AAGLO Summary 4: Standardised testing of Graduate Learning Outcomes in higher education

December 2011

AAGLO Project interest in standardised testing

The possibility of a role for standardised testing of both generic and discipline specific graduate learning outcomes in the Australian context has been raised in a TEQSA (2011) discussion paper and through Australia's participation in the OECD Assessment of Higher Education Learning Outcomes (AHELO) project. Standardised testing as an approach to the assurance of student achievement of graduate learning outcomes (GLOs)\(^1\) have therefore been considered by the ALTC AAGLO - Assessing and Assuring Graduate Learning Outcomes project team as one aspect of their investigation of two key questions:

- What types of assessment tasks are most likely to provide convincing evidence of student achievement of or progress towards graduate learning outcomes? and,
- What processes best assure the quality of assessment of graduate learning outcomes?

Our investigation involved a literature review, discussion with representatives of the OECD AHELO and Collegiate Learning Assessment (CLA) project teams and consultation with members of the AAGLO reference group whose responses are embedded in this summary.

A review of standardised testing

The practice of standardised testing on a national or international scale is more common to the school sector than to higher education. Nevertheless the experience of the school sector is useful in describing practice and identifying issues that have emerged.

Morris (2011, 5) in an extensive review of standardised testing in OECD countries defines them as “tests that are designed externally and aim to create conditions, questions, scoring procedures and interpretations that are consistent across schools”. She distinguishes between standardised tests with high stakes for students – where results determine access to or graduation from an institution or program of study - and those with no stakes for students – but with possible stakes for teachers, programs or institutions if linked to funding arrangements. Decisions to implement standardised testing are attributed to five primary drivers: 1) New public management; 2) Standards based assessment; 3) International competition; 4) Increasing demand for 21st Century Skills; 5) Test industry pressure (ibid, 7).

Morris (ibid, 10) argues that “the purpose behind a standardised test should guide the rationale for the assessment and feed into the design and implementation of the test as well as steer the use of the test results”; and that inferences drawn from specific test results will not be accurate or valid for purposes other than that for which the test was designed. In addition to tests with high stakes for students’ academic or professional careers, there are four distinct purposes identified for administering standardised tests, each associated with implications for use of test results.

<table>
<thead>
<tr>
<th>Standardised tests purpose</th>
<th>Use of test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>To monitor and evaluate the education system</td>
<td>To inform policy</td>
</tr>
<tr>
<td>To hold the education system (and/or its components) accountable</td>
<td>To reward or sanction institutions or teachers</td>
</tr>
<tr>
<td>For public information (i.e. by those outside the education system)</td>
<td>To compare and rank schools</td>
</tr>
<tr>
<td>For formative purposes</td>
<td>To monitor, support or advocate</td>
</tr>
</tbody>
</table>

The practice of using single, large-scale, standardised tests to serve multiple purposes – e.g. the determination of national standards and the determination of rewards or sanctions for teaching quality – is considered problematic in that any increase in the ambiguity of result validity has the potential to undermine the assessment system (ibid).

Test design and development involves numerous decisions too detailed for a elaboration in this summary. They relate to the scope of the test, development of test items, test frequency and timing, sample or census-based participation, the method of analysis of test results - norm referenced, criterion referenced or growth measures - the use of ICT, implementation and scoring. There are also decisions related to if and how results are reported -are they to be publicly available? aggregated or disaggregated? adjusted to account for factors outside institutional control?

Advantages of standardised testing

Many standard tests such as the US Collegiate Learning Assessment and the Australian Graduate Skills Assessment focus on generic skills such as critical thinking, written communication and problem-solving, though the AHELO project has discipline specific tests for engineering and economics in development in addition to its adaptation of the CLA (Tremblay 2011).

\(^1\) Graduate learning outcomes: the intended learning outcomes of a particular university program as specified by an institution or discipline.
Agencies engaged in the development and implementation of standardised tests claim a range of significant educational benefits. These include the provision of comparisons with similarly situated institutions to use as the basis for a formative approach to the enhancement of students’ educational experiences and learning outcomes. AHELO also aims to inform students, governments and employers about the effectiveness of resource deployment and graduate employability (AHELO 2011, 2) while it is suggested that information reported from the Graduate Skills Assessment is used for fields of study comparisons, measures of growth, early identification of students in need of support, entry into postgraduate courses or for seeking employment after graduation.

AAGLO reference group (ARG) members expressed a number of reservations as well as commenting that:

Standardised testing measuring differential between skills upon commencement in schools and graduation could form part of an overall analysis of the University's skills improvement mechanisms, both as embedded in the curriculum and in the form of other language and academic skills programs (ARG1).

Introducing a standard test across the whole of Australia will allow for ease of benchmarking (ARG2).

I can see a possible use of common tests for assessing small aspects of knowledge-based discipline specific learning outcomes in early foundational aspects of a higher education degree (ARG3, emphasis in original).

To date the use of major standardised tests is voluntary, a factor that providers consider crucial to effective use.

Issues around the use of standardised testing

Standardised testing is a complex undertaking that has limitations and possibly undesirable consequences (Banta 2006, 2007a, 2007b; TEQSA 2011; Sadler 2011) as well as potential benefit.

Limitations of standardised testing

It is questionable whether generic skills can validly be assessed “in isolation from the discipline specific contexts in which students have gained their education and socialisation” (ARG4), as “very little is being written about what tests of generic skills are actually measuring and with what accuracy (ARG5)”.

Doubts have been raised regarding the risk that “if absolute student scores are not appropriately controlled, they will fail to distinguish the contribution of the actual teaching from the entry standards of the institution” (TEQSA 2011, 12). In the United States it has been reported that “institution level correlation between student scores on the tests of generic skills and entering SAT/ ACT scores is so high that prior learning accounts for at least 2/3 of the variance in institutional scores” (ARG5), and that the effects of age, gender, socioeconomic status, race/ethnicity, college major, sampling error, measurement error, test anxiety and student motivation to perform account for a proportion of the remainder. “We must argue for multiple measures of institutional effectiveness ... and ... specific guidance for improving curriculum and instruction” (ARG5, emphasis in original).

Undesirable consequences

There is evidence that standardised tests have led to unintended and undesirable consequences in the school sector and there is potential for their replication in higher education, particularly when student test results are used in accountability systems or published in ways that encourage unjustifiable or demoralising comparisons.

Morris (2011) reports an increase in strategic behaviours such as ‘teaching to the test’. This can involve both item teaching - teaching test taking skills and using test or similar items as instructional material - or organising the curriculum around test items rather than a body of content. In either case, this can inflate scores without reflecting an actual increase in student understanding and therefore provide a misleading measure of student achievement. “If this is the case then it is not the instrument itself that needs examining but the way in which is used in the higher education sector” (ARG2).

A related consequence is the narrowing of the curriculum where efforts to improve test scores will narrow instruction to those aspects of the curriculum most likely to be tested. “A perverse outcome that could result is the incorporation into curricula of generic communication or problem-solving courses for example, at the expense of skill development in areas such as a clinical communication skills or skills in specific approaches to solving engineering problems” (ARG4).

When test results have implications for institutional or teacher sanctions and rewards there is evidence of manipulation of the student population through the exclusion of low performing students from taking the tests. International research and media coverage of the Australian school NAPLAN testing has reported instances of cheating in which teachers have changed student responses, filled in blank answers, allowed additional time, or provided students with answers.

In addition “Universities draw on people from vastly different demographics students, and having standardised testing that is only done upon exit or graduation can disincenivtise universities from enrolling students from disadvantaged backgrounds or students who have attained a lower ATAR score and are, bluntly assumed to have a lower grasp of generic skills” (ARG1).

Contribute to the AAGLO Project

You are invited to contact the project manager if you would like to contribute information or opinion on the issue of standardised tests or to register your interest in being included in planned project consultation activities.

AAGLO Project Team

Simon Barrie (The University of Sydney):
Clair Hughes (The University of Queensland):
Geoffrey Crisp (RMIT University):
Anne Bennison (The University of Queensland):
a.bennison@uq.edu.au
References

AHELO Project website www.oecd.org/edu/ahelo


Collegiate Learning Assessment website http://www.collegiatelearningassessment.org/


