How well do Australian universities assess their students?

Stephen C. Russell
University of South Australia, Adelaide, Australia
Stephen.russell@unisa.edu.au

Abstract: Work carried out by Russell in 2008 collated the assessment guidelines of Australian universities into an ordered set of assessment requirements, and determined the overall support for each one. This work has been supplemented with new data, and refined to remove overlaps and ambiguities, and to improve the clarity of the assessment principles.This paper correlates the good assessment principles stated in the university assessment guidelines with the perceived outcomes reported in large scale student surveys to determine the relative successes of universities in achieving their assessment goals.

Introduction

It is well known that the foundations of good teaching, learning and assessment practices were developed in the 1970s and 1980s by people like Marton & Säljö (1976), Biggs (1979), and Entwistle & Ramsden (1983). Since that time experiments have gradually uncovered which assessment practices actually achieve the desired learning outcomes (see Suskie, 2006). There is, however, a suspicion that a gulf exists between theoretical knowledge of good assessment practice, and the actual fulfilment of the espoused principles.

According to Boud (1990) there are three potential gaps between assessment and theory. There are gaps between:
- what academics do in assessing students, and what they do in their own work
- assessment tasks and the real world
- assessment practices and academic ideals and goals of higher education.

The first gap occurs because teachers can be more concerned with presenting course content than actually assessing the learning objectives that they set. This is often because teachers fail in defining the learning outcomes that satisfy the objectives (Ramsden, 2003; Glaser, 1991).

The second gap highlights the tendency of teachers to use assessment tasks that are not authentic. So the students fail to make the connection between the tasks that they are asked to carry out, and the tasks that they perceive as being part of the real world, or of their future profession.

The third gap is addressed in this paper, but covers the other two gaps in effect. This gap is between how universities say they assess students, and how their lecturers actually assess students. This gap is seen as being a serious threat to education in the Western World. Race (2003) and Angelo (1996), for instance, have pointed to serious deficiencies in assessment practices in UK and US, respectively.

This state of affairs prompted Russell (2008) to examine the published assessment guidelines of Australian universities. Russell constructed an ordered set of good assessment principles, and measured the levels of support for these principles by Australian universities. This analysis was further refined to remove ambiguities and overlaps, and test the consistency of the findings (described in detail in a later paper).

The current paper addresses the potential gap between assessment theory and assessment practice through measuring the success of Australian universities in carrying out a selection of their published principles of good assessment practice.

Methodology

The first issue in addressing how well Australian universities assess their students is to determine clear statements of their principles of good assessment practice. This was carried out through accessing all the assessment guidelines available on the thirty-eight university websites listed by the Australian Education Network (2009). The principles were extracted from the published documents according to standard system engineering requirements analysis methods (as described in, for instance, *Systems Engineering Fundamentals*, 2001). The resulting statements were simple unitary requirements addressing just one issue, and were objective (as far as possible), non-ambiguous and non-overlapping. Since no guideline is ever intended to be mandatory, the 'should' imperative rather than the 'shall' imperative was used. Also, a concession was made to the flexible nature of common English usage by listing in brackets alternative words to the selected subject. These additional words are used in different guidelines, but appear to have a common intended meaning.

Votes were assigned to each of the requirements according to the number of universities that supported it, and the requirements were prioritised according to the number of votes received. The highest priority requirements, those with 20 votes or more, are listed in Table 1.

Table 1. The highest priority assessment requirements

No.	Derived Requirements	TOTAL
	Goals and Purpose	
1.1	Assessment should guide and encourage effective approaches to learning	29
1.2	Assessment should measure expected learning outcomes	38
1.3	Assessment should measure/encourage higher order learning (analyse, synthesise, evaluate)	21
1.4	Assessment should support students develop generic skills (graduate qualities, graduate capabilities)	22
1.5	Assessment should certify required performance or academic standards (competency, accreditation)	34
1.6	Assessment should be authentic (focus on application of knowledge and skills to workplace performance standards)	20
1.7	Assessment should be reliable (consistent, accurate)	35
	Assessment Criteria	
2.1	Assessment should be transparent (effectively communicated to students)	22
2.2	Assessment should be based on pre-determined and clearly articulated criteria that describe standards of knowledge, skills, competencies and/or capabilities (clearly stated tasks)	39
2.3	Assessment should be valid (aligned with desired teaching outcomes)	30
	How students learn	
3.1	Assessment should be flexible (use a range of assessment contexts and tasks, allows choice)	34
3.2	Assessment should be inclusive and equitable for all students (fair)	35
	Feedback	
4.1	Assessment should return clear and constructive feedback on the students' work	38
4.2	Assessment should return feedback in a timely manner	35
4.3	Assessment practices should be improved based on assessment outcomes, student feedback, review	23
	Workloads	
5.1	Assessment should involve reasonable workload for students (coordinated across programs)	27

The second issue was to determine how well each university achieves their professed assessment objectives. In order to determine this it was necessary to identify a rich and statistically significant (if possible) set of data that evaluates the success of a number of the assessment objectives. There are several possible alternatives available to public access, but it was decided that the most promising was the CEQ (Course Experience Questionnaire) – as reported in the University of South Australia CEQ Database (2009).

The tests that best fitted a selection of the highest priority requirements are shown in Table 2, where the number of votes for a particular derived requirement are shown in column 2.

Table 2 – Tests used in this research

No.	Votes	Derived Requirements	Test number/ Question texts[*]
1.1	28	Assessment should guide and encourage effective approaches to learning	CEQ/GSS 43 My program helped me to develop the ability to plan my own work
1.4	21	Assessment should support students develop generic skills (graduate qualities, graduate capabilities)	CEQ/GSS TOTAL + CEQ/GQS TOTAL
2.1	21	Assessment should be transparent (effectively communicated to students)	CEQ/CGS 28 I usually had a clear idea of where I was going and what was expected of me in this program
2.2	38	Assessment should be based on pre-determined and clearly articulated criteria that describe standards of knowledge, skills, competencies and/or capabilities (clearly stated tasks)	CEQ/CGS 08 It was always easy to know the standard of work expected.
4.1	37	Assessment should return clear and constructive feedback on the students' work	CEQ/GTS 01 The staff put a lot of time into commenting on my work CEQ/GTS 02 The teaching staff normally gave me helpful feedback on how I was going

* CGS (Clear Goals and Standards Scale), GQS (Graduate Qualities Scale), GSS (Generic Skills Scale), GTS (Good Teaching Scale)

According to the *Graduate Course Experience* (2007), there were 115,057 valid CEQ responses in 2007. Respondents were asked to express agreement or disagreement on a five-point response scale (strongly disagree; disagree; neutral; agree; strongly agree) to 13 core items, and from a subset of the remaining 36 optional items. This means that only a subset of universities answered some of the items used to test the success of assessment objectives in this research, irrespective of whether they agreed with general principle in question.

Graduate Course Experience (2007) noted that the fields of education explain most of the variation in the scores. It was, therefore, essential to remove the effects of the fields of education from the dataset in order to clarify the influence of the institutions on the results. This was done by normalising the datasets for each field of education selected for study (Creative Arts; Engineering and related technologies; Health; IT (Information Technology); Management and Commerce; Natural & Physical Sciences; Society and Culture) against their mean values across all the universities – giving percentage differences from the mean for each field of study, for each university. The results for each field of study were then averaged together to give an overall figure for the university, for the relevant test question.

Results

The results are shown in Table 3, where the tests are labelled according to their requirement numbering shown in Table 2. Results are presented in terms of the difference from the mean of the

'percentage agreement' metric, which reflects the percentage of responses in the 'agree' or 'strongly agree' rating scale categories on the survey form.

Here it can be seen that tests T2.2 and T4.1 were answered by all universities. These results indicate a natural dispersion of order 4-5% in achievement of objectives, and means are close to zero. This is as expected since every university is trying to achieve the same objectives, and some will do better than others.

For the remaining tests, only a proportion of universities deliberately voted for an assessment principle. From these results it is clear that there is no significant difference between those universities that voted for a principle, and those that did not.

Table 3 Average scores for all universities with and without a vote for an assessment requirement. Standard deviation is recorded under STDEV.

	No vote	Vote	STDEV*
T1.1	-0.49%	+0.21%	3.5%
T1.4	-0.21%	-0.25%	3.7%
T2.1	+1.67%	-1.20%	4.3%
T2.2		+0.07%	3.5%
T4.1		+0.09%	5.4%

    * Standard Deviation

The question remains as to whether universities are able to influence their staff in such a way as to cause a systemic effect in the quality of their assessment outcomes – presumably through some other means than publishing assessment guidelines? In an attempt to explore this question using the data acquired in this research, the outcomes of the CEQ test data for requirement 4.1 – *Assessment should return clear and constructive feedback on the students' work* – were plotted for each individual field of education. An excerpt from these data is shown in Figure 1. This bar graph shows clearly that some universities (anonymised in the graph) demonstrate a positive outcome for all fields of education (in this graph – A7, D2, and C5), while others show a negative correlation (in this graph – E1, C1, A1, and C6).

Conclusions

The first results of this research are the list of the declared good assessment principles by Australian universities, and the degree of support for these principles. While many of the principles can be recognised as being well supported in the literature, it would be interesting to determine just how strong the correlation is, and perhaps how up to date it is. In order to answer this question it would be necessary to perform a similar analysis of the literature as is carried out and reported here. The scope of this endeavour is rather broad, so it may require the imposition of strict constraints in order to address more limited but focussed objectives.

The main thrust of the present paper is to address the question of how well universities actually succeed in achieving their objectives in good assessment practice, as published in their assessment guidelines. It is clear from the data that there is no correlation between what a university states in their assessment guidelines, and the outcomes as determined from the CEQ survey.
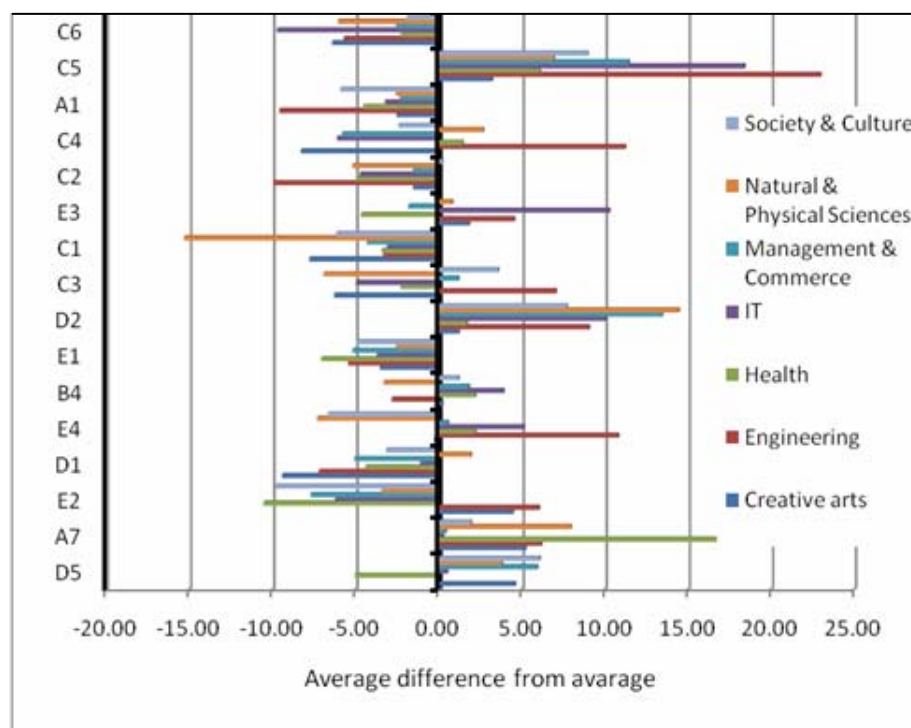
Figure 1. An excerpt from the test for assessment requirement 4.1 (*Assessment should return clear and constructive feedback on the students' work*) – which was T4.1 (*The staff put a lot of time into commenting on my work;* AND *The teaching staff normally gave me helpful feedback on how I was going*), plotted by subject area.

How does this relate to the potential gap between academic goals and actual practices? All we can say from this analysis is that no matter what a university says about good assessment practice, the simple statements made in the guidelines seem to have no effect whatsoever. However, there is a suggestion from Figure 1 that universities are able to cause systematically positive outcomes for all fields of education. The question is, then, what do these universities do to positively affect the assessment outcomes of their students? Conversely, what are some universities doing wrong to cause systematically negative outcomes?

References

Angelo, T. 1996, 'Transforming assessment: high standards for higher learning,' *AAHE Bulletin*, April, 3-4

Australian Education Network 2009, accessed at <http://www.australian-universities.com/list/> on 03 July 2009.

CEQ Database 2009, UniSA website, accessed at <https://my.unisa.edu.au/unisainfo/reporting/cgi-bin/ppdsisapi.dll?DC=Q&nia=Run&nid=48e3bd00222b11ddbcc0a2eb2c9c869a&nic=%2FUnisa%20Public%2FEvaluation%2FGCEQ%2FCEQ%20National%20Trend&nih=0&back=%2Funisainfo%2Freporting%2Fcgi-bin%2Fupfisapi.dll%3Fxmlcmd%3D%3CGetPage%3E%3CTemplate%3Emain.utml%3C%2FTemplate%3E%3C%2FGetPage%3E%26id%3D8b1e03c4bd8211d796e289b1e2dd2bca > on 3 July 2009.

Biggs, J 1979, "Individual differences in study processes and the quality of learning outcomes", *Higher Education*, 8, 381-394

Boud, D 1990, 'Assessment and the promotion of academic values,' *Studies in Higher Education*, vol. 15, no. 1, pp. 101-111.

Entwistle, N. & Ramsden, P. 1983, *Understanding Student Learning*. London: Croom Helm.

Glaser, R 1991, 'Expertise and assessment' in Wittrock, M C & Baker, E L (Eds) *Testing and Cognition*, Prentice Hall, New Jersey.

Graduate Course Experience 2007, *The report of the Course Experience Questionnaire*, Graduate Careers Australia Ltd, Victoria, Australia.

Marton, F. & Säljö, R. 1976, On qualitative differences in learning. :I Outcome and process", *British Journal of Educational Psychology*", 46, 115-127.

Race, P, Brown, S & Smith, B 2005, *500 tips on assessment*, 2ⁿᵈ edn, RoutledgeFalmer, London and New York.

Ramsden, P 2003, *Learning to teach in higher education*, 2ⁿᵈ edn. RoutlidgeFalmer, London and New York.

Russell, S. 2008, *Assessment in Australian universities: what they say they do to engage students*, Australian Technical Network (ATN) Assessment Conference, Adelaide

Suskie, L 2006, *Five dimensions of good assessment*, Middle States Commission on Higher Education, US, accessed at <http://planning.iupui.edu/page/download/?key=129728430&path=/consult/conferences/national/06/handouts/monday/**suskie**.pdf > on 28 Oct 2008.

Systems Engineering Fundamentals 2001, Defense Acquisition University Press, Virginia, US