Abstract
We know much about higher education, but very little about assessing student knowledge and skill in ways which are generalisable beyond specific subjects or courses. This paper argues for greater progress in this rapidly developing area of higher education. It suggests approaches for enhancing the assessment of student knowledge and skill, and considers how these might underpin a model of assessment standards. Preliminary observations are made in conclusion about what might be required to bring such changes into practice. The analysis is by no means intended to be exhaustive, but rather to provide a discursive snapshot of a few main trends and possible developments.

Keywords: student outcomes, assessment methods, academic standards

1. Measuring Outcomes in Higher Education
Demonstrating what individuals know and have gained through years of higher learning has long been of interest to them and their communities. Yet uncertainty still surrounds efforts to produce generalisable measures of the outcomes of higher education. This paper argues that there is an increasing need to move beyond the obstacles which have hindered progress in this area and produce measures of knowledge and skill which are generalisable beyond localised individual or institutional contexts.

The paper considers possible improvements which might be made to assessment practices and outlines a model of assessment standards which such improvements would help to support. Many approaches in this area have been proposed and tested over the decades, although none of the approaches have fully addressed the demands placed upon them. There are grounds for optimism, as noted in a series of recent reviews (Miller & Ewell, 200; Dwyer, Millett & Payne, 2006; NCPPHE, 2006), because current insights into higher education learning along with contemporary assessment methods say much about how we might produce generalisable measures of higher education outcomes.

Based on the conclusions advanced by Spellings (2006) and Ischinger (2006), it is assumed in this paper that interest in ‘outcomes’ will grow as higher education plays an ever-more important role in individual and community life. Governments no longer supply institutions with single-line budgets, focusing on the input side and leaving the processes and outcomes up to institutions. Increased regulation and prescription of institutional practice has grown with increases in the costs and value of higher education to advanced service economies. A further shift in individual and community focus towards outcomes seems very likely as quality processes can be largely assumed, and it becomes more strategically important to assure the competence and capability of highly trained individuals (James, 2003; Economist, 2006).

It is acknowledged at the outset that the terms ‘outcomes’ or ‘outcomes-based’ can be considered ‘dirty words’ in higher education. These terms are often seen to represent a position...
which, at its best, embraces futile attempts to measure intangible aspects of academic learning by reducing them to discrete competencies or, at its worst, reflects a pernicious agenda to prise the assessment of academic standards away from academics or institutions. Universities, academic knowledge and higher education pedagogy definitely do incorporate unique complexities to which any measurement of outcomes must be sensitive. It is unlikely that measuring outcomes in higher education is impossible, however, or that it cannot or should not be done in ways which are valid and value-adding to learners, teachers and institutions.

2. A Focus on Knowledge and Skill Outcomes

The outcomes of higher education are many and varied, but this paper focuses on student-level knowledge and skill. Individuals attend university to learn, and the development of knowledge is arguably the most basic outcome of university study. The focus and scope of this knowledge will of course vary across individuals and contexts. It will range from awareness of broad ideas, knowledge of basic facts, and understanding of sophisticated processes and systems.

While much knowledge will be specific to various courses and disciplines, individuals are also expected to acquire general knowledge through study. This may be knowledge about ethical standards and processes, of social contexts and systems, or of how best to manage personal learning. It is these more general kinds of knowledge that are often used to characterise the ‘higher-order’ or ‘value-added’ outcomes of university study.

In addition to knowledge, higher education should help people develop skills so that they can use their knowledge in productive ways. Given the nature of academic learning, many of the skills learned in higher education will likely be psychological rather than behavioural. As with knowledge, many are likely to be specific to particular contexts and fields of study. Others, however, will be more general and potentially form part of most, if not all, courses.

Many taxonomies of such generic graduate skills exist, often produced to link with various qualifications frameworks or definitions. A shortlist of the most common of these generic skills may include: critical and evaluative thinking, analytical problem solving, innovative and creative thinking, written and verbal communication, collaborative work, numeracy, technology literacy, independence and initiative, and planning and management. Citizenship or social and political awareness might be added to this list. Interestingly, many if not most of such lists omit the basic skill of knowledge development.

A skill for learning may be one of the most important outcomes of higher education. Most graduates will take up forms of knowledge work that require re-learning, updating and expanding their knowledge and skills. Graduates need to engage in formal learning throughout their lives, and it is important that they have the necessary skills and interests to do so. Oddly, while a capacity for learning and an appetite for inquiry are clearly basic and important outcomes, these are rarely taught explicitly at university.

3. Enhancing the Assessment of Knowledge and Skill

While there are pockets of excellence, developing generalisable measures of individual knowledge and skill outcomes remains a major challenge for higher education. There have been enormous advances in assessment over the last hundred years, much of which is summarised or exemplified by Linn (1989), Keeves (1998), AERA (1999), NRC (2001) and OECD (2003). Important aspects of this assessment methodology, however, have yet to be applied to higher education. Universities and academics are responsible for monitoring and assuring academic standards, and it is critical that continuous efforts are made to enhance the standard of assessment itself.
Currently, many tasks are developed by individual teaching staff for use in specific subjects whose content may change in various ways from year to year. Staff often develop such resources in relatively short periods of time, for localised purposes and with limited resources or knowledge of assessment methodology. As a result, student knowledge and skill is often measured using uncalibrated tasks with unknown reliabilities and validities, scored normatively by different raters using unstandardised rubrics and then, often with little moderation, adjusted to fit percentile distributions which are often specified *a priori* by departments, faculties or institutions.

It is feasible to develop validated assessment tasks for a large number of higher education subjects. In general, large subjects based on a single textbook, which take a ‘shrink-wrapped’ approach, could well be accompanied by assessment materials. These materials could incorporate formative assignments for continuous assessment as well as validated examinations or items. The tasks themselves could be supported by notes for managing the assessment, analysing data, interpreting results and reporting achievement. A degree of flexibility would presumably need to be designed into the tasks to both encourage and support local adaptations. These assessments could be designed to fit different levels and fields of study, and may include performance tasks, portfolios, open ended questions, constructed response items, and multiple choice questions. The validated tasks for these mass subjects could take many different forms, their defining characteristic being that they are designed to optimise the measurement, diagnosis and enhancement of learning.

Many higher education subjects are specialised in nature or small in scale, and it may not be feasible to develop fully validated assessments. It is important, nonetheless, that the resource-consuming nature of assessment task design does not inhibit the use of high quality assessment activities. In such instances, the most appropriate approach may be to train academic staff. An awareness of basic principles of assessment design and advanced practice would develop the capacity of teaching staff to enhance their own assessment tasks and activities. It would also have more general pedagogical benefits, of course, by requiring academics to think not just about what and how they will teach, but about what students are expected to learn and how they should be assessed.

The training of teaching staff in assessment could be coupled with a process of assessment task review, in which technical experts or academic colleagues offer feedback on assessment tasks and approaches, and ensure that tasks are of appropriate quality. Of course, this currently happens for many courses and assessments, but the process is by no means universal. The largely individualised development of assessment tasks makes it difficult to develop informed and generalisable criteria which map out thresholds of increasing performance. It is difficult, as a result, for institutions to assure that quality of the tasks which are themselves used to set academic standards.

Moderation processes might be used to ensure the generalisability of assessment standards and outcomes. In general, moderation involves teaching staff reviewing samples of student work to assure the comparability of standards across contexts. Such moderation may be conducted on an *ad hoc* basis, as often already occurs. It is preferable to design robust and scalable management systems, however, to ensure that outcomes can be quality assured. Moderation could be managed by a cross-institutional agency, as in senior secondary education, or perhaps by a cluster of discipline-specific agencies. It might involve statistical calibration processes to help equate standards, highlight unusual scores and manage moderation processes.

The assessment of subject-specific skills might largely mirror the assessment of knowledge, with appropriate differences such as a greater emphasis on performance-based forms of assessment. Perhaps the greatest point of departure is in terms of the assessment of generic skills. While widely marketed in course descriptions and objectives, generic skills are rarely taught
explicitly in university courses. They are likely characterised and nuanced by the fields in which people study, encounter and practice their skills. These skills are dynamic and tacit in nature, therefore, and this has implications for how they are assessed.

Objective tests can be used to measure critical thinking, problem solving and numeracy skills. Such tests have become increasingly popular over the last decade. Specific tests might be conducted to measure these skills, or validated items could be embedded within more routine forms of assessment. While resource-demanding to produce and administer, these tests have the advantages of providing objective estimates of each participant’s performance and of moving beyond subjective self assessment. Several tests exist to measure adult generic skills, including numeracy and literacy, and a few have even been applied in international studies. Tests might include items to assess aspects of, but not all, other interpersonal, communication and technology skills.

Performance assessment would seem to be the best means of assessing interpersonal, communication and technology skills, which are more social and behavioural in nature. An individual’s capacity to work and communicate with others, for instance, could be measured by observing their participation in group tasks and class presentations. Technology literacy could be assessed by observing how people interact with computers and other technologies.

Individual self reports have been shown to play a particularly useful role in seeking feedback from students on the extent to which they participate in productive learning practices, and hence demonstrate a skill for learning. But self reports can also be used to capture learners’ perceptions of the extent to which their university study and experience has developed their generic skills. Individual self reports might be obtained via a survey of later year, graduating or exiting students or, possibly at the end of each subject.

4. A Model of Assessment Standards

It has been argued that only limited progress has been made in assessing university students’ achievement and capability in ways which are generalisable beyond individual subjects or courses. This paper has asserted the need for progress in this area, and outlined approaches for enhancing the assessment of student knowledge and skill which could advance current practice. Much would likely flow from developing and implementing such improvements, which would play an important role in monitoring and enhancing academic standards. Much work, however, would be required to bring many of the ideas sketched out above into practice.

A good way to begin would be to map out a valid and feasible model of assessment standards. Such a model could consist of an indicator framework, a series of performance standards, a suite of measures to support these indicators, and an approach consisting of methods, such as those outlined above, for gathering data on each measure. Such models have been developed by several institutions around the world, and the essential building blocks already exist in many institutions. Harnessing and developing this activity would be a productive way of making early progress in this area.

Generalisability should be built into such models by ensuring that the indicators reflect high-level educational outcomes such as those advanced in this paper. While the precise selection of such outcomes and the emphasis placed on each is likely to vary across contexts, it is important that they have the sort of systemic and student relevance as those mapped out above. The selection of indicators is important, as it contains a values statement, determines the ultimate validity and feasibility of the model, and will likely influence future policy and practice. Agreeing on a set of outcomes indicators, both within institutions and systems, is an important activity in itself.

Once selected, the indicators should be positioned within a standards framework which captures varying levels of performance. These levels would be linked to the underpinning
measures and desired benchmarks, and described in ways which are useful to students, teachers, institutions and industry. Some indicators are likely to be binary in nature, while others are likely to reflect several levels of increasing performance. Certain aspects of such a framework are likely to be locally defined, but others might well be underpinned by more generalisable learning or employment taxonomies. Ideally, the framework would be designed at a level of generality which enables it to be relevant to all year and qualification levels, and to all fields of education.

Each indicator framework would be underpinned by one or more measures. These would ideally be common within but also across institutions, although this may not always be possible or may take some years to achieve. It is likely to be necessary in certain instances to develop processes for equating or standardising different measures, to ensure that a certain level of performance on one measure is aligned in appropriate ways with performance on another. Clearly, it would be productive to work with existing measures, although these, as outlined above, may not always be sufficient to meet the revised needs placed upon them.

As discussed, developing robust measures of student-level outcomes is likely to require various objective forms of assessment. For many courses, it is likely that assessment tasks would need to be developed which are either standardised in nature, or which have been passed through a robust production, scoring or moderation process. Any changes in how outcomes are reported will most likely involve the specification of marking rubrics and mark equation or conversion tables. Developing generic scoring rubrics, forms of task peer review, or simply training teaching staff in assessment methods are all means of making initial progress in this area. As canvassed, it may be beneficial to conduct stand-alone or embedded generic skills assessments which, in addition to providing measures of individual performance and potential in targeted areas, provide data which could be used for the purposes of moderation.

5. Bringing the Advances into Practice

A series of new policies and practices will likely be required to support any extensive movement towards the generalisable measurement of student-level outcomes. Many might be developed from existing pockets of activity, or transferred from school systems in which they have been developed and tested over many years. Certain forms of cultural change would likely be required, along with significant investments in training and systems. These changes and investments may be costly, but their value lies in the significant information that they would provide on individual learning and growth.

It is essential that academics and institutions themselves take the lead in developing this growing aspect of higher education. This is not just because institutions have the authority to accredit their own programs and ensure academic standards and underpinning quality assurance processes. It is vital that progress in the measurement of student outcomes builds rather than breaks the link held by teachers and institutions between the development, dissemination and assessment of knowledge. It is important that any new measurement of student learning and development is itself collaborative in nature, given the broader individual, social and economic roles such measures will play.

It has only been possible in this paper to offer a preliminary analysis of the specification and measurement of student-level outcomes. No attempt has been made, for instance, to consider how outcomes data might be best analysed, or how it might be reported to individuals, institutions or government. Among others, many important questions surround the value of absolute performance measures compared with measures of individual growth and added-value. While graduates and employers often review trends in performance over the duration of a degree, this perspective is rarely captured explicitly in academic transcripts or reports. This paper has also not attempted to situate the proposed conception of outcomes within a broader
framework of academic standards. This would require defining ‘academic standards’, and determining how student-level outcomes could function as indicators of these for management or quality assurance purposes. Further, the paper has not examined many of the important pedagogical and course management considerations implied by the propositions made about outcomes measurement.

There are many such possibilities for further development, but this paper has focused squarely on the developing generalisable measures of knowledge and skill outcomes in higher education. It is assumed that much would flow from significant progress in this area. Monitoring the quality of education is, or should be, linked in important ways to a capacity to demonstrate that individual learning and development has taken place.

References
Author/s: Coates, H

Title: Developing generalisable measures of knowledge and skill outcomes in higher education

Date: 2007

Citation: Coates, H. (2007). Developing generalisable measures of knowledge and skill outcomes in higher education. Australian Universities Quality Agency

Persistent Link: http://hdl.handle.net/11343/32195

File Description: Developing Generalisable Measures of Knowledge and Skill Outcomes in Higher Education