

AAGLO Summary 8: e-Assessment issues in the effective assessment and assurance of GLOs

The AAGLO project and e-assessment

The focus of the *AAGLO - Assessing and Assuring Graduate Learning Outcomes* project is the 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 (GLOs)? and,
- What processes best assure the quality of assessment of graduate learning outcomes?

Though the pedagogical principles for e-assessment are the same as for any other form of assessment, the growing availability of accessible technology has expanded the options for the assessment of graduate learning outcomes.

e-Assessment and GLOs

The Report on Summative E-Assessment Quality (REAQ) was published in the UK by JISC and examined what academics thought would constitute quality in e-assessment. Academics indicated that high quality was associated with:

- Psychometrics (reliability, validity);
- Pedagogy (mapping to intended learning outcomes);
- Practical issues (security, accessibility).

The REAQ Report emphasised that e-assessment was not envisioned solely as a method of convenience for testing large student cohorts, but rather that the medium provided an additional method for designing tasks that were authentic to the learning outcomes being tested. Academics recognised that more sophisticated task types needed to be developed to evidence achievement of or progress towards many GLOs. Quantitative methods of determining validity and reliability were seen as crucial to quality assurance, but the available methods for determining validity and reliability tend to rely on task types that privilege selected response questions (for example MCQs). Academics expressed concern that additional qualitative methods for measuring quality needed to be available for general use in the online environment.

There is an extensive literature in the discipline of measuring validity and reliability, including Classical Test Theory and Item Response Theory (particularly the Rasch Model) (Baker, 2001; McAlpine, 2002; Wang,

2004). These techniques are applicable to e-assessments using selected response formats, but are less well developed for assuring quality of GLOs that require constructed response formats (Crisp, 2012). Academics tend to rely on accumulated discipline-based history about what constitutes an acceptable standard, rather than use quantitative statistical principles. The key validation tool for the majority of current assessments tends to be academic acumen rather than quantitative evidence (Knight, 2007).

There are recent trends towards using a variety of e-assessment tasks to gather more diverse evidence to inform assessment decision making. For example, the use of simulations, role-plays, scenarios, e-portfolios, blogs and serious games can provide assessors with evidence on aspects of performance that are not easy to capture using conventional assessment approaches and may be used to provide evidence of the demonstration of more complex GLOs (Crisp, 2012).

The assessment of role-plays, simulations and scenarios would normally involve divergent, constructed responses where students document their reflections on their actions within a virtual environment. These virtual activities tend to assess a student's affective capabilities rather than their acquisition of discipline content; they also tend to involve complex real world problems that do not have a prescribed solution. Role-plays, simulations and scenarios allow students to explore the complexities of an issue and the need to consider multiple stakeholder perspectives when proposing a solution to a problem.

Allowing students to manipulate data, to examine the consequences of their responses and to make informed decisions about potential solutions are all consistent with assessing advanced skill development in students, as described by the higher levels of the SOLO taxonomy (Biggs & Tang, 2007). Simulations and sophisticated digital tools allow students to construct multistructural and relational responses to questions.

E-assessment can offer new opportunities to assess 21st century skills through the design of tasks that require Web 2.0 creative activities; interactive tasks that include branching and decision points such as role plays and scenario based activities; and through the use of global communication tools, including blogs, wikis and discussions boards. Academics can use experiential and task-based assessments that include the appropriate use of virtual worlds to capture evidence of performance

rather than the recall of information; authentic assessment activities in virtual should involve criteria relevant to student performance, rather than the recall or manipulation of content knowledge in isolation from context (Richardson & Molka-Danielsen, 2009; de Freitas & Neumann, 2009).

e-Assessment standards

Many issues specific to e-assessment are technical in terms of task compatibility across different computer devices or operating systems and staff development for academics to design tasks appropriate to the online medium. e-Assessment standards, as they exist today, tend to be designed around the construction and delivery of summative high stakes exams and the sharing of content across different operating systems, rather than about pedagogical quality. The British Standards Institute (BSI) has published a series of standards for e-learning and e-assessment:

- BS ISO/IEC 23988:2007: a code of practice for the use of information technology (IT) in the delivery of assessments;
- BS 8426: a code of practice for e-support in e-learning systems;
- BS 8419-1 and -2: Interoperability between metadata systems used for learning, education and training.

BS ISO/IEC 23988:2007 makes recommendations for the use of ICT to deliver valid, fair and secure assessments and to collect and score participants' responses for high stakes examinations.

A global e-learning standard using XML (Extensible Markup Language, defines a set of rules for human-readable and machine-readable documents) has been published by IMS (IMS QTI). The IMS QTI (Question and Test Interoperability) specifications use standardised XML code to define how to represent assessment content and results, how these can be stored and exchanged between systems, including common learning management systems. The specifications facilitate assessments being authored and delivered on multiple platforms without needing to be rewritten. These specifications and standards tend to be more relevant to vendors of learning management systems, publishers or organisations producing e-assessment content, rather than academics.

e-Assessment guidelines

The International Test Commission published its Computer-based and Internet delivered testing Guidelines; they are designed predominantly for commercial or professional associations designing, delivering and validating e-assessments (ITC, 2005).

The Scottish Qualifications Authority has produced general guidelines for e-assessment that are designed for managers of assessment centres as well as practitioners who are looking to create and deliver e-assessments (SQA, 2007). These guidelines tend to be tips and

suggestions related to good assessment practices and link to much of the existing literature.

FREMA is the Framework Reference Model for Assessment project and defines how the components of assessment, for example the services associated with the construction, delivery and recording of e-assessments, interact in an online environment (FREMA). What is interesting about the FREMA model is it defines an ontology for the e-assessment domain as well as a set of e-assessment concept maps describing the entities and processes involved in e-assessment.

The University of Dundee Policy and Procedures for Computer-Aided Assessment have been in use since 2002 and regularly revised. Although these procedures are written for a specific institution, they are generally useful because they highlight the importance of planning the pre and post components of assessment sessions, including an analysis of the quality of the questions and how well the questions related to the learning outcomes.

In Australia, the Australian Flexible Learning Framework and the National Quality Council have published a detailed guide on e-assessment for the VET sector (AFLF, 2011). The guidelines are intended to be used by AQTF auditors as a reference when evaluating the e-assessment offerings of institutions.

Conclusion

As academics become more familiar with the use of evidence-centred assessment design principles we will witness a more coherent alignment between the learning outcomes articulated for a course, the learning activities set by the academic and the assessment tasks completed by the student (Shaffer *et al.*, 2009). This more coherent alignment will allow GLOs to be assessed by the most task type, whether that task is completed online or offline.

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)

References

- AFLF (2011). E-assessment guidelines for the VET sector.
http://www.flexiblelearning.net.au/shared/docs/e_assessment_guidelines_for_the_vet_sector.pdf
- Baker, F. B. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. <http://edres.org/irt/baker>
- Biggs, J., & Tang, C. (2007). Teaching for quality learning at university: What the student does (3rd ed.). Maidenhead, Berkshire: Open University Press
- BSI. <http://www.bsigroup.com>
- Crisp, G. (2012). Learning, Teaching, and Assessment Using Technology. In Handbook of College and University Teaching. A Global Perspective. (Ed) James E. Groccia, Mohammed A. Al-Sudairy and William Buskist. Chapter 31.
- de Freitas, S., Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. Computers and Education. 52(2), 343-352.
- FREMA <http://www.frema.ecs.soton.ac.uk/overview>
- IMS QTI
http://www.imsglobal.org/question/qtiv1p2/imsqti_oviewv1p2.html
- ITC (2005). International Guidelines on Computer-Based and Internet Delivered Testing.
<http://www.intestcom.org/guidelines>
- Knight, P. (2007). Fostering and assessing 'wicked' competences. <http://www.open.ac.uk/cetl-workspace/cetlcontent/documents/460d1d1481d0f.pdf>
- McAlpine, M. (2002). A Summary of Methods of Item Analysis. CAA Centre, University of Luton.
<http://caacentre.lboro.ac.uk/dldocs/Bp2final.pdf>
- REAQ
http://www.jisc.ac.uk/media/documents/projects/reaq_finalreport.pdf
- Richardson, D. & Molka-Danielsen, J. (2009). Assessing student performance. In J. Molka-Danielsen and M. Deutschmann (eds), Learning and Teaching in the Virtual World of Second Life. Pp 45-60. Trondheim: Tapir Akademisk Forlag.
- Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Franke, K., Rupp, A. A., Mislevy, R. (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. The International Journal of Learning and Media. 1(2), 33-53.
- SQA (2007)
http://www.sqa.org.uk/files_ccc/guide_to_best_practice.pdf
- Wang, L. (2004). New methods for CBT item pool evaluation. Annual Meeting of the American Educational Research Association (AERA), April, 2004. San Diego, CA.
<http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/wa04-02.pdf>