

ASSESSING THE SCIENCE KNOWLEDGE OF UNIVERSITY STUDENTS: PERILS, PITFALLS AND POSSIBILITIES

Susan M. Jones

Tasmanian Institute of Learning and Teaching
University of Tasmania, Hobart, Australia
S.M.Jones@utas.edu.au

Abstract

Science content knowledge is internationally regarded as a fundamentally important learning outcome for graduates of bachelor level science degrees: the Science Threshold Learning Outcomes (TLOs) recently adopted in Australia as a nationally agreed framework include “Science Knowledge” as TLO 2. Science knowledge is commonly assessed using traditional examinations, tests and/or quizzes, but such forms of assessment can be problematic. A key contributing issue is the emphasis on “content” in many science curricula. Frequently, a focus on transmission of knowledge is mirrored in an over-reliance on traditional ways of teaching and an over-dependence upon summative assessment: students respond by relying on rote learning at the expense of developing a deep understanding of science concepts. The challenge is, therefore, to design teaching approaches that foster active learning, and, critically, to match these with rigorously designed and meaningful assessment tasks that support higher level learning of science knowledge.

Keywords

Assessment, content knowledge, examination, learning outcomes, multiple choice, rubrics

Introduction and Background

Science content knowledge is regarded as a core learning outcome for graduates of bachelor level science degrees, but meaningful assessment of science knowledge is still problematic, despite the wealth of available literature. There are two major areas of concern. The first is the degree to which “knowledge” is privileged as a major component of the typical undergraduate science curriculum and the second is the sometimes uncritical predominance of traditional forms of assessment as class sizes and, therefore, teaching academics’ workloads continue to increase. This paper aims to present an overview of why a more effective and rigorous approach to assessment of science knowledge is mandatory and discusses some of the challenges and possibilities inherent in the design of appropriate assessment tasks. As Hanauer and Bauerle (2012) pointed out, assessment reform is the key to facilitating innovation in science education.

Articulating graduate learning outcomes

In recent years, there has been an international thrust towards overt documentation of expected standards for graduates in specified fields of higher education, coupled with expectations that such learning can be evidenced, for external peer review and/or for formal quality assurance purposes. In the United Kingdom, for example, the Quality Assurance Agency for Higher Education (QAA) publishes Subject Benchmark Statements which describe the nature and characteristics of a degree in a given subject area and set out expectations of standards for the abilities and capabilities that graduates will have acquired. However, the Subject Benchmark Statements provide non-prescriptive guidance as to how programs should be delivered to ensure that graduates meet those expectations. Similarly, Tuning¹, a movement that began in Europe but has since been taken up in the United States of America, Latin America, Africa and Russia, is a process that aims to articulate what a student knows and is able to do in a given discipline at the point of graduation. Tuning documentation provides frameworks that establish clear learning expectations for students in given subject areas. Such documents can be used both as the basis for curriculum design and for quality assurance purposes.

In Australia, there has recently been considerable impetus towards defining and describing threshold (i.e. pass level) learning outcomes for graduates in particular discipline areas. The Australian Learning and Teaching Council (ALTC)-sponsored Learning and Teaching Academic Standards (LTAS) Project and later spin-off projects funded by the Office for Learning and Teaching (OLT) have provided a suite of nationally agreed threshold learning outcomes (TLOs) for graduates of a range of disciplines and at a range of educational levels. The TLOs for a particular discipline provide important insights into the qualities, attributes and skills that are considered germane to a tertiary level qualification in that discipline in the current environment.

The place of “knowledge” in an undergraduate science curriculum

In an era when it has become fashionable to dismiss the acquisition of knowledge as being of lesser importance than the acquisition of skills, the nationally endorsed Threshold Learning Outcomes (TLOs) for bachelor-level degrees in Science (Jones, Yates & Kelder, 2011) demonstrate that content knowledge still holds a central place in the core science curriculum. Science Threshold Learning Outcome 2 is *Scientific Knowledge* (Jones et al., 2011, p. 11), and is framed as follows:

Upon completion of a bachelor degree in science, graduates will:

2. *Exhibit depth and breadth of scientific knowledge by:*
 - 2.1 *demonstrating well-developed knowledge in at least one disciplinary area*
 - 2.2 *demonstrating knowledge in at least one other disciplinary area.*

¹ See <http://www.unideusto.org/tuningeu>

The *Good Practice Guide for Science Threshold Learning Outcome 2* (Jones, 2013) provided some discussion of how the phrase *demonstrating well-developed knowledge in at least one disciplinary area* should be interpreted. Jones (2013) suggested that students should acquire a body of knowledge that reflects “a coherent knowledge and understanding of the core principles and concepts of that disciplinary area” (p. 4) but noted that the flexibility of science bachelor degree programmes, both within and between institutions, means that it is not usually appropriate to mandate particular areas of knowledge at graduate level.

It is difficult to find a direct international comparison for the Australian Science TLOs, which were framed so as to apply across all bachelor degrees in science. Other national or internationally agreed frameworks such as the QAA Benchmark Statements² or the European Tuning Reference Points for the design and delivery of degree programmes³ do not provide over-arching documents for science but, instead, focus at the level of specific disciplines or subject areas. However, some useful comparisons can still be made. For example, the Subject Benchmark Statement for Biosciences (2007), while acknowledging the vast diversity of degrees under this disciplinary banner, stated that the subject knowledge common to all biosciences degree programs will include “engagement with the essential facts, major concepts, principles and theories associated with the chosen discipline” (p. 12), and that the “teaching and learning strategy should be designed to encourage a progressive acquisition of subject knowledge” (p. 16).

Similarly, the QAA Subject Benchmark Statement for Chemistry (2007) included amongst the Chemistry-related cognitive abilities and skills that Chemistry graduates will demonstrate “the ability to demonstrate knowledge and understanding of essential facts, concepts, principles and theories” (p. 11). The benchmark standards articulated in this document require that “a basic knowledge and understanding of the content covered in the course is evident” (p. 13) in all pass-level graduates. As another example, the Tuning Reference Point for Earth Sciences (2009) requires that bachelor level (first cycle) graduates have acquired the key competency: “a broad knowledge and understanding of the essential features, processes, history and materials of System Earth” (p. 19). An examination of Tuning Reference Points documents for other disciplines within science shows that they include similar statements.

At the more general level, in a recent call to arms regarding current cultural norms around science education, Anderson et al., (2011) contended that a university-level science education should ensure that students acquire “broad content knowledge” (p. 152), as well as developing analytical skills and an understanding of scientific research process, inspiring curiosity and preparing students for lifelong learning. There is, therefore, an internationally held consensus view that a tertiary-level science education must include the acquisition of discipline-specific knowledge.

² See <http://www.qaa.ac.uk/ASSURINGSTANDARDSANDQUALITY/SUBJECT-GUIDANCE/Pages/Honours-degree-benchmark-statements.aspx>

³ See <http://www.unideusto.org/tuningeu/home.html>

Assessment of science knowledge

If we acknowledge that content is a core learning outcome of a university science degree, then effective strategies for meaningful assessment of students' science knowledge are required. This is particularly relevant as Australia moves towards full implementation of a national standards framework as the regulatory instrument for the quality assurance of degree programmes (Krause, Barrie & Scott, 2012). As universities are increasingly called upon to monitor and assure academic standards, the standard of assessment of learner outcomes must continue to be improved (Coates, 2012). From the educational perspective, Boud and Associates (2010) stressed that assessment is central to curriculum design because it frames how and what students learn. In particular, they proposed that assessment is most effective when it is designed to focus students on learning, when it is recognised as a learning activity that requires students to engage on appropriate tasks, and when students benefit from useful and informative feedback. How is this premise currently applicable to the assessment of science knowledge and are there significant issues that need to be addressed?

The “content-heavy” science curriculum

The prominence that content *per se* is traditionally given in an undergraduate science curriculum is increasingly recognised as being problematic. Science courses are frequently criticised for being content-heavy. Worryingly, science curricula in general have remained typically static with the emphasis being on content itself rather than students' ability to apply that content knowledge (Matthews & Hodgson, 2011; Stokstad, 2001). The sheer pace of discovery in modern science means that students are frequently overwhelmed and discouraged by the volume of content in their undergraduate courses while teachers struggle to incorporate new material within already crowded curricula (Hoskins & Stevens, 2009). Indeed, a survey of life sciences faculty members showed that they believed that the (perceived) need to cover content mitigated against the teaching of science process skills despite their rating the acquisition of such skills as being very important for their students (Coil, Wenderoth, Cunningham, & Dirks, 2010).

This continued focus on transmission of knowledge is mirrored in an over-reliance on traditional ways of teaching and an over-dependence upon summative assessment. Furthermore, most textbooks, particularly those for early year levels, present scientific information from a “step-by-step accumulation of knowledge” viewpoint (ignoring the fact that the scientific progress is not necessarily linear) (Hoskins, 2008, p. A40), and therefore foster a rote-learning approach. This is despite compelling evidence that encouraging undergraduate students to be actively engaged in their own learning results in higher levels of understanding and knowledge retention than traditional lectures and laboratory classes (DeHaan, 2005; Stokstad, 2001).

Approaches to assessment of science knowledge

How is science knowledge usually assessed?

James (2003), speaking about the Australian higher education system in general, noted that there was a strong emphasis on summative assessment using final examinations coupled with a tendency to over-assess in an attempt to cover prescribed subject content matter. In science undergraduate courses, summative assessments that are “heavy on the testing of content knowledge” continue to dominate (Hanauer & Bauerle, 2012, p. 36).

Two recent Australian studies have probed the ways in which science undergraduates’ knowledge is assessed. A survey of students in the final year of a Bachelor of Science or Biomedical Science at two Australian research-intensive universities by Hodgson, Varsavsky and Matthews (2013) revealed that examinations were a dominant form of assessment and that students rated them most important (88.8%) of all forms of assessment as a method of assessing scientific knowledge. It is worth noting that these students were not asked to comment on their perception of the quality of the assessment type but upon its prominence across their program of study. After surveying Chemistry teaching at twelve Australian universities, Schultz, Mitchell Crow and O’Brien (2013) reported that examinations constitute, on average, about 50% (i.e. represent 50% of the percentage marks awarded) of total assessment in undergraduate Chemistry courses, with practicals being the next most important (mean 28%). The survey suggested that multiple choice testing represents an average of 31% (range 15-49%) of assessment at first year level but is little used in second or third year Chemistry courses. It also showed that online assessment is more likely to be used in first year classes. Similar patterns of assessment practice are likely in other science disciplines.

As further examples, documents on learning, teaching and assessment produced by Tuning Europe for a range of disciplinary areas⁴ show that summative written (open or closed book) and/or oral examinations are considered to be the cornerstones of assessment in undergraduate science courses. It appears, therefore, that traditional examinations continue to be a very significant component of assessment in undergraduate science courses. It can be asked, however, if they provide meaningful assessment of students’ science knowledge?

Traditional examinations as assessment tasks

Traditional forms of knowledge assessment such as unseen closed book examinations have been heavily criticised for being driven by the (perceived) needs of teachers rather than students, and for producing “passive consumers” (Falchikov, 2005, p. 37). The types of questions set in an examination strongly influence students’ study strategies. Examinations, together with tests and quizzes

⁴ Available from <http://www.unideusto.org/tuningeu/documents/teaching-learning-a-assessment.html>

(generally in-class assessment tasks of shorter duration than formally invigilated examinations) often encourage surface learning and discourage retention of knowledge across courses or year levels. If students are only tested on factual recall, then they will learn at that level (Crowe, Dirks, & Wenderoth, 2008; Hanauer & Bauerle, 2012). Furthermore, students rarely receive any useful feedback on their performance in examinations (Hanauer & Bauerle, 2012) so they are unable to make further sense of what they have learned – or not learned. Additionally, traditional examinations favour students who happen to be skilled at dealing with time-constrained assessment tasks (Race, 1999), and only represent a snapshot of an individual's performance on that day – which may be influenced by a number of external factors such as their current state of health (Race, Brown, & Smith, 2005).

The validity of such assessment can also be brought into question: do examinations merely assess whether students can write about what they have read and have been able to remember (Race et al., 2005)? Even those students who perform well in traditional examinations may in fact have a poor grasp of key (or threshold) concepts (Boud, 1990). Hughes and Magin (1996) explained this apparent contradiction using a framework originally devised by Biggs and Collis (1982) who considered that students progress through five stages of ascending complexity as their understanding of unfamiliar material grows and as they move from incompetence to expertise. This Structure of the Observed Learning Outcomes (SOLO) Taxonomy defines five stages of levels of understanding:

1. *Prestructural* – lack of coherent grasp of the material but where isolated facts or skills may be acquired.
2. *Unistructural* – a single relevant aspect may be mastered.
3. *Multistructural* – several elements are mastered separately.
4. *Relational* – several relevant aspects are integrated into a theoretical structure.
5. *Extended Abstract* – stage of expertise in which the material is mastered within its own domain and in relation to other knowledge domains.

Thus, the development of knowledge and understanding is considered to take place along a continuum with students moving from simple unstructured knowledge and understanding to the complex structured and sophisticated knowledge that provides the basis for expert performance (Hughes & Magin, 1996). An emphasis upon testing recall of factual knowledge will not therefore differentiate between students at different stages along this spectrum. Hughes and Magin (1996) presented some useful case studies, four of which are from undergraduate science or engineering, that provide practical examples of strategies with which to assess higher level understanding.

Bloom's Taxonomy (Bloom, Krathwohl, & Masia, 1956) is another very useful framework that can be employed in assessment design for undergraduate science courses (Momsen et al., 2013). Bloom's Taxonomy is often reconstructed pictorially (see, for example, Lord & Baviskar, 2007) as a hierarchical triangle: knowledge (the base level), comprehension, application, analysis, synthesis, and

evaluation (the highest level). Elements of the first level assist with understanding of the next level and so on. The revised version of Bloom's Taxonomy (see Krathwohl, 2002) is particularly relevant in a discussion of science knowledge because it presents the "knowledge dimension" with four (rather than the original three) sub-categories: factual knowledge, conceptual knowledge, procedural knowledge and metacognitive knowledge.

Students often have difficulty demonstrating their competence at the higher cognitive levels of Bloom's Taxonomy which require deep conceptual understanding of disciplinary knowledge (Crowe et al., 2008). Most test or examination questions focus on the lower levels of knowledge and comprehension. Knowledge involving understanding, application and attitudes is rarely assessed (Lord & Baviskar, 2007), but assessment that only requires recall and summarisation of factual knowledge *per se* may not discriminate between students who are at different stages of mastery of the material, as described above for the SOLO taxonomy. Similarly, examinations may disadvantage students if teaching approaches are not aligned to the cognitive challenge of the examination questions: if classroom activities focus mainly on facts and details, but the examination is aimed at a higher cognitive level, then students will tend to perform poorly because they have not been given the opportunity to practice working at that level and to develop a deep understanding of the material (Crowe et al., 2008).

The scope of this paper does not allow a critique of current approaches to undergraduate science teaching. However it is critical that instruction and assessment be aligned, and that both send clear non-conflicting messages about learning expectations and the nature of knowledge in the relevant science discipline (Momsen et al., 2013). Crowe et al., (2008) provide an excellent example of a science undergraduate teaching program designed to enhance student learning (of biology) through the implementation of Bloom's Taxonomy. Their "Blooming Biology Tool" (BBT) is used to determine the level of Bloom's Taxonomy assessed by questions on biology-related topics and versions of the BBT are available for both staff and students. Application of the BBT assists teaching academics to develop better questions and more appropriate learning tasks and helps students develop their own metacognitive skills. Case studies involving an undergraduate physiology course, a biology workshop and a cell biology laboratory class demonstrated that implementation of a BBT-based approach enhanced students' mastery of the subject material.

Multiple choice questions for testing science knowledge

Multiple choice questions (MCQs) are frequently employed as an "effective and efficient method of assessing students' content learning" (Fellenz, 2004, p. 703) and there is an extensive literature on the design of MCQs, their advantages and disadvantages (see, for example, Haladyna, Downing, & Rodriguez, 2002). Such tests are easy to mark (with automated electronic marking possible), which is an

attraction for the assessment of very large classes, and they have the potential advantage that test performance is not reliant on (English) writing ability.

Overwhelmingly, MCQs are used to test recall of factual information (Fellenz, 2004), but Lord and Baviskar (2007) point out that multiple choice questions can be constructed so as to assess higher levels of learning, thus moving science students towards understanding, rather than merely recall, of information. In contrast, Crowe et al., (2008) strongly supported the use of short essay or similar questions in examinations in order to ensure that higher order cognitive skills are being assessed. However, when Palmer and Devitt (2007) analysed both MCQs and modified essay questions (MEQs) used for summative testing in a clinical medical undergraduate course and classified them according to a modified Bloom's Taxonomy. They found that, while over 50% of both MCQs and MEQs tested only factual recall, the MCQs used in these tests were actually better at testing higher order skills than the MEQs.

Such studies highlight the need for teaching academics to be more informed about effective assessment design and more critical of their own approaches to assessment of science knowledge, particularly at undergraduate level. As one example, Schultz (2011) described an innovative approach to assessment of science knowledge. Non-multiple choice randomised assignments aligned with the teaching activities and intended learning outcomes are delivered to a very large (> 300 students) first year class via a learning management system. Advantages of this tool include instant marking, automated mark entry and avoidance of cheating. Importantly, the students can take advantage of unlimited practice questions and receive formative feedback on their learning. Importantly, the questions are designed specifically to address the higher levels of Bloom's Taxonomy.

Other strategies for assessing science knowledge

As an assessment strategy, examinations focus primarily on assessing content knowledge whether at the level of factual retention or at higher cognitive levels. As noted, they are often the most significant item(s) of assessment in core (i.e. compulsory) science units in terms of percentage marks allocated per task. Other commonly employed assessment tasks include practical reports, review essays, posters and oral presentations. Such tasks are usually designed specifically to develop and demonstrate students' acquisition of discipline-specific skills, communications skills or understanding of the processes of science as articulated in the Science TLOs (Jones et al., 2011) but relevant marking rubrics may include criteria relating to the demonstration or application of content knowledge. For example, a universal rubric for assessing students' scientific reasoning skills via scientific writing tasks (Timmerman, Strickland, Johnson, & Payne, 2011, p. 519) requires that the Introduction to the written piece be assessed against the criterion Accuracy, with the descriptor:

Content knowledge is accurate, relevant and provides appropriate background including defining critical terms.

One may argue that such assessment tasks cannot assess the full spectrum of content knowledge covered in core science curricula, and it is important to remember that the Science TLOs, including TLO 2 *Science Knowledge*, apply integratively to the entire degree program (Jones et al., 2011) rather than to single units of study. However, a carefully planned assessment regime can help to move students' thinking and their understanding of content knowledge to higher levels of Bloom's Taxonomy. For example, first year biology students exposed to an innovative guided inquiry curriculum with a diverse suite of assessment tasks (including oral and poster presentations) self-reported using less memorisation and recall and more application and judgment/evaluation than their predecessors in a traditionally taught biology course (Goldey et al., 2012). Such curricula are often put in place in order to improve students' scientific literacy, quantitative and/or communication skills and may not be core or mandatory units. It is telling, therefore, that such curricula may have profound impacts upon students' disciplinary knowledge and conceptual understanding of science.

In summary, assessment that focuses primarily on the lower cognitive levels mitigates against the development of the critical thinking and problem-solving skills that science graduates must acquire (Momsen et al., 2013). Assessment programs for science undergraduate courses therefore need to provide constructively aligned assessment, include both summative and formative assessment, allow assessment of a range of knowledge types, use a range of different assessment tools and address both higher order thinking and the application of knowledge to real-world contexts (Hanauer & Bauerle, 2012). Importantly, such strategies need to be applied from the very first year of a science undergraduate program (Stokstad, 2001).

Conclusion and Recommendations

The Science discipline has already made important and significant moves towards a shared understanding of desired learning outcomes for Australian bachelor level graduates. This paper has focused on identifying some of the key issues around effective and defensible assessment of Science TLO 2. As Coates (2012) pointed out, "developing assessments of performance that simultaneously provide sound information to students, institutions and systems remains a major challenge for higher education" (p. 14).

Academics in science-related disciplines must take collective responsibility for developing, sharing and peer-critiquing best practice in assessment of their students' science knowledge. To achieve this, the still common impediment of strongly content-driven and traditionally taught undergraduate curricula must be overcome. In the light of these challenges, the following recommendations are made:

I: that science academics take individual responsibility for ensuring that undergraduate science curricula are founded upon teaching approaches that foster active learning.

2: that science academics take individual responsibility for designing meaningful assessment tasks that support higher level learning (Fellenz, 2004) and provide incontrovertible evidence of student achievement.

3: that science faculties have in place mechanisms for constructive discussions about the curriculum among academics teaching into all levels of an undergraduate degree program, and a process for internal, formative, peer review of assessment practices and outcomes.

4: that the national science disciplinary community initiate and develop a system of external peer review of assessment of graduate learning outcomes for bachelor-level degrees in science.

References

- Anderson, W. A., Banerjee, U., Drennan, C. L., Elgin, S. C. R., Epstein, I. R., Handelsman, J., ... & Warner, I. M. (2011). Changing the culture of science education at research universities. *Science*, *331*, 152-153. doi: 10.1126/science.1198280
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO Taxonomy*. New York: Academic Press.
- Bloom, B.S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of Educational Objectives: The classification of educational goals*. New York: David McKay.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, *15*(1), 101-111. doi: 10.1080/03075079012331377621
- Boud, D. & Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney, Australia: Australian Learning and Teaching Council. Retrieved from <http://www.olt.gov.au/resource-student-assessment-learning-and-after-courses-uts-2010>
- Coates, H. (2010). Defining and monitoring academic standards in Australian higher education. *Higher Education Management and Policy*, *22*(1), 41-58. Retrieved from http://logincms.uws.edu.au/__data/assets/pdf_file/0004/169168/Coates_HEMP_2010_2.2.2.pdf
- Coil, D., Wenderoth, M. P., Cunningham, M., & Dirks, C. (2010). Teaching the process of science: Faculty perceptions and an effective methodology. *CBE- Life Sciences Education*, *9*, 524-535. doi: 10.1187/cbe.10-01-0005
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to enhance student learning in biology. *CBE-Life Sciences Education*, *7*, 368-381. doi: 10.1187/cbe.08-05-0024
- Dehaan, R. L. (2005). The impending revolution in undergraduate science education. *Journal of Science Education and Technology*, *14*(1), 252-269. doi: 10.1007/s10956-005-4425-3
- Falchikov, N. (2005). *Improving assessment through student involvement. Practical solutions for aiding learning in higher and further education*. London: Routledge Falmer.
- Fellenz, M. (2004). Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Evaluation in Higher Education*, *29*(6), 703-719. doi: 10.1080/0260293042000227245

- Goldey, E. S., Abercrombie, C. L., Ivy, T. M., Kusher, D. I., Moeller, J. F., Rayner, D. A., & Spivey, N. W. (2012). Biological Inquiry: A new course and assessment plan in response to the call to transform undergraduate biology. *CBE- Life Sciences Education*, 11, 353-363. doi: 10.1187/cbe.11-02-0017
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334. doi: 10.1207/S15324818AME1503
- Hanauer, D. L., & Bauerle, C. (2012). Facilitating innovation in science education through assessment reform. *Liberal Education*, 98(3), 434-41. Retrieved from https://www.aacu.org/liberaleducation/le-su12/hanauer_bauerle.cfm
- Hodgson, Y., Varsavsky, C., & Matthews, K. E. (2013). Assessment and teaching of science skills: whole of programme perceptions of graduating students. *Assessment & Evaluation in Higher Education*, 39(5), 515-530. doi:10.1080/02602938.2013.842539
- Hoskins, S. (2008). Using a paradigm shift to teach neurobiology and the nature of science - a C.R.E.A.T.E.-based approach. *The Journal of Undergraduate Neuroscience Education*, Spring 2008 6(2), A40-A52. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23493428>
- Hoskins, S. G., & Stevens, L. M. (2009). Learning our L.I.M.I.T.S.: less is more in teaching science. *Advances in Physiological Education*, 33, 17-20. doi: 10.1152/advan.90184.2008
- Hughes, C., & Magin, D. (1996). Demonstrating knowledge and understanding. In P. Nightingale et al., (Eds.), *Assessing learning in universities* (pp. 127-161). Sydney, Australia: University of New South Wales Press.
- James, R. (2003). Academic standards and the assessment of student learning: some current issues in Australian higher education. *Tertiary Education and Management*, 9, 187-198. doi: 10.1080/13583883.2000.9967016
- Jones, S. M., Yates, B. F., & Kelder, J.-A. (2011). Science learning and teaching academic standards statement. Strawberry Hills, Australia.: Australian Learning and Teaching Council. Retrieved from <http://www.olt.gov.au/resource-learning-and-teaching-academic-standards-science-2011>
- Jones, S. M. (2013). *Good practice guide (Science), Threshold Learning Outcome 2: Scientific knowledge*. Sydney, Australia: Australian Government Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-learning-and-teaching-academic-standards-science-2011>
- Krathwohl, D. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice*, 41(4), 212-218. doi: 10.1207/s15430421tip4104_2
- Krause, K.-L., Barrie, S., & Scott, G. (2012). Mapping learning and teaching standards in Australian higher education: An issues and options paper. Retrieved from http://www.itl.usyd.edu.au/projects/aaglo/pdf/TL%20Stds_Issues%20Paper.pdf
- Lord, T., & Baviskar, S. (2007). Moving students from information recitation to information understanding: Exploiting Bloom's taxonomy in creating science questions. *Journal of College Science Teaching*, 5, 40-44. Retrieved from <http://www.eos.ubc.ca/research/cwsei/resources/Lord%26Baviskar-Blooms.pdf>
- Matthews, K. E., & Hodgson, Y. (2011). Evidencing learning standards in science: Graduate perceptions of gaining knowledge and skills at two research-intensive

- universities. *ACSME Proceedings 2010: Teaching for diversity-challenges and strategies* (pp.114-120). Retrieved from http://www.qsinscience.com.au/wp-content/uploads/2011/12/Matthews-Hodgson_-2011.pdf
- Momsen, J., Offerdahl, E., Kryjevskaja, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE-Life Sciences Education*, 12, 239-249. doi: 10.1187/cbe.12-08-0130
- Nicol, D. (2007). E-assessment by design: Using multiple-choice tests to good effect. *Journal of Further and Higher Education*, 31, 53-64. doi: 10.1080/03098770601167922
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research Paper. *BMC Medical Education*, 7(49). doi:10.1186/1472-6920-7-49
- QAA (The Quality Assurance Agency for Higher Education). (2007). *Subject Benchmark Statement for Biosciences*. Retrieved from <http://www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Subject-benchmark-statement-Biosciences.aspx>
- QAA (The Quality Assurance Agency for Higher Education). (2007). *Subject Benchmark Statement for Chemistry*. Retrieved from <http://www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Subject-benchmark-statement-Chemistry-2007.aspx>
- Race, P. (1999). Why assess innovatively? In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education* (pp. 57-70). Buckingham, UK: The Society for Research into Higher Education.
- Race, P., Brown, S., & Smith, B. (2005). *500 Tips on assessment* (2nd ed.). London: Routledge Falmer.
- Schultz, M. (2011). Sustainable assessment for large science classes: Non-multiple choice, randomised assignments through a learning management system. *Journal of Learning Design*, 4(3), 50-62. doi:10.5204/jld.v4i3.80
- Schultz, M., Mitchell Crow, J., & O'Brien, G. (2013). Outcomes of the Chemistry Discipline Network mapping exercises: Are the threshold learning outcomes met? *International Journal of Innovation in Science and Mathematics Education*, 21, 81-91.
- Stokstad, E. (2001). Reintroducing the Intro Course. *Science*, 293, 1608-1610. doi: 10.1126/science.293.5535.1608
- Timmerman, B. E., Strickland, D. E., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment and Evaluation in Higher Education*, 36(5), 509-547. doi: 10.1080/0260293090354099
- Tuning Project (2009). *Reference points for the design and delivery of degree programmes in Earth Science*. Retrieved from <http://www.unideusto.org/tuningeu/publications/252-reference-points-for-the-design-and-delivery-of-degree-programmes-in-earth-sciences.html>