledge and for new applications of knowledge. It is here that the most highly skilled members of the workforce are educated and here too that the intellectual base for new knowledge-intensive industries is formed.

But higher education in a modern democracy does more than this. By deepening understanding of health and social issues, and by providing access to higher levels of learning to people from all backgrounds, it can enhance social inclusion and reduce social and economic disadvantage. By engaging with scholars from other countries and educating people from other countries, it helps create a nation confident and engaged both with its geographic region and the wider community of nations. By helping sustain and renew other institutions through its capacity to develop knowledge and skills, higher education acts as a cornerstone of the institutional framework of society.

At the same time, its traditional personal development role also remains critical and relevant. Higher education can transform the lives of individuals and through them their communities and the nation by engendering a love of learning for its own sake and a passion for intellectual discovery.

The functions of higher education

The central place of higher education in modern Australia derives primarily from the traditional functions of universities in western countries. Universities must show evidence of internationally competitive performance against all aspects of the following functions.

- Developing and disseminating advanced-level knowledge and skills through teaching and scholarship:
  - to provide for self-fulfilment, personal development and the pursuit of knowledge as an end in itself;
  - to provide the skills of critical analysis and independent thought to support full participation in a civil society;
  - to prepare leaders for diverse, global environments; and
  - to support a highly productive and professional labour force.

- Generating new knowledge and developing new applications of knowledge:
  - by undertaking basic and applied research;
  - by developing high-level research skills; and
  - by exchanging and transferring knowledge and its applications with industry and society.

Through the exercise of these functions and related activities, the higher education system in modern Australia also makes essential contributions to:

- developing and maintaining a just, civil and sustainable society
  - by playing a key role in the development and maintenance of the nation’s legal, economic, cultural and social institutions;
  - by assisting to develop the capacity of Australia to function effectively in the community of nations and of individual Australians to be global citizens; and
The succeeding sections take, to my mind, a utilitarian view of these aspirations. Their proposed implementation by universities to provide assurance of academic standards seems to me to be deeply flawed. I return to this aspect later. However, two main principles emerge in "The functions of higher education" (above): the civic role of the university and its role in the economy of its society.

The civic role of the university

Checkoway\(^3\) noted that “civic engagement is essential to a democratic society”. One of the laments of modern society concerns the shortcomings related to this aspiration.

An analysis of civic participation by young people, 14 years old (typically ninth graders), conducted in 28 countries in 1999, showed high levels in a range of relevant activities by students in Australia and the US\(^4\). However, an examination of civic engagement within the 17-25 year old Australian cohort reported increasing levels of disengagement from fundamental democratic duties such as enrolment and voting\(^5\). One of the issues touched on by the report was the credibility citizens placed on the media, whether print or digital. However, the report did not go further to enquire about their capacity to form rational opinions from their own enquiries. This surely must be an attribute of an educated person.

How can universities contribute? The phrase “informed public debate” springs to mind. John Newman, writing in 1873, captured the role of the university in facilitating participation in society’s affairs\(^6\):

“But a University training... is the education which gives a man a clear conscious view of his own opinions and judgements, a truth in developing them, and a force in urging them. It teaches him to see things as they are, to go right to the point, to disentangle a skein of thought, to detect what is sophistical, and to discard what is irrelevant.”

---


Perhaps, today, we would add nuances related to the contingent nature of knowledge and hence the need for moderation, tolerance and liberalty. However, generally this is a theme that has been taken up by many others: for example, Broady, to whom we return shortly, Anderson and Colby et al.

Anderson puts a different slant on the issue:

"Yet it is precisely the purpose of the university to set the standards of truth seeking for a society, to stipulate the rules that distinguish good sense from nonsense, truth from error, excellence from mediocrity."

The university exerts this influence in a number of ways: through the manner in which learning and teaching are conducted, the normal academic discourse between colleagues both here and in other institutions nationally and internationally and in the nature of our public engagements. To me, there is an undercurrent of dependability in all of these transactions and one can think of any number of occasions when members of the public have been influenced by the dependability of either their own reasoning or their assessed quality of information provided by others, for example:

- in local politics, the “children overboard” affair, the missing weapons of mass destruction that impelled us to war;
- the exposure of the risks of thalidomide;
- the climate change debate.

The Oxford English Dictionary defines "dependable" as "that may be depended on; trustworthy, reliable" and "trustworthy" as "worthy of trust or confidence". Taken together, these do not give an impression of infallibility or immutability, but rather of information of a quality on which one’s own actions could be rationally based. It is hard to imagine a functional society in which dependable information has no value. Two examples suffice. In financial terms, Titman and Trueman demonstrated that there was a sound theoretical basis for the belief that “auditor and investment banker quality provides information about the firm’s true value.” This reflects the perceived dependability of the information about the firm and, I believe, can be used as an analogy for a broad range of situations. Levy and Razin also demonstrated that democracies rarely fight each other, largely because of their ability to communicate reliable information, both within and outside their boundaries. That reliability is established, inter alia, by the sort of informed public discourse referred to earlier.

---

If dependability is crucial, then it follows that there is a premium for graduates and staff of any university whose ability to produce knowledge, to lead debate and to influence opinion constructively can be relied upon to produce a resolution of issues, even in the face of new and complicated challenges. While recognizing that not all students or all staff will reach this level of achievement, I propose that it is with those who do that we align the concept of high academic standards.

This concept then encompasses the highest intellectual ability and rigour, possessed by such students and staff in each discipline. It needs not to be seen as immutable, because the challenges are constantly changing and possibly becoming much more demanding. A generation or two ago, society did not have to consider its response to climate change, peak oil or human cloning. Equally, in my opinion, students enter university with a greater awareness of such issues, are potentially better prepared and as a result capable of greater achievement.

This formulation also permits a clear definition of a passing grade. Might I suggest that the minimum standard of achievement of any graduate should be such that his or her advice, judgement or opinion, offered at any time in their career, could be depended upon in the resolution of problems familiar in the context of the discipline and that he or she would be able to identify both the limits of their competence and the expertise required to resolve the problem.

There is a value that can be attached to both the highest and the lowest academic standards. There is a cost associated with poor judgements, whether it be a poor decision to wage war or some more mundane issue. The value of demonstrated competence is therefore at least the avoidance of such costs. The additional value of the highest level of achievement is the value associated with intellectual innovation and breadth of professional vision, in providing solutions to problems previously thought intractable or the synthesis of novel possibilities for opinion or action.

In summary, there is a value to society in its universities achieving academic standards in a range bounded on the one hand by demonstration of dependable competence and, on the other, by the highest levels of intellectual rigour and originality that can practically be achieved.

Marks\(^\text{12}\) considers that academic standards are public goods. Their chief characteristic in this respect is that they are ‘non rival’; that is, any number of people can access or use them simultaneously. The foregoing discussion is consistent with this view. The free-riding he refers to is behaviour by individuals to access the benefit of the reputation for high academic standards without investing in their maintenance. An example: a student fails to contribute to a groupwork assignment, but expects to share in the mark earned by the remainder of the group. There are many other possible examples. The danger is

that if the incidence of free-riding is high, the quality of the standards, both in actuality and perceptually, falls. Thus, there is a value to students, and staff, in ensuring they actively work to maintain high academic standards. The value of the individual is deeply interwoven with the reputation of the institution.

The economic role of universities

Spigeleman\textsuperscript{13}, in an address concerning John Newman, illustrated the growth of knowledge as an economic outcome, by pointing out that real estate, the bastion of wealth until quite recently, is now greatly overshadowed in worth by intellectual property. Intellectual property has a value that reflects the innovative power it can exert on society.

The term “knowledge economy” is recent, appearing possibly for the first time in Peter Drucker’s book in 1969\textsuperscript{14}. Two distinct asset categories can be identified: knowledge and skills\textsuperscript{15}. This is an important distinction. Knowledge is non-rival in that many people can use the same idea simultaneously without compromising the access of others. Knowledge and ideas therefore have an undeniable public dimension. In contrast, skills are rival and individual. Each member of society acquires and deploys, to a greater or lesser extent, skills characteristic of them. Thus, individuals present themselves to the market with specific skill sets and are valued accordingly. These skills are not easily transferred to others, unlike ideas, except by learning. The aggregation of skilled individuals represents society’s human capital.

Knowledge has the characteristic that it is difficult and expensive to generate, but easy and cheap to distribute. Parallel to this, experience shows that generally, the individuals (or individual enterprises) who generate and freely promulgate ideas are poorly rewarded compared to society in general. Hence, the profusion of companies whose basis is intellectual property protected by patents and the like. There is some argument that the move by universities to do the same is inimical to their social mission.

The important corollary of the above is the emergence of private sector generation of knowledge, for example by widely used search engines. One inevitable result of this is that universities no longer enjoy the monopoly on knowledge generation they may once have had. In this context, what is the role of universities in the knowledge economy?


Certainly, universities still conduct research and prepare professional workers, but these are roles which are no longer unique to the university.

Conceicao and Heitor offer a solution to the question. Their analysis of the functions of research and teaching in universities leads to the statement

“We begin by analysing the university function of teaching, which contributes to the accumulation of knowledge, specifically of skills, through the formal process of learning through education, or ‘learning by learning’. This process, following the analysis earlier, is divergent: a university education combines the transmission of codified knowledge by the teachers with the individual characteristics of the students, in a process in which the interpretation of ideas leads to the accumulation of unique skills. Given this situation, each student can profit from these skills in the future.”

I would add that society, as well as the student, will benefit. From their assertion, it follows that there will be subtle differences in the way individuals perceive the world and some may be more predisposed to visualise new conceptual propositions.

I propose that an important aspect of the role of universities in the knowledge economy lies in their ability to expose students simultaneously to the processes of codification and integration while rewarding their rigorous development of individual intellectual skills and understanding\(^\text{16}\). This seems to me to be the true engine room of innovation in the knowledge economy – an economy which, without innovation, will stagnate and stifle.

There is another contribution that a university education can make to the private sector. Broady points out that a common feature of the best academic cultures and the world of business is the need to make the best possible informed decision.

“To listen to some natural scientists, one might suppose that in the humanities and the social sciences anything counts as valid. But in these fields too there are criteria for evaluating the validity of evidence and conclusions that are no less scrupulous, though perhaps less definitive, than those that operate in some of the natural sciences. They too have well-attested procedures for testing evidence, for assessing its weight, for judging whether it has been justly used to support a given conclusion, and so forth. One makes statements within the conventions of rational argument so that not everything that is said is regarded as of equal validity and one can discriminate between evidence and conclusions of differing degrees of acceptability. That is why, even though it may not be possible to subject ‘judgement’ to scientific appraisal, it can still be evaluated rationally. This is precisely what happens in the law, for example, when a judgement is made between two contending interpretations of a case. There the rules that govern what is admissible evidence, whether the evidence is accurate and validly related to the conclusions that are drawn from it, are very stringent indeed; and though legal judgements may sometimes be wrong, there is no denying that in the law a clear and strict procedure has been evolved for deciding which of two opposing judgements is the more likely to be valid, even when there can be no absolute certainty in the matter. For

\(^{16}\)Broady makes the point that there is an asymmetry between knowledge and understanding. One cannot understand a thing without knowing about it, but the reverse is not necessarily true.
even though, in fields that depend upon judgement, we cannot get the conclusive proof of an experimental science, we can discriminate between a worse and a better judgement; and that is certainly well worth doing. But to do so needs the rigour of verbal argument rather than that of scientific methods. In my opinion, it is in developing the capability for argument by tongue and by pen that we may be able to develop an education that is more relevant to the requirements of practice.”

I take this to mean that, wherever ideas are rigorously contested, the intellectual goals and processes are qualitatively similar. In terms of the application of these skills to the larger world, either in the university or embodied in its students, Broady comments:

“The relevance of these observations for practice is clear. One cannot use knowledge effectively unless it is understood. In the practical conduct of war, the military commander seeks from his intelligence officers not simply knowledge of the enemy’s dispositions but also a judicious interpretation, an understanding of that information made in the light of other data, some of which may not be fully consistent. Information alone does not make decisions. Only when that information is properly understood can practical decisions be made. Here again, the requirements of good education coincide with those of useful practice. Hence, an education that is concerned with developing what education ought to develop, namely understanding, is more relevant to practice than one which, in Muir Wood’s words, crams the student or is over-specialized. To understand is to be able to argue; to argue to a conclusion is to be prepared to decide; and deciding judiciously is what is required in good practice.”

This section has identified two ways in which a university contributes to the knowledge economy: it is a unique environment in which individual skills and perspectives are encouraged and it develops the practice of judicious decision-making which allows those individual attributes to be beneficially deployed. The unique role of the university in the knowledge economy is therefore that it teaches the philosophy and practice of knowledge generation. The essential nature of competence and the desirability of high standards of intellectual achievement in its graduates are axiomatic if that role is to be adequately discharged.

**Standards for our University**

The task now is to ask what the foregoing arguments mean in terms of defining academic standards at this University. There are a number of factors which may influence the decision.

The academic standards any higher education institution accepts cannot be divorced from its mission. Our university is embracing, as its statement of purpose, the following:

“We aim to create and sustain a university that will, for the benefit of both Australia and the wider world, maximise the potential of the brightest researchers and the most promising students, whatever their social or cultural background.”

This implies both a broad acceptance of the roles for the university and very capable academics and students. If the highest academic standards which we expect are, as
previously noted, those to which all staff and students can aspire but only a few achieve, the capability of our academic community and its purpose suggest our highest standards of academic achievement will be demanding and be accepted as so by informed commentators both within and outside the university. We may also assume that the definitions we adopt regarding what is routine in our disciplines and which thus define the lowest standards of academic achievement we are prepared to accept, may be appropriately demanding. We return to this point later in discussing accreditation and academic standards.

To the extent that student potential defines the highest standards, we might also expect that standards defined for each discipline will be such as to stretch the best students enrolling in those disciplines. There should be some expectation that emphasis on high standards will attract the most promising students to each discipline, leading, in turn, to continuing elevation of the standards. The common characteristics across the disciplines will be the challenge to the most capable students, the continuous scrutiny of the standards and their modification to maintain that level of academic challenge.

Our standards will therefore be distinctive, characteristic of an elite university and comparable to those used by leading international higher education institutions. Other Australian universities may have different standards, appropriate to their mission, staff and student capabilities.

Tangible academic standards
To proceed, we need a means of making the concept of academic standards more concrete. The following brief description will be elaborated in succeeding sections, but should suffice for the moment. Academic standards were defined by the Graduate Studies Programme of the UK as “Explicit levels of academic attainment that are used to describe and measure academic requirements and achievements of individual students.

Figure 1 Academic standards depend on the institution’s mission and the capability of its staff and students.
and groups of students.\textsuperscript{17} Sadler\textsuperscript{18} describes four ways in which standards might be made explicit: tacit knowledge, numerical scores, narratives and exemplars and develops the argument that the combination of narrative and exemplars may be the optimal method. However, this expression can be very individualistic. The expression is generalised in an environment where all, teachers and students, become expert assessors of quality and consensus is developed by informed conversation. The latter is a familiar skill exemplified by the conversations had, for instance, in selection committees where the final candidate has to be chosen from two roughly equal contenders, or in a grant review panel, where a final application has to be chosen from a pool. In both these examples, criteria of quality in the relevant context are available, exemplars are at hand and experts from different backgrounds can compare qualitatively different candidates or applications through a shared and agreed assessment of quality.

The same principles make academic standards tangible, comparable and transmissible:

- narrative description supported by exemplars;
- calibration by informed conversation among
- expert assessors of quality from related, but not necessarily identical backgrounds.

\textsuperscript{18} See footnote 1.
Externalities and equivalences

It seems to me entirely plausible that the phrase "academic standards" is one of the most commonly used in educational literature. The meaning is therefore somewhat context-specific, so the purpose of this section is to provide an explanation of the most common situations in which we will have to contemplate the meaning of the phrase. It is therefore not an exhaustive review of the literature in the area.

Recently, an environmental scan of standards setting approaches was published\textsuperscript{19}, from which I refer to the following figure (Figure 2).

\textbf{Figure 2} A conceptual framework for setting expectations and measuring academic achievement.

\textbf{Threshold levels}

Our interest is initially focussed on B, subject area descriptions. These typically describe threshold levels for learning outcomes, in essence the minimum acceptable level of academic achievement in that discipline for the university sector. As an example, the Subject Benchmark Statements in the UK are described as "Subject benchmark statements set out expectations about standards of degrees in a range of subject areas. They describe what gives a discipline its coherence and identity, and define what can be expected of a graduate in terms of the abilities and skills needed to develop

understanding or competence in the subject." In the UK, in addition to the threshold level descriptions, two additional levels of performance may be specified:

- threshold performance is the minimum required to gain an honours degree. Students at this level will be able to demonstrate an acceptable level of ability and skills;
- typical performance is that expected of students at the lower/upper second class boundary. Such students will demonstrate definite competence and skills;
- excellent performance is demonstrated by students gaining first class Honours. These students will have a range of competencies and skills at an enhanced level.

Recall that "honours" is in the UK context in these statements.

Another check: when we describe standards, we generally are providing expectations of academic performance. In contrast, when we report standards, we are generally referring to levels of student achievement that have been attained. It is entirely possible for expected standards to remain high and constant, while achievement standards decline, although this would be a worrying circumstance for any faculty.

An issue of debate in which this university must engage is, if such a system of national descriptors of academic standards is imposed on Australian higher education institutions, to what extent are we bound to it? What freedoms exist for us to expect our students to achieve higher standards? I would argue that if national standards are not treated as minima, then opportunities for individual universities to excel at either the institutional or discipline level and demonstrate this, will be lost. In this respect, it is interesting to note that the Graduate Standards Programme in the UK recommended that each institution should clarify their own threshold standards.

Professional accreditation

By my count, thirteen of our faculties have some form of professional accreditation which provides for their graduates a basis for a right to practise. Should we accept, as has been suggested from time to time, accreditation of our courses as a definition of our academic standards in that discipline?

I would argue not, for the following reasons. First, "professional accreditation is primarily concerned with standards of competence, evaluated against criteria of fitness to practice". This argues for them being considered as threshold standards, in that

---

every graduate wishing to practice in that discipline must meet the requirements. In this case, they would be situated either in A or B in Figure 2.

Second, there may be tension between the learning requirements of the profession and those of the university. On the one hand, the profession requires competence in professional skills. On the other hand, as was argued in earlier sections, we believe our graduates should be capable and practised in critical thinking and related skills that allow them to be effective citizens, innovators and leaders in their fields. These two sets of requirements are not necessarily congruent, which can provide rare challenges for faculties. However, no blanket statement can be made as there are cases of accreditation acting to enhance academic standards. As only one example, I note the accreditation process for the Association to Advance Collegiate Schools of Business (AACSB) requires demonstration of continuing academic success by staff, efforts to continuously improve academic outcomes and appropriate demands for rigour and effectiveness in assessment, among others.

Graduate attributes

Our University’s policy describes graduate attributes as:

"Policy Outline"

This policy sets out the generic attributes which characterise graduates of the University of Sydney. Generic attributes go beyond the disciplinary expertise or technical knowledge that has traditionally formed the core of most university courses and describe the qualities that also equip graduates for their role in society and the world of work.

"Background"

The University of Sydney has identified three overarching graduate attributes—Scholarship, Lifelong Learning, and Global Citizenship—which reflect the research intensive nature of the University, its scholarly values in relation to research-led teaching, and the place of its graduates in a global society. These overarching attributes represent combinations of clusters of more specific attributes, which can be interpreted or contextualised differently in different disciplinary domains. These are in turn supported by generic foundation skills and abilities underpinned by basic competencies."

A quick scan of websites shows great similarity: for instance, the University of Queensland notes:

"The statement of Graduate Attributes describes a set of dispositions that are applicable beyond the specific disciplinary context in which students develop them.

---


It is expected that each bachelor degree program will develop the core attributes listed below, along with any specific professional competencies required. The core attributes reflect and build upon the culture of inquiry and innovation that is part of a research university, as well as embedding the need to exhibit effective interpersonal skills and a broad understanding of social and ethical responsibilities.  

The foreword to a recent report on the National Graduate Attribute Project emphasises that many universities, for some time, have shared a similar concept of "desirable qualities in their graduates". This has been expressed in policies such as those quoted above. However, the question we ask here is: are graduate attributes synonymous with academic standards?

The answer, I think, is no. The definition of standards incorporates the concept of levels of attainment or performance. The detailed description of generic attributes does not, operating instead as thresholds. For instance, from our policy:

"Research and Inquiry: Graduates of the University will be able to create new knowledge and understanding through the process of research and inquiry. This might be understood in terms of the following:
- be able to identify, define and analyse problems and identify or create processes to solve them
- be able to exercise critical judgement and critical thinking in creating new understanding
- be creative and imaginative thinkers
- have an informed respect for the principles, methods, standards, values and boundaries of their discipline and the capacity to question these
- be able to critically evaluate existing understandings and recognise the limitations of their own knowledge."

However, there is every reason to expect that statements of graduate attributes should inform and be incorporated into descriptions of academic standards. There is a close alignment between the intent of the graduate attributes statements and the qualities of university graduates that were discussed earlier (pages 3 and 6). This position is also reflective of the location of the reference to graduate attributes in Figure 2 and seems natural but, as Barrie et al note, the history of universities checking to make sure that graduates actually possess the desired qualities at appropriate levels is a chequered one.

---

27 See footnote 26.
Learning Outcomes

The question of whether learning outcomes constitute academic standards is not easily answered as these terms and others are used interchangeably in the literature. Formal definitions include:

“A learning outcome is the specification of what a student should learn as the result of a period of specified and supported study.”

and

“Learning outcomes are statements of what a learner is expected to know, understand and/or be able to demonstrate at the end of a period of learning. They are explicit assertions about the outcomes of learning – the results of learning.”

A description of the Tuning process, a component of the Bologna program, offers:

“In the language of Tuning, learning outcomes and competencies have particular meanings. Learning outcomes are statements of expectation – what the graduate is expected to know, understand and/or be able to demonstrate upon completion of the ‘learning’ (whether that be at the level of unit of study or award). In contrast, competences are what is ‘obtained’ by the student, overall, and their development may not be readily ‘located’ within particular units of study. The competencies of a graduate may exceed the ‘required level’ described by the learning outcomes.

In this sense, learning outcomes in Tuning describe threshold-level expectations.”

To me, a strict interpretation would be that learning outcomes are not academic standards, but are closely related to them. Adam makes the point that the orientation of education systems to learning outcomes is global and probably represents a genuine paradigm shift. We should therefore accept it as persistent and pervasive. Later in this paper, the close relation between the language used in the specification of learning outcomes and for standards in holistic assessment will become apparent. The difference is largely that assessment standards are explicitly concerned with describing how much better or worse than the threshold level a student has performed.

28 See footnote 19, frontispiece.
31 See footnote 19, p.7.
32 See footnote 30.
As well, inspection shows that at higher levels of curriculum design, for example, at the degree level, there is a strong overlap between graduate attributes and the sort of learning outcomes generally encountered.

Finally, learning outcomes, properly defined and attested, should simplify the process of credit transfer between universities. The concern of faculties to which students are transferring is genuinely gaining a clear idea of the student’s capabilities which are not always apparent from a content-oriented subject outline. Clear explication of learning outcomes would help resolve this problem.

**Relationships**

While Figure 2 provides some indications of the relationships between different approaches to specifying academic standards, local structures allow greater specificity. This was foreshadowed in the Bradley Review, Recommendation 19, which proposed, *inter alia*:

"That the Australian Government adopt a framework for higher education accreditation, quality assurance and regulation featuring:

- accreditation of all providers based on their capacity to deliver on core requirements including:
  - an Australian Qualifications Framework with enhanced architecture and updated and more coherent descriptors of learning outcomes;
  - new quality assurance arrangements involving the development of standards and implementation of a transparent process for assuring the quality of learning outcomes across all providers of higher education;

- an independent national regulatory body responsible for regulating all types of tertiary education. In the higher education sector it would:
  - carry out quality audits of all providers focused on the institution’s academic standards and the processes for setting, monitoring and maintaining them. This would include auditing the adoption of outcomes and standards-based arrangements for assuring the quality of higher education;"

Within Australia, the Australian Qualifications Framework (AQF) is presently redefining the structure of awards recognised under its auspices. As well, the new higher education regulatory authority (not yet well defined at the time of writing) is under development, with the obligations outlined in Recommendation 19 of the Bradley Review. The drivers for this are elaborated in Section 4.1 *et seq* of the Bradley Review:

---

33 See footnote 2, page xx. I have omitted some of the elements of this recommendation not germane to the current argument.

"Strengthened accreditation and quality assurance processes are needed to ensure that students receive the best possible education and that employers can have confidence in the quality of education provided to their current or potential employees. Strengthening the sector’s general regulatory, accreditation and quality assurance systems will also enhance Australia’s position in international education.

In addition, given the substantial public funding of the higher education sector, the Australian community needs to be assured that it is getting value for its contribution. A more deregulated and demand-driven funding system involving substantial additional public investment is proposed in Chapter 4.2, in which higher education providers have the flexibility to set their own entry criteria for students. In order to be implemented effectively, these reforms will require a rigorous system of accreditation and quality assurance to ensure that standards are maintained. Given that the standards required in universities underpin quality across the rest of the higher education system, it is imperative that the Australian community has confidence in the standards of its universities and that there is a transparent, national system in place to assure these same standards are required of all providers of higher education.

The current arrangements are complex, fragmented and inefficient."

The theme of ensuring confidence in the learning outcomes of graduates by stakeholders external to universities is clear. It is entirely understandable that perceptions such as unrestrained plagiarism35 and widespread grade inflation36 may have undermined confidence in academic standards.

Recommendation 19 (above), in fact, captures a solution to these problems in which the standards set by government agencies and universities work as a nested, but open-ended, specification of academic standards. In this view, the learning outcomes described in the levels in the revised AQF act as minimum specifications for the award at any level. Congruent with these minima or no less than them are others described in discipline accreditation standards or subject benchmark statements (Figure 3), forming a nest of threshold standards. Academic standards for a given discipline within any university must have lower bounds at least level to the outer thresholds and preferably

above them. The upper bounds should be restrained only by the context of that university.

We have recently discussed the nature of the cultural hero our university should accept as a true reflection of its mission. Models such as the geek, leader, thinker and worker were considered. Given the legitimate diversity of such models and hence missions and of student demographics across the sector, consideration of Figure 1 implies that there should be great diversity in academic standards between universities, above the lower bounds set by the thresholds.

The burden on universities is to be able to define and defend their standards as meeting the above criteria and to be able to demonstrate that their assessment methods genuinely assess student learning against those standards.
Where should standards be specified? Where should assessment occur?

Assessment and the aim of learning

Graduates provide perhaps the most important vehicle for the benefits of a university to be transferred to its society. The importance of an accurate statement of the quality and capability of our graduates so that we can account for our activities to our stakeholders is therefore obvious. In the previous sections, we have described the quality or standards of our graduates as goals to be achieved or aspired to at graduation. The concepts of high standards and passing grades previously articulated (page 5) are effectively visible at completion. The concepts of, for instance, AQF learning outcomes and graduate attributes, equally make sense only at the completion of study.

There is, consequently, a compelling logic that summative assessment should occur immediately prior to graduation so that graduate quality and capability can be holistically evaluated and warranted. Some universities do this: Oxford University for instance. Setting academic standards against which graduate academic achievement can be measured at this stage seems both natural and problematic. The problems are those associated with assessing the high-level attributes used to describe desired qualities in graduates. These offer little opportunity for objective metrics to be used as the basis of assessment. Peter Knight recognised these difficulties and suggested that, instead of universities warranting competence or quality in graduates, graduates should be able to make claims, supported by evidence, about their qualities on which universities offer comment\(^\text{37}\). The standards against which this evidence might be judged would inevitably share semantics with, for instance, graduate attributes or graduate learning outcomes. The process of assessment would be an exercise in professional judgment.

Most, if not all Australian universities have chosen a different route to warranting graduate quality, in which graduate quality is the accretion of continuous specific assessments made throughout the period of study. There are at least two major problems with this model. First, a single overall grade for the degree, as calculated for the GPA or the UK Honours, is simply uninformative when compared against the breadth of graduate qualities we now expect\(^\text{38}\). Yorke also notes that different students may follow quite different trajectories to arrive at the same final grade. Students may also "shop" for the easiest subjects to maximise their grade.

The second problem is more fundamental. This model of success by accretion rests on an assumption that the aggregation of scores from any series of specific assessments throughout candidature for instance, in each unit of study, is a valid holistic assessment


of the graduate that can be considered alongside the description of graduate qualities. I do not believe this assumption has been rigorously tested.

There are, nonetheless, some checks on the assumption in the typical Australian undergraduate degree. Where Honours, whether an appended year or integrated into an undergraduate degree, is awarded on the basis of a substantial work of scholarship or research and eligibility for Honours has to be demonstrated on merit, there is some element of a holistic appraisal of graduate quality. There is also some reliance on the market to provide post-hoc assessment of graduate quality. However, this approach has two major flaws. First, there is commonly a tension in the relative weightings universities and employers place on generic and professional skills. Second, in a market where all universities use the same model for student assessment, employers may face a Hobson's choice.

It is worth asking here whether the planned Australian Higher Education Graduation Statement (AGHES) will provide a sufficiently detailed holistic statement of graduate quality. The answer is probably no, in the form currently planned for the AGHES. The AGHES is based on the European Diploma Supplement, which "describes the content of the qualification and the structure of the higher education system within which it was issued" as well as including a copy of the student's academic record. It therefore has a focus on describing the academic system and is subject to the same flaws in providing a holistic perspective of the student as the traditional transcript.

**Curriculum design and assessment**

The emphasis on learning outcomes noted earlier will have a substantial effect on curriculum design and assessment practices. First, because learning outcomes for a degree or course are specified as goals to be achieved by graduation, there is a natural incentive to design curricula in a top down manner. If we assume that there should be a substantial learning value attached to study at university, it follows that the experience, knowledge and skills possessed by a student at admission will be far inferior to those expected at graduation. The curriculum should therefore allow the student to experience a trajectory of intellectual development through the degree. The individual units of study should themselves be designed to contribute to this trajectory and the ultimate achievement of the learning outcomes specified for the degree.

Curriculum design is therefore likely to be dominated by concern for what the student, at any stage, needs to experience and learn. The whole should have strong elements of coherence and of collective teaching, in which teachers contribute their efforts as needed. The coherence may be greater in professional rather than generic courses, because of external specification of desired skills and knowledge. This approach is in some contrast to traditional curriculum design, where units of study may be offered

---

more as a showcase of a particular lecturer’s discipline, often with the aim of attracting Honours or postgraduate students.

The organisation of a curriculum that aims to assist students to develop towards the expected graduate outcomes might encompass two broad principles. The first is that of the spiral conception of development:

"Bruner (1960) identified one of the functions of the curriculum as the progressive introduction to the learner of concepts and frameworks at increasingly complex levels. Fundamental and structuring ideas would be encountered repeatedly in a spiral process involving the redefinition of fundamental ideas and concepts at evermore sophisticated levels of understanding and application. In Bruner’s words:

... to be in command of these basic ideas, to use them effectively, requires a continual deepening of one’s understanding of them that comes from learning to use them in progressively more complex forms. (Bruner, 1960, p. 13)"

The thrust of this principle is not that students progress through different stages of learning (enactive, iconic, symbolic) through the curriculum, but that the depth of sophistication with which the curriculum is understood and tackled by the student progressively increases from first year to graduation.

The second principle is that the circumstances in which students demonstrate their level of mastery should offer the maximum opportunity to show that they have integrated their learning in a holistic way and are capable of transferring their understanding to new contexts. I am reminded here of a colleague to whom I believe I could attribute the statement "I don’t try to teach them to think like biologists: I try to get them to be biologists."

This is where professional courses may have a natural advantage, in that they lend themselves to the design of curricula that encompass these principles. Crosthwaite et al\textsuperscript{42} provide one example of such a design, exemplified in their Figure 2. In each year of the course, there is a substantial opportunity for students to interact in team projects where they are invited to demonstrate their ability to act as engineers. As the authors note,

"As students move through the programme, the projects progressively increase in the extent of open-endedness, ambiguity, uncertainty, complexity, technical challenge, scale, breadth, creativity and all such factors that contribute to emphasizing and building real professional practices."


Could one envisage an approach to summative assessment within such a curriculum design that gave greater weight to demonstrations of mastery in the integrated, authentic projects and relatively lighter weight to assessment within individual units of study? This would have a number of advantages.

The inclusion of graduate attributes in the scope of both learning and assessment is natural. Indeed, the enabling conception of graduate attributes would be a necessary condition for successful student achievement of the learning outcomes. This is explicitly demonstrated in the pedagogical model of Humphries and Jolly, shown in Figure 4.

![Pedagogical model of Humphries and Jolly (2003).](image)

Boud promotes the engagement of assessment with practice:

> "What is practice and why might it be a useful prompt for considering assessment? In one sense, practice is simply the act of doing something in a particular situation, for example, analysing particular kinds of problems and applying the results to make a change. That is, it is a description of the everyday acts of practitioners, and most

---

a wealth of practice contexts that could be used. At this point, I have suggested a model of assessment in a degree that places greatest summative weight on special integrative units that offer the student opportunities to demonstrate mastery, at an appropriate level, across the breadth of learning outcomes for the degree. This leaves the proposition that other individual units of study that may introduce and enhance key concepts, technical skills and information structures may have a much lighter assessment load. In one real sense, this is familiar territory. Routinely, research students moving into new discipline areas are asked to take or audit graduates find themselves in positions in which they can be regarded as practitioners of one kind or another even if they are not involved in what has traditionally been regarded as professional work. However, analysing problems in the generic sense is not a useful representation of practice. Practice will involve analysing particular kinds of problems in a certain range of contexts. What is important for practice is what is done, by whom and in what kinds of setting."

As Boud points out, the archetypal example of this is apprenticeship, but higher education is not apprenticeship. "The challenge for higher education is to develop knowledge and skills in a context in which the opportunities for practice and the opportunities to see others practice are not restricted, but where high-level competencies are required." Boud lists a number of benefits from considering assessment in the context of practice, including: locating assessment in authentic contexts; establishing holistic tasks rather than fragmented ones; requiring the use of standards appropriate to the task; and, quite germane to this discussion, moving away from an exclusive emphasis on individual assessment in each unit towards development of assessment tasks throughout a program and the linking of activities from different courses.

As noted earlier, professional courses may have an advantage in having a distinct professional profile as their reference. However, we should qualify that statement with two observations. First, it is possible that the emphasis that accreditation may have on acquisition of necessary professional competences will not give sufficient recognition to the development of academic generic attributes (Jill White, personal communication). Curriculum development will have to integrate the latter into the development of the graduate.

Second, developers of generic courses may assert that no such professional development allows definition of useable practice contexts that they could use is identifiable. I would refute that on two grounds. The first of these is that a university course must have, as one of its aims, the intellectual development of the individual. This cannot be achieved by accident, from allowing the student to intellectually graze at random across a landscape of units of study. There must be at the very least some concept of the generic intellectual development that can be discerned and tested at the end of each year. Further, graduates of generic courses find careers as geologists, historians, art critics, economists, physicists, anthropologists and so on. There is, in fact, a wealth of practice contexts that could be used.

At this point, I have suggested a model of assessment in a degree that places greatest summative weight on special integrative units that offer the student opportunities to demonstrate mastery, at an appropriate level, across the breadth of learning outcomes for the degree. This leaves the proposition that other individual units of study that may introduce and enhance key concepts, technical skills and information structures may have a much lighter assessment load. In one real sense, this is familiar territory. Routinely, research students moving into new discipline areas are asked to take or audit
subjects that provide them with essential new skills and knowledge. In many cases, the summative assessment of this is implicit in the assessment of the final thesis. That is, the success of the auditing is seen in the benefit it offers the final work. Can one translate that ethos to a broad range of degrees? The answer to this question must incorporate at least several elements.

What is assessed in individual units and how is this done?
As noted above, most units of study would be enabling, introducing and enhancing key concepts, technical skills and information structures that will be drawn upon in the integrative units. Assessment, if any, in the enabling units would have a primary role in ensuring that students had adequately developed the skills and accurately integrated the concepts into their growing information structures. The content of this is discipline-specific. One could envisage ungraded (pass-fail) assessment of skills acquisition, particularly in performance contexts (human movement, jewellery construction, laboratory practice, etc) and graded assessment of integration of knowledge into information structures. The latter, of course, assumes that knowledge gained in prior units is still current: in other words, all the units of study on offer have an underlying coherence.

Our major generic degrees offer a sense of intellectual development in the majors offered, which, for instance, allow a student to assert a depth of competence in a particular discipline area, say, anthropology. There is opportunity for students to take elective units outside the area of the major, but in the pursuit of an individual intellectual interest. This might allow, for instance, the anthropology major to develop a measure of competence in a language. This raises an interesting question of how academic mastery in the electives needs to be assessed. There must be some assessment, adequate to assure competence. The question may have two possible answers: the demonstration of competence could be woven, in an individualistic way, into the holistic summative assessment or we could take up Peter Knight’s suggestion of allowing students to make claims as to their competence in these areas in addition to the core summative assessment.

How will progression be judged?
In most of our degrees, all units of study have effectively equal weight in judging academic achievement as a basis for student progression through a degree. If the focus of assessment is biased towards special integrative units, then logically, progression ought to be based mainly on success in these. This would give the integrative units somewhat of the characteristic of barrier units if the integrative units were prerequisites for enabling units from the succeeding year. Thought needs to be given to how far the academic levels of enabling and integrative units can be de-synchronised

---

46 See footnote 37.
and what possibilities exist for remediation of integrative units in Summer and Winter Schools.

**Will a changed assessment practice distort learning?**

There is a belief in universities that students focus their learning on what is to be assessed, that the declaration "this isn’t in the exam" means it won’t be studied and "this next bit will be in the exam" is a reliable way of focusing student attention. Gibbs\(^47\) (2006) writes

"Students are strategic as never before and they allocate their time and focus their attention on what they believe will be assessed and what they believe will gain good grades. Assessment frames learning, creates learning activity and orients all aspects of learning behaviour. In many courses, it has more impact on learning than does teaching."

If the greatest summative assessment weight is on the integrative units, will students adequately learn from the enabling units? I suggest there are two broad aspects to a positive answer to this question.

First, there is an apparent hierarchy of assessment practices that support the sort of student learning that leads to the understanding necessary to transfer mastery from the enabling to the integrative units. Gibbs and Simpson\(^48\) report on studies that show better student academic performance from assessment by assignment than by final examination. Scouller\(^49\) reported that preparation for assignments encouraged students to use deep learning approaches compared to multiple choice questions that fostered a surface learning strategy. Hence, the choice of an appropriate assessment task may encourage a constructive approach to learning.

Second, the curriculum should offer motivation to students to learn. A well-designed curriculum that fosters students’ sense of self-efficacy, responds to their interest and gives them a clear sense that the learning is meaningfully related to their goals should elicit a strong, positive motivational response (see, for example, Keller 2008\(^50\), Glynn et al 2005\(^51\)). As well, there is a strong sense of values that should encourage student learning. Nisan (1992) noted:

"The average student in a good school tends to do the work expected of him or her, or at least takes care to devote a minimum level of attention to studies, even when a subject does not arouse high intrinsic motivation and even when rewards and punishment are


not salient. What, then, is the source of such students' hard work? An answer to this question (is)...that the students share the belief of the curriculum designers that the program is desirable and valuable. This perception...has motivational force." 52

Placing a greater summative weight upon assessment of integrative units clarifies what we as academics see as valuable in the curriculum. There is no doubt that we have always been interested in the development of the graduate as a whole individual, but assessment only of individual units has meant that students have had to work out the integration for themselves. The signal they have broadly heard from us is that we did not value it enough to assess it meaningfully.

To summarise this chapter, the growing importance of external perceptions of graduate quality, as demonstrated by students at completion of their studies, places demands on universities to be able to warrant quality in these terms. The problem is deploying a curriculum design with an overall assessment process that allows the university to assure this quality of learning. With the emphasis on holistic review of student achievement, assessment of individual units of study does not offer an adequate assessment process. Nor does a curriculum that offers a measure of integrated assessment to a selected subset of its student cohort, e.g., our Honours, offer the needed assurance. This chapter offers a suggestion for a broad approach to curriculum design that would support and encourage an integrated approach to learning and opportunities for holistic assessment. It is to the latter we now turn.

Towards holistic assessment

The thrust of the first chapter was that our graduates, to give effect to the benefits of their university education, need to enter the world of work, there to assume some role that fulfils their conception of themselves in that world. We should have assisted by offering a learning experience that we believe equips them with skills, concepts and understanding that will allow them to live effectively in that role. Those skills, concepts and understanding combine the generic attributes of graduates with those specific to their chosen discipline. Bok makes the point that it is not only what has occurred in the classroom that is important in the growth of the graduate, but a range of other encounters that encompass what we call "the student experience". All this is represented in the single individual we call the graduate.

In that statement, the difficulty of assessment is laid bare. Any assessment of the quality and capability of our graduates must be capable of assessing skills, concepts and understanding in an individual. The complexity of the situation is compounded not only by the sense of integration of knowledge, skills and experience in any student, but also by the idiosyncratic differences in understanding that will separate them.

At this point, I would like to share a quotation from Joughin:

"One difficulty with assessment as a term in educational contexts is that its usage often departs from how the term is understood in everyday usage. The Oxford English Dictionary (2002) is instructive here through its location of educational assessment within the broader usage of the term. Thus it defines "to assess" as "to evaluate (a person or thing); to estimate (the quality, value, or extent of), to gauge or judge" and it defines assessment in education as "the process or means of evaluating academic work". These two definitions neatly encompass two principal models of assessment: where assessment is conceived of quantitatively in terms of "gauging" the "extent of learning, assessment follows a measurement model; where it is construed in terms of "evaluation", "quality", and "judgement", it follows a judgement model. Hager and Butler (1996) have explicated the distinctions between these paradigms very clearly (see also Boud in Chapter 3). They describe a scientific measurement model in which knowledge is seen as objective and context-free and in which assessment tests well-established knowledge that stands apart from practice. In this measurement model, assessment utilises closed problems with definite answers. In contrast, the judgement model integrates theory and practice, sees knowledge as provisional, subjective and context-dependent, and uses practice-like assessment which includes open problems with indefinite answers. Knight (2007) more recently highlighted the importance of the distinction between measurement and judgement by pointing out the common mistake of applying measurement to achievements that are not, in an epistemological sense,
measurable and noting that different kinds of judgement are required once we move beyond the simplest forms of knowledge. Boud (2007) has taken the further step of arguing for assessment that not merely applies judgement to students’ work but serves actively to inform students’ own judgement of their work, a skill seen to be essential in their future practice.

Assessment as judgement therefore seems to be at the core of assessment, and its immediate object is a student’s work. However, one further step seems needed. Is assessment merely about particular pieces of work or does the object of assessment go beyond the work? Two other definitions are instructive. Firstly, in the highly influential work of the Committee on the Foundations of Assessment, Knowing What Students Know: The Science and Design of Educational Assessment, assessment is defined as “a process by which educators use students' responses to specially created or naturally occurring stimuli to draw inferences about the students' knowledge and skills” (Committee on the Foundations of Assessment, 2001, p. 20).

Secondly, Sadler (in a private communication) has incorporated these elements in a simple, three-stage definition: “The act of assessment consists of appraising the quality of what students have done in response to a set task so that we can infer what students can do, from which we can draw an inference about what students know.”

From these definitions, the irreducible core of assessment can be limited to (a) students’ work, (b) judgements about the quality of this work, and (c) inferences drawn from this about what students know. Judgement and inference are thus at the core of assessment, leading to this simple definition:

To assess is to make judgements about students’ work, inferring from this what they have the capacity to do in the assessed domain, and thus what they know, value, or are capable of doing. This definition does not assume the purpose(s) of assessment, who assesses, when assessment occurs or how it is done. It does, however, provide a basis for considering these matters clearly and aids the discussion of the relationship between assessment and learning.”

Under these circumstances, the validity of assessments that attempt to measure capability is open to question. The complexity of the attributes that need to be measured and the expectation that individual students will most likely give different, but still highly competent answers to the same broad assignments all point to the observation that, generally, measurement is not possible: we are in the domain of professional judgement.

The issues are familiar. Measurement is obvious where a trait such as typing speed, or precision of machining is to be tested. Measurement is obviously farcical in assessing the quality of the work. The idea that students might give

---

quite different answers to the same broad assignment and be judged, nonetheless, to have done work of equal quality is also familiar. Think of Honours research assignments. We oblige students to choose different topics, but the broad assignment is the same. We believe that we can, even so, judge these separate works in terms of their inherent quality.

What I need to do in the remainder of this chapter is tease this analogy apart, so that we may see the conditions under which such an apparently subjective and idiosyncratic activity as judgement can serve both to provide valid assessment and assessment that is, within the restraints imposed by the complexity of the competence we are trying to assess, pretty reliable. Because assessment should assist learning, I need to show that the assessment process motivates students and facilitates their learning. The process, needless to say, also needs to be one in which teachers are keen to be involved and which has the least apparent imposition upon them.

**Criteria and standards**
Academics, in fact, most people, constantly exercise professional judgement. In our lives, we do it when we sit on selection committees to hire new staff, when we promote our colleagues and when we assist granting agencies to decide the best applications to fund. Lamont57 has recently published a description of the peer grant review process in the humanities and social sciences in the US, which describes the strengths and weaknesses of the system. Broadly, issues of personal and disciplinary bias and inconsistent application of criteria are among the sources of weakness that we need to consider. A strength is the collegiality and discourse between panel members that are major components in decision making.

The use of criteria and standards in judgement is inescapable. Earlier (page 1), I gave a definition of standards. To accompany it, I offer the following definition of a criterion: "a basis for comparison; a reference point against which other things can be evaluated."58 In the following discussion, I will differentiate between criteria as the characteristics or attributes of work that form the basis for judgement and standards as the extent to which those criteria have been met. Thus criteria and standards can be applied very mechanistically (and measurably), as in the case of sunglasses. According to AS/NZS 1067:2003 – Sunglasses and Fashion Spectacles, for the criterion of transmittance of ultraviolet radiation in the spectral band 280-315nm, the standard is that the passage of no more than 5% of the incident light through the lens is permitted59.

There are other criteria that apply to these items, which cover other attributes which do not overlap with the criterion quoted above. For each of these, a standard which is either met or not is described, along with a method of measuring that performance. This

standards specification provides a process for ungraded assessment (Pass/Fail) for items that might be described as sunglasses. A pass is required against all criteria for the overall assessment to be a Pass and for the manufacturers to be entitled to label the items as sunglasses.

There may be circumstances in a university education where such unambiguous criteria and standards apply ("student will extract the correct tooth using approved techniques?"), but it is easy to imagine other occasions where neither the criteria or standards exercised in making a judgement of student performance appear to have that level of clarity. Under these circumstances, what are the requirements for an assessment system in which all may have confidence?

- the criteria used in rendering the judgement need to be capable of articulation;
- the standards used in rendering the judgement need to be capable of articulation;
- there needs to be a process by which the standard of performance is judged.

Broadly, two approaches have been used recently to apply these principles to student assessment, depending on whether the problem is approached giving the criteria priority (criterion-referenced assessment – CRA) or the standards (standards-referenced assessment – SRA). Though both are attempting the same problem and hence have similarities, there are differences in process that differentiate them and offer guidance to the circumstances in which either may be the preferred choice.

**Criterion-referenced assessment**

In typical CRA assessment, the criteria are specified *a priori*, then standards are described for each of the criteria, against which performance can be judged. The result of this exercise is often called a marking schema or matrix. The example of the sunglasses could be considered to lie near the unambiguous end of a spectrum of ambiguity: the criteria are unambiguous, non-overlapping or non-interactive, the rules for aggregation are clear (must pass all criteria), the criteria form an exhaustive set, the standards can be precisely stated and performance against them accurately measured.

In the academic situations where conditions like these, or close to them, prevail, then CRA provides a viable model for assessment. This is often referred to as *analytical* assessment.

In these cases, the specification of criteria and standards provide assistance to learning. Because the criteria are specified *a priori*, students can know what the important characteristics of the assessment task are and the standards against which their performance will be judged. This is an important component of student learning, as not only can a high standard of performance be visualised before assessment, but students can easily identify, after assessment, where their performance needs to be improved if further assessment is required.
However, most academic assessment situations lack this sort of clarity. An analysis of the effectiveness of CRA under these conditions has been provided by Sadler, who defines the problem in the following terms:

"This ... is relevant to a specific range of approaches for marking or grading student responses to a particular class of assessment tasks. These tasks all require divergent or 'open' responses from students. This means that there is no single correct or best answer, result or solution. Response formats include term papers, essays, written assignments, field and project reports, seminar presentations, studio and design productions, specialised artefacts, professional performances, clinical consultations, creative works and client interviews. Divergent tasks dominate in a range of disciplines and professional academic programmes, and account for a significant proportion of assessment activity in higher education. They are intended to provide opportunities for students to demonstrate sophisticated cognitive abilities, integration of knowledge, complex problem solving, critical opinion, lateral thinking and innovative action.

Producing a response requires abilities in both design and production, the objective being to allow considerable latitude for creative solution, analysis or expression. There are no formal techniques or recipes which, if followed precisely, would lead to high-quality responses. Divergent responses are referred to as 'works,' regardless of their material form. Constructed responses to examination items may, if they are substantial enough, also qualify as 'works,' but responses to objective test items do not."

Sadler identifies two broad problems with CRA in these circumstances. The first is that preset criteria cannot account for the range of possible divergent answers. The chosen criteria are but a discretionary sample from a much larger population of potential criteria that could be drawn upon to judge the student response. This leads to real problems for assessors. What should their response be to an essay, for instance, that is perceived to be brilliant, but which addresses few of the pre-selected criteria? On the one hand, they could mark it down for that reason, but given the expectation of a range of correct answers, that seems unjust. Alternatively, they could ignore the lack of fit to the preset criteria and award it the grade they feel is justified. However, this violates the fundamental principle of this approach to assessment, that the set of criteria, explained beforehand to students, represents the basis for assessment of performance.

Experience also shows us that students can produce essays that appear to meet all the criteria (use of English sound; mention made of major wars of the period, etc), but which do not give an impression of scholarship, coherence or deep understanding of the question. Again, the assessor faces the question of what mark, one reflecting "hitting" most of the criteria or one reflecting a lack of deep understanding, is the best judgement of the quality of the work.

The second problem lies in the way performance in different criteria is aggregated to arrive at the overall grade. This has a number of subordinate problems. First, in the sunglasses example, it was easy to identify criteria that were discrete, non-overlapping and non-interactive. This is not always the case in the assessment of academic work. In the example case, the rules for aggregation were simple: the item must pass all criteria and failure at one is an overall failure. Rules for combining results from assessment of student work against a range of criteria are rarely simple to specify. Finally, it is not impossible for two different assessors to arrive at the same overall grade, but have quite divergent ideas about performance on the separate criteria.

What has been observed in many cases, is that assessors initially adopt a holistic approach to judging the quality of the work. The overall quality of the work is judged first and marks are apportioned across the criteria to add up to the total. Another issue that confounds the problem is that linguistic expression of standards can fail to adequately describe the nuances associated with different levels of learning. O'Donovan et al describe how "different tutors in the Business School had used the assessment grid, and, consequently, exactly the same grade/level descriptions for a first year undergraduate module and a masters level module, apparently without any difficulty whatsoever." These same authors also observe that increased specification of criteria and standards in CRA is unlikely to solve the fundamental problems.

**Standards-referenced assessment**

Standards-referenced assessment gives priority to holistic assessment of a work against agreed standards of performance. The term, as I use it in this paper, was described by Sadler but has since been embraced in other works. The characteristics of SRA are that it relies on the capacity of teachers to make expert judgements of quality in academic work, it is referenced to standards that can be articulated, examined and defended and which are represented by a combination of verbal description and tangible exemplars and that it rests, at its most effective, on assisting students to develop the same expertise in assessing quality as their teachers.

Teachers can make sound professional judgements about the quality of a work. The exercise of this judgement is central to the way academics work. Sadler gives a defence of the validity of the ability of experts to recognise quality and be able to offer a

---


63 See footnote 1.


66 See footnote 65, page 57.
reasoned defence of this in each particular case, while not necessarily being able to clearly articulate the general case.

Criteria are an essential part of the specification of both the standards and the judgements. General pre-set criteria are an essential part of the rubric specifying a standard. The exemplars chosen provide concrete referents. However, in explaining a judgement, an assessor may, in addition to the pre-set criteria, draw upon other emergent criteria to which the assessor chooses to refer in order to justify, against the standard, the grade awarded to the work. The choice of emergent criteria may vary from case to case. I will describe an approach to the setting of standards in the next chapter, but they meet the definition provided on page 10.

For standards to have credibility, they must be accessible and agreed upon by all who have a stake in their application. The rubric together with the exemplars provides an accessible form. Conversation provides the mechanism for all stakeholders to exchange ideas about them, decide on changes so that they truly represent the desired standard and agree upon them.

Teachers are obvious stakeholders who also bear the responsibility of drafting the standards, so the first level of conversation must be among them. We have all encountered conversations in promotion and appointment committees as an essential part of arriving at a decision. Lamont describes the conversational processes involved in peer-review of humanities grant applications:

"...I have been concerned with how scholar-evaluators serving on peer-review panels come to agreements through interactions. My study examined not only how panelists justify their judgments, but also how the processes and rules of practice set the stage for a sense of legitimacy...Panelists do not simply enact the rule of meritocracy...; they engage in a genuinely social - that is, interactional - micro-political process of collective decision making. They draw emotional and cognitive boundaries between the work they appreciate and the work they do not, and they do so within relationships of exchange and deliberation...Trying to remove subjectivity entirely from evaluation is doomed to failure, because the evaluation process in intersubjective. The panelists' sense of the legitimacy of the process is as tied to unwritten customary rules that they themselves produced (and reproduce) as it is to broader norms of universalism and professionalism."

Lamont goes on to suggest that an understanding of the evaluation process in peer-review may be helpful in other contexts, to which I add the setting of standards. I cannot prove, but suspect, that the diversity of professional interests among academic staff is such that no group we ever assemble to deliberate on standards in a given context will not have some overlap into other related contexts. Shared membership and discourse

---

68 See footnote 57, especially pages 246-7.
could, conceivably, lead to a sharing of standards across and hopefully between faculties.

Teachers also have a role in engaging the wisdom of the world of work in the setting of standards. Particularly for the standards that apply to final, integrative summative assessment at completion of a degree, the engagement of discipline advisory boards in the review of standards may be a constructive bridge to practice.

The crucial conversations, however, are between teachers and students. There are two key objectives to these conversations. First, engagement with the standards allows students to see, explicitly, the quality of work to which they should aspire. My own experience is that students will be keen to view the exemplars, not to copy, but to understand. As well, poor performance by students can be explained, not in terms of what went wrong, but what needs to be improved in their study and understanding, for the quality of their work to improve. Again, exemplars provide strong referents for these conversations. Second, students should be guided to develop their own skills as expert assessors of quality. This is now viewed by many authors as an important step in assisting students themselves to understand where they stand in their current competence in relation to the expectations of performance specified in the standards.

I have used the term "conversations" to cover actual conversations with students, of the sort we typically enjoy with research students, feedback and formative assessment. Let us assume for the moment that all three terms essentially cover the same act: exchange of information between student and some other that assists the student to understand how his or her learning can be improved towards a specified goal.

Nicol and Milligan provide seven principles of good feedback practice, which I summarise from their article below: Good feedback practice

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of reflection and self-assessment in learning;
3. delivers high-quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape their teaching.

---


I would add one other:

8. encourages active and critical, not passive, learning.

The last point is taken from the first of Gee's principles but also aligns with one of the major tenets of recent work by Goodyear and Zenios, which explores the role of discussion between peers and teachers as an essential environment for knowledge-building, the ability "to create new knowledge in an ever changing world which demands recurring refinement of existing frameworks and their subsequent adaptation to new situations and environments".

The active engagement of students in the feedback process has a range of benefits for their learning experience. Price and O'Donovan show how engagement of students in self- and peer-assessment as a feedback process helps their learning. Their work describes a CRA environment, but the approach can be equally well applied to an SRA environment. A thoroughly illustrative example of the power of self-assessment is given by Gibbs.

As well, it is my impression that effective feedback processes should motivate students for a number of reasons. They should have an enhanced sense of self-efficacy and of control of their learning, as well as reduced exam anxiety. The relevance of their studies to their goals beyond graduation should also be apparent. I find it difficult, if not impossible, to believe that a curriculum designed towards overall learning outcomes as described in the previous chapter and the discussions described immediately before, could not achieve the "constructive alignment" promoted by Biggs.

The picture which has emerged for me is one in which learning, in the broadest sense of the word and assessment, in the form of contestation of ideas in discussion or some formative process are inseparable in a continuing process of intellectual development, particularly for the student, but also for the teacher. This has been captured diagrammatically by Rust (Figure 5). His article refers mainly to CRA, but again is eminently applicable to SRA.

**Drawing it together**

I have used the idea that summative assessment of graduate quality and capability is assessment of a complex reality at a high level of integration, where student responses

---

71 See footnote 70.
75 See footnotes 50 and 51
will be divergent, to press the case for holistic judgement as the most appropriate principle. I have also argued that standards-referenced assessment is likely to be the most appropriate manifestation of this principle. I now ask: is SRA also an appropriate form of assessment for individual units of study, or even for assessment elements within units of study?

The answer, as it is so often in academia, is: it depends. A unit of study in an overall curriculum design, or perhaps an element of that unit of study, might aim to teach a necessary performance competency, e.g., ability to identify the bones in the human body. If the expectation on students is that they must display a minimum level of competency and this can be measured, then students could be marked as having achieved a pass or fail. In one sense, this is a very simple form of SRA, with perhaps one or a few criteria and a single standard of performance for each that must be achieved. Little assessor judgement is required and the assessment problem could be recast as CRA.

My guess is that most teachers of units of study would wish to test the depth of a student’s learning of key skills, concept and acquisition of understanding, not by

![Figure 5 A social-constructivist assessment process model by Rust.](image)
reproduction of memorised rubrics, but by seeing how well students can apply their knowledge to real world problems, admittedly simulated and simplified to suit the assessment, but still couched in authenticity. I would imagine that, in these cases, learning objectives are such that divergent student responses to assessment are expected. My belief is that SRA provides the most appropriate assessment process in these situations. Its reliance on professional judgement aligns well with teachers’ expertise and its implementation is, to that extent, familiar and straightforward.

One of the important questions, to which I have already alluded, is where is summative assessment most efficiently placed in the design of the curriculum? I have opened the possibility that we might choose not to summatively assess all units of study. Let me suggest such an instance.

Sadler has defined learning as follows:

"For my money, learners can be said to have learned something when three conditions are satisfied. They must be able to do, on demand, something they could not do before. They have to be able to do it independently of particular others, those others being primarily the teacher and members of a learning group (if any). And they must be able to do it well. Assessment of learning should be directed towards gathering evidence for drawing inferences about capability under these conditions, not the scaffolded conditions." 78

Suppose one of the skills we want the students to master is writing a professional report. In a first semester, students are introduced to the practice of report writing, with workshops and a cumulative assignment, where students can submit first drafts of sections for critique and feedback. Students are allowed to resubmit their work as a single report at the end of the semester, after the feedback has been taken into account, for further assessment and grading. In terms of Sadler’s definition, that grading is not a true measure of the students’ learning, because it is scaffolded by the feedback. The grading is useful formative assessment, but a true summative assessment of student learning would be to require them, in the following semester, to produce a new professional report, without assistance on its production.

---

Implementing assessment

Setting standards
The crucial first step in standards-referenced assessment is the specification of standards. Sadler79 and Cooney80 have provided, independently, similar advice on how this can be practically achieved. There is also a strong alignment with the standards-setting procedure outlined in Hardison and Vilamovska81. The process is as follows. For a given assessment task already graded, separate out the student works into grade bands, that is, those that failed, passed, were of credit standard and so on. Starting, say, with those awarded a High Distinction, develop a verbal description that captures the essence of the reason for that grade. Pick out several exemplars that typify the grade. Sadler’s original discussion of exemplars82 showed that one is not enough, but, together with verbal descriptions, a limited number can be effective.

Repeat the task for works awarded a Distinction. Now test the standards (verbal description plus exemplars) to see that a differentiation of the grades is possible. The specification of the standard is refined until this test can be passed. The procedure is repeated for other grades and their boundaries.

The question of direction may be important. As described above, the process is from the top down. However, the most difficult standard of all may be the one that distinguishes a pass grade from a fail. What is a just adequate level of mastery? If this is the most difficult step, then it follows that one should start there and progressively move upwards, highlighting the improvements in mastery that justify progression to higher grades.

Sadler made an important observation about exemplars in his 1987 paper:

“Furthermore, because exemplars are concrete and chosen at a particular point in time, they may incorporate elements of fashion, cultural tradition, or current technology, virtually ensuring that they gradually go out of date. As a consequence, the practical utility of a set of exemplars deteriorates unless it is revised periodically. In addition, exemplars may encourage convergent responses and inhibit creative experimentation if they are interpreted literally as models to be copied. The permissible limits to variation then go unexplored.”

The process, as outlined, provides a step in the task of establishing standards and implementing standards-referenced assessment in existing, stable units of study.

79See footnote 67, pages 192-3.  
82See footnote 1, pages 201-207.
However, in those cases where a new unit of study is proposed, the standards will have to be established *de novo*. Any unit of study has a purpose, expressed in its learning outcomes. The process described above rests on the assumption that, in the context of work already graded, the assessment tasks were aligned with the learning outcomes defined for the unit of study. Therefore, the relationships between the assessment tasks, the standards of expected performance and the learning outcomes should be discernible and can act as a guide to the construction of standards for new units, based on the intended learning outcomes.

**Reviewing standards**

Comments in the previous chapter underpin the strategy of calibrating standards by conversation. This is not a trivial exercise, but one to be approached with the same rigour as selecting a new professor, say. The conversations should ask: in the context of the learning outcomes for the unit of study and the degree in which it is embedded,

- do the standards for a pass grade conform to the broad expectations expressed on page 5?
- do the standards for the highest grade stretch and challenge our brightest students?
- are the standards appropriate for the level of the unit of study? That is, compared to standards in the same area for years above and below, does a trajectory of increasing expectation of intellectual development emerge?
- are the standards commensurate with those for other units of study at the same level?

Conversations may seem a ‘soft’ process. However, Hardison and Vilamovska found that their step involving conversation to form a consensus within each panel generally resulted in a hardening of the standard. They also used review panels of four to five members, which suggests the impost of the review process on faculties should not be great, particularly if it is not done every year, but at some lower frequency. The inclusion of staff from other cognate units, above, below and alongside the one under review, would provide the comparisons outlined above. Students, too, can contribute to reward of the standards (see Figure 5).

Given that “standards” can apply both to the expectations as well as the actualities of student performance, another aspect of review of standards would be scrutiny of the standards achieved by students. Norm-referenced assessment, in which ranked students are graded according to an agreed distribution, typically a normal distribution, eliminates this necessity, but at the cost of weakening, if not eliminating any link between descriptions of quality and student grades. Its other major disadvantage is that any student’s grade is not determined solely on that student’s performance, but includes the performance of the other students in the class. See Biggs\(^83\) for a more complete

---

\(^{83}\)See footnote 76, pages 144-147 and page 194.
critique of the disadvantages of norm-referenced assessment. Like Biggs, I do not consider it appropriate for the task of assessment at this university.

That being said, some distributions of student grades against the standards will justify scrutiny. Consistent poor performance by all students may point to standards that are pitched too high or failure to effectively communicate the expected standards to students and allow them to adequately test their ability against the standards prior to summative assessment. Consistently good performance may point to standards that are pitched too low, so that the best students are not adequately challenged. There may be a sound reason for this. The learning required in some competency-oriented units of study may simply not challenge most of our students. In this case, it may be more correct to have a single standard of satisfactory performance and offer only pass/ fail grades.

The important issue here is that the processes of setting and reviewing standards should allow all stakeholders the opportunity to comment and agree on the standards a priori. If our teaching and the engagement of the students is such that many can demonstrate a high level of achievement against challenging standards, then this is surely a great success.

**Aggregating marks.**

This is surely a matter of great contention for any system of assessment, not just standards referenced assessment. The almost universal practice of simply adding marks both within and across units of study has no real basis except tradition. The result is something approaching a single mark (think for example of Year 2/3 WAM as eligibility for Honours) that obscures all detail of what the student can actually do. Some courses which include mandatory thresholds with pass/ fail grades do not include those achievements in the calculation.

The problem is non-trivial but must be tackled on academic grounds rather than convenience. Within a unit of study, if there are multiple assessment tasks, consideration must be given to how they align with learning outcomes, whether they are formative or summative in intent, or whether they all must be passed are questions that must be answered before a solution to the aggregation of marks can be decided.

If our intent is to warrant student competence and offer a grade that recognises a high order of student academic achievement for the unit of study, then we should be able to arrive at a judgement of the student’s performance in the unit, taken as a whole. Biggs offers a useful guide to good practice here. He proposes three models for aggregating performance across multiple assessment tasks, to arrive at a holistic grade for the unit. This appears to me to align well with the tenets of standards-referenced assessment.

---

84See footnote 76, pages 190-194.
Biggs also offers a protocol for converting an overall grade (e.g., HD) into a mark. This is perhaps not so important for the lower grades, but may provide an essential identifier for, say, a prize recipient.

If there are multiple assessments within a unit, it may be sufficient to return grades (e.g., credit) to students, rather than marks. This allows the assessor to later make a sound overall judgment of the level of achievement for the unit as a whole, according to the agreed protocol. This judgement will, almost inevitably, have some holistic aspect to it, so that the final judgement need not be precisely the same as an arithmetic sum or average of component scores. Because the learning process has informed students about the exercise of professional judgement, they should be able to both understand the assessor’s decisions and make accurate inferences about their own performance.

A major problem remains and is not susceptible to easy resolution. What is the best process for making an accurate declaration about the student’s overall performance in the degree? Is it simply the process for units of study, writ large, or some other deliberative process? The simple addition of numeric scores does not seem a good fit alongside the accumulation of diverse skills, knowledge and understanding implied by the learning outcomes for the degree.

Scaling

The holistic judgement of grades for a unit of study does not demand that the final allocation of grades is determined in a single pass. Prudent assessors will check their judgement by asking, at any grade boundary, two questions with reference to the standards:

- For the better of the two grades, does the worst student performance receiving that grade truly deserve this grade rather than the poorer grade?
- For the lesser of the two grades, does the best student performance in this grade truly not deserve the better grade?

When these two questions can be answered affirmatively with reference to the standards, the judgement may be considered sound. Clearly, involvement of more than one assessor in this evaluation should improve its reliability. The above process is an important check on the application of the standards and is not considered scaling in this paper.

Scaling may be necessary in a norm-referenced assessment environment where the numeric grades must be adjusted arithmetically to ensure the correct distribution of grades is achieved. In a criterion-based or standards-referenced assessment environment, such scaling has no place.

However, things can go wrong! An assessment may fail in practice because it fails to adequately test student performance against expected standards. Typically, student
results will be skewed to one end of the spectrum or the other. Scaling a failed assessment to “correct” the problem seems illogical. A better outcome would be to accept that this test is invalid as presented and check to see if either:

- The form of that test item allows student performance against it to be scored pass-fail, or
- The overall judgement of student performance for the unit of study can be performed adequately by ignoring that test, possibly with some adjustment to the standards used. Such an adjustment would have to be recorded and students notified.

Non-academic issues
The application of a late penalty with some deduction of marks is a common practice. However, it and similar practices are really attempts to modify student behaviour and place student behaviour on the same scale as academic achievement. There is a logical paradox here, which has been elaborated by Sadler\(^{85}\) along with other considerations such as the appropriate response to plagiarism in work submitted for assessment. The suggestions offered in the paper may have useful implications for curriculum design.

Forms of assessment
Standards-referenced assessment can be applied, as an over-arching philosophy, to many forms of assessment, but not all. Biggs\(^{86}\) offers a useful catalogue of assessment tasks, as do Dunn et al\(^{87}\). Williams\(^{88}\) provides a useful approach to use of examinations that may reduce the stresses normally associated with exams, without reducing their effectiveness at probing deep understanding.

Moving from norm-referenced assessment
Norm-referenced assessment, used at a university, should be attempting to provide the same opportunities for students to demonstrate a profound mastery of their subjects as does standards-referenced assessment. Divergent responses should be the norm and their quality should be judged against criteria and standards appropriate to the assessment task. In that sense, there is relatively little additional effort required to move to a standards-referenced environment.

At the end of the process that judges student work, the norm-referenced approach should have the student works rank ordered in terms of quality. Now, instead of putting grade boundaries at positions in the rank dictated by proportions, the grade boundaries are located where reference to the standards indicates. The process used is the same as

\(^{85}\) Sadler, D.R. (2009). Fidelity as a precondition for integrity in grading academic achievement, Assessment & Evaluation in Higher Education, xx ,

\(^{86}\) See footnote 76, pages 165-175.


that described earlier for checking grade allocation under “Scaling”. That is, in terms of allocating grades, only a different final step is required.