BACK TO THE FUTURE: FACILITATING ASSESSMENT DESIGN, COLLABORATION, AND OUTCOMES ALIGNMENT WITH AN ASSESSMENT FRAMEWORK

Philip J. MacKinnon[a], Kristy C. Osborne[a], Alan Hayes[b], Louise Lexis[c], Deanne H. Hryciw[d], Michelle Towstoless[e], Kathy Tangalakis[e,f], on behalf of the members of the National Bioscience Assessment Collaboration (NBAC)[g]

Corresponding Author: Philip J. MacKinnon (philip.mackinnon@acer.org)
[a]Australian Council *for* Educational Research, Assessment and Psychometric Research Division, Camberwell, Melbourne, VIC 3124, Australia
[b]Institute for Health and Sport, College of Health and Biomedicine, Victoria University, PO Box 14428, Melbourne, VIC 8001, Australia
[c]Department of Physiology Anatomy & Microbiology, La Trobe University, Melbourne, VIC 3086, Australia
[d]School of Environment and Science, Griffith University, Nathan, QLD 4111, Australia
[e]First Year College, Victoria University, PO Box 14428, Melbourne, VIC 8001, Australia
[fF]Institute for Sustainable Industries & Liveable Cities, Victoria University, PO Box 14428, Melbourne, VIC 8001, Australia
[g]The NBAC includes 35 academics from 10 Australian tertiary institutions.

Keywords: assessment, assessment framework, bioscience, core concepts, intended learning outcomes, PISA, threshold learning outcomes

ABSTRACT

The National Bioscience Assessment Collaboration team constructed and piloted an assessment framework for the biosciences. The framework contains five dimensions, including an adaptation of the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment framework and physiology discipline core concepts.

Pilot mapping of a 72-question physiology exam, containing multiple choice and short answer questions, showed that reliable mapping of assessment to most dimensions of the framework is possible. It also showed that the exam items were skewed on most dimensions of the framework, suggesting a bias to memory rote recall, which has previously been identified as a systemic issue across the biosciences.

The assessment framework can help diagnose where assessment improvement is required in a subject or across a degree and facilitate improvements to assessment design and possible assessment collaboration and benchmarking between institutions.

Proceedings of the Australian Conference on Science and Mathematics Education, 29 September - 1 October 2021, pages 71-77, ISSN 2653-0481

ENSURING ASSESSMENT IS LINKED TO LEARNING OBJECTIVES

In its purest form, assessment measures how well a student performs against a list of criteria. It is a tool for measuring whether the student has achieved a certain level of competency, knowledge and skill. This information is not only vital for the educator but for the student. In the tertiary environment, well-designed assessment should tell the educator and the student, whether the subject learning outcomes (LOs) have been achieved. Assessment can also reassure employers that graduates have the desired attributes to be accredited to a profession and perform professional duties safely and competently (Male, 2021; Boud & Dochy, 2010).

The mode and style of assessment included in a subject influences the type of learning promoted by the subject (Hackling, 2012). In this way, assessment has the dual purpose of measuring achievement and promoting learning (Boud & Dochy, 2010). Any subject with a primary focus on traditional assessment that requires the recall of large amounts of content knowledge under strict, timed test conditions, will promote memorisation and rote learning (Hackling, 2012; Race, 2007). Students' development of competencies, such as techniques and critically evaluating information, are not easily assessed or nurtured in a subject whose assessment is purely knowledge recall (Hackling, 2012). Subject and degree LOs require the nurturing and development of critical thinking skills and a reliable assessment system to measure students' performance in this skill – see, for example, the degree LOs for Science from the Learning and Teaching Academic Standards (Jones, 2011). There are five-degree LOs for Science, and LO *3: Critically analyse and solve scientific problems*; requires science graduates be able to gather, synthesise and critically evaluate information; as well as plan and design investigations; select and apply optimal techniques for conducting investigations; and interpret and draw conclusions from their collected data. To achieve *LO3*, science subjects must nurture, develop, and assess critical thinking skills.

Assessment that provides evidence of the student successfully achieving the LOs needs to be considered at the outset of the subject and overall degree design. Boud and Dochy's 5th proposition for tertiary assessment reinforces this ideal and emphasises the inclusion of a variety of relevant assessment types, tasks, and means of deployment (Boud & Dochy, 2010). Timely deployment of assessment throughout the subject can provide students with a valuable measure of how their learning is progressing and provide educators with a snapshot of which areas the student may require support to reach the degree LOs. Timely assessment feedback often has the most significant influence on ensuring improved student achievement (Hattie, 2009). This feedback must be clear, specific and detailed, with sufficient time for students to utilise the feedback within the subject (Boud & Dochy, 2010; Race, 2007).

The development and implementation of high-quality assessment requires a significant amount of time, resources, insight and collaboration from academics in their respective field, as highlighted by Boud and Dochy (2010), in the 6th proposition for tertiary assessment. Tools, such as the Assessment Review Tool (The Higher Education Academy, 2012), have been developed to determine the assessment's quality. Senior university management can use such tools to ensure high-quality assessment and measure students' progress against the desired graduate attributes. Subject and degree coordinators and teaching academics can also use the tools to ensure assessment items meet the requirement to adequately measure subject LOs. The mix of summative and formative assessment can ensure feedback is scaffolded and provided in a timely and detailed manner.

In particular, the Assessment Review Tool asks questions of teaching academics such as:

- B1.4. Is assessment and feedback planned within and across degrees to ensure appropriate student preparation and practice before the summative assessment occurs?
- B2.2. Does the assessment design process ensure valid assessment of the learning outcomes?
- B2.4. Are assessment decisions about design, development and variety made within a degree context and focused on degree learning outcomes?
- B6.4. Is the potential for bias in professional judgements acknowledged?

The tool provides a Likert scale for academics to judge the quality of their assessment against the questions. A section also provides evidence for the score and notes potential actions to improve the score. Some methods for assessing the quality of assessment and gathering evidence is provided in the resources in Section 4 (The Higher Education Academy, 2012).

One method for ensuring an assessment is indeed valid and can assess the subject and degree LOs is to collaborate with academic colleagues (Boud & Dochy, 2010; Race, 2007). Colleagues can use their tacit knowledge in the discipline, as well as knowledge of the desired graduate attributes to assess: the level of difficulty of an assessment; that the assessment is designed correctly to assess the desired subject LO; and that overall, the assessment has the right balance of items to assess all the desired degree LOs (Race, 2007; Sadler, 1987). However, to ensure students, degree accreditation bodies and future employers, that the subject and its associated assessment provide evidence of students achieving the desired LOs and graduate attributes, a rigorous external standards-referenced assessment design is highly desirable.

BUILDING AN ASSESSMENT FRAMEWORK FOR THE BIOSCIENCES

In 2019, the National Biosciences Assessment Collaboration (NBAC) commenced constructing an assessment framework for the biosciences. An assessment framework provides a conceptual map of the knowledge and skills to be assessed, which academics can use to ensure that assessment measures the appropriate student attributes. The map should at least include the LOs for the coursework but could also include disciplinary core concepts. It would also be desirable to include the competencies required for successful performance as a graduate, such as higher-order thinking and problem-solving skills.

An assessment framework can help provide objective data on student performance, identify where improvements are required, and aid in improving teaching and learning practices. Notably, an assessment framework can show that all subject and degree learning outcomes have been adequately assessed, giving confidence that students have all of the required knowledge and skills (Australian Council *for* Educational Research, n.d.).

Precedents show the potential for using assessment frameworks in higher education. Assessment frameworks are routine for large-scale testing and school assessments. The Australian Medical Assessment Collaboration (AMAC) project is one such example (Australian Medical Assessment Collaboration, 2012). The AMAC project team compiled and validated a bank of multiple-choice questions (MCQs) to assess medical students' clinical knowledge, competencies and higher-order skills — as prescribed by the medical graduate attribute framework (Australian Medical Assessment Collaboration, 2012). The use of these high-quality test items and evaluation of the results has made possible national benchmarking across undergraduate medical degrees in Australasia. Collaboration and benchmarking of assessment between medical schools in Australasia have continued beyond the initial project and funding, demonstrating the sustainability of assessment projects. Similar frameworks, funded by the Organisation for Economic Co-operation and Development (OECD), were developed in the disciplines of economics and engineering (OECD, n.d.).

For the bioscience context, the NBAC adopted the OECD's Programme for International Student Assessment (PISA) 2018 Science framework (OECD, 2018) due to its simplicity, practical orientation and alignment with the Australian Biomedical Science degree LOs (Macaulay & Poronnik, 2014).

One of the advantages of the PISA framework is that it is a good representation of science and assesses more than just disciplinary knowledge—it can thus help ensure that assessment is not just assessing memory recall. The PISA framework conceives science as requiring three competencies underpinned by three types of knowledge. The competencies cover not only the ability to use current theoretical models, but to also design and critique scientific practice and to be able to make sense of the results of scientific enquiry.

The only modification that the NBAC has made to the framework in moving it from assessing 15-year-olds to assessing tertiary coursework, was to extend the second competency to evaluating, designing *and conducting* scientific investigations. Thus, the practical component of bioscience coursework can be captured by the framework.

In addition to PISA competencies and knowledge, we included four other dimensions:

1. The subject LOs: mapping can show the extent to which assessment addresses each subject LO. The subject LOs used in this exercise are those from an introductory human physiology subject, see Table 2.
2. The 14 biomedical degree LOs (Macaulay & Poronnik, 2014) which specify the expected knowledge and skills for a graduate of a degree in biomedical science. The biomedical degree LOs are in five domains: the understanding of science, scientific knowledge, inquiry and problem solving, communication and personal and professional responsibility. The first three domains of the LOs align with, and are complementary to the PISA framework. The final two domains — communication and personal and professional responsibility — extend beyond the PISA competencies and knowledge.
3. The five levels of Bloom’s Taxonomy (Conklin, 2005) which indicate the cognitive demand required of students to complete a question.
4. A list of the disciplinary core concepts which integrate disciplinary knowledge into a coherent whole by identifying and describing the underlying physical processes. Mapping assessment to

the disciplinary core concepts may show how they are understood and applied by students. For this exercise we used a draft version of the core concepts of physiology, recently agreed to by national consensus and in the Australian context (Tangalakis - personal communication) and adapted from those previously published (Michael, Cliff, McFarland, Modell, & Wright, 2017), see Table 2.

The number of dimensions in the NBAC framework is likely to be more than required. For example, the PISA framework and the Biomedical degree LOs overlap, meaning that all elements may not be necessary to map assessment. In addition, subject or degree coordinators could adapt the framework for other purposes by adding or removing elements.

However, to pilot the framework, we included as many elements as practicable to observe how each element performed.

Furthermore, the framework should be complemented by psychometric analysis of assessment item performance. Together, these will show what skills and knowledge are being assessed and the students' performance.

PILOTING THE NBAC ASSESSMENT FRAMEWORK

We trialled the NBAC assessment framework with an exam from an introductory human physiology subject whose LOs are shown in Table 2. This exam had been previously administered to students as an end-of-semester summative exam in one of our Australian institutions. The exam consisted of 60 MCQ and 12 short answer questions (SAQ).

Four academics independently categorised each of the questions in the exam using a Qualtrics questionnaire, and the results were collated and analysed in Microsoft Excel. Three of the academics are experienced university physiology educators. One of the three wrote 25% of the exam paper used for mapping and was coordinator of the subject for the year in which the exam paper was administered. The fourth mapper is a biochemist, an experienced university educator and assessment expert. All are members of the NBAC.

We examined two aspects of the mapping:

1. the consistency of the mapping by the four academics of each assessment item against each of the six dimensions.
2. the distribution of assessment items against each of the dimensions

Consistency of mapping
Mapping consistency will show the reliability of the framework for categorising assessment items. There must be confidence that academics familiar with the content will categorise items similarly.

Table 1: Mapper agreement on each dimension of the NBAC assessment framework. We define mapper agreement as occurring when all mappers select the same option for a dimension of the framework. The table presents the percentage of MCQs and SAQs for which there was agreement for each of the dimensions. For example, the four mappers agreed in their mapping of an MCQ to the subject LOs for 85% of the MCQs.

Framework Dimension	MCQ (%)	SAQ (%)
Subject LO	85	75
PISA Competency	85	92
PISA Knowledge	53	83
Biomedical degree LO	68	67
Core concept	2	8
Bloom’s	55	17

The results are encouraging for using the PISA competencies and knowledge types and LOs as part of the NBAC framework. As shown in Table 1, there was a high level of agreement between mappers for the subject LOs, PISA Competency and biomedical degree LOs for both MCQ and SAQ. The PISA Knowledge dimension had a lower level of agreement for MCQs but a high level for the SAQ.

Agreement when mapping MCQs was similar for Bloom's as for PISA Knowledge. But there was a low level of agreement for Bloom cognitive levels on the SAQs. There were only 12 SAQs in the exam, suggesting that a larger pilot may resolve whether the consistency differences between MCQ and SAQ are real.

The lowest agreement was for the core concepts, with full agreement of the four mappers on less than 10% of MCQs and SAQs. The reasons for this require exploration. One problem is that questions will often need an understanding of multiple core concepts, and thus mappers are selecting one of several core concepts that the question could be mapped to. Also, the relationship between the core concepts needs examination. For example, rather than being equal alternatives, there may be a hierarchy of core concepts.

Mapping reliability on all dimensions would likely be improved by providing training for academics to use the framework, especially for dimensions with which they are less familiar, such as the PISA Competency and Knowledge dimensions.

Distribution of items on the framework

Item distribution on the framework will be an important analytical tool helping to show that students have been adequately assessed on the desired dimensions of the framework. It could be used for this purpose at subject or degree level. The location of items on the framework will also help when comparing assessment items and instruments to each other.

The distribution of mapping choices on each dimension of the framework is shown in Table 2 and is an indicator of the extent to which the exam covers the options for each dimension.

The items in the pilot physiology exam show a restricted distribution on all dimensions with items predominantly aligning with just one of the three subject LOs, the PISA explain competency and disciplinary knowledge domains; and 75% of the items in the exam aligning with one of the 14 biomedical degree LOs. Finally, over 90% of the items were mapped to the Bloom's knowledge and comprehension levels.

The mapping results are consistent with our observations of bioscience assessment from several Australian institutions which suggest examinations focus on simple explanations and recall of disciplinary knowledge, essentially requiring memorisation and rote learning. One caveat to this conclusion is that the skewed distribution on the exam may be acceptable providing other assessment ensures that the remaining desired attributes are assessed.

All of the core concepts were selected by the mappers to varying degrees. Most frequently selected was the concept of structure/function, which is consistent with the most frequently selected subject LO relating to structure and function. Mappers were unable to map a relatively high number of items to any of the core concepts. This needs further investigation as it suggests the concepts' underlying examinable material are not included in the core concept list or mapper familiarisation of the 'unpacked' sub-themes relating to each core concept, is required.

To construct exam items that map to other elements of each of the assessment framework dimensions, could be a focus of future assessment creation. It is possible to write exam items, including MCQs, that demand higher-level cognitive skills for completion and test knowledge of more biomedical degree LOs and discipline core concepts.

Table 2: The distribution of mapping choices on each dimension of the framework for the MCQ and SAQ items for all four mappers. It is expressed as a percentage of total number of choices made by all mappers for each of the dimensions of the framework. Choices were counted for all items regardless of agreement between mappers.

Subject LO	**MCQ (%)**	**SAQ(%)**
1. Describe and relate the structure and function of the cardiovascular, urinary, respiratory and gastrointestinal systems and the role of genetics to normal physiological processes.	92	77
2. Describe how the cardiovascular, urinary, respiratory and gastrointestinal systems act and interact to maintain a constant internal environment (homeostasis).	8	21
3. Develop and demonstrate requisite skills in experimental techniques, recording and critical analysis of data and report writing.	0	0

PISA Competency	**MCQ (%)**	**SAQ (%)**
Explain	93	94
Evaluate	0	0
Interpret	7	6

PISA Knowledge	**MCQ (%)**	**SAQ (%)**
Procedural knowledge	10	4
Epistemic knowledge	2	0
Disciplinary knowledge	88	19

Biomedical Degree LO	**MCQ (%)**	**SAQ (%)**
2.1 Demonstrate well-developed knowledge in at least one disciplinary area in the biomedical sciences.	74	75
2.2 Demonstrate knowledge in other disciplinary areas contributing to the biomedical sciences.	2	0
2.3 Demonstrate integration of knowledge from across the disciplines contributing to biomedical science.	17	25
3.1 Identify, critically analyse and solve problems in the biomedical sciences by collecting, accurately recording, analysing, interpreting and drawing conclusions from scientific data.	4	0

Draft Core Concepts of Physiology	**MCQ (%)**	**SAQ (%)**
Cell-cell communication - The function of the organism requires that cells pass information to one another to coordinate their activities.	0.5	0.0
Cell membrane - Plasma membranes are complex structures that determine what substances leave/enter cell. They are essential for cell signalling, transport, function.	3.6	2.7
Movement of substances - The transport of substances along gradients (ions, molecules, blood, and gas) is a central process at all levels of organisation in the organism.	6.3	0.0
Structure/function - The function of a cell, tissue, or organ is determined by its form. Structure and function (from the molecular level to the organ system level) are intrinsically related to each other.	40.7	43.2
Homeostasis - The internal environment of the organism is actively maintained within narrow limits by the function of cells, tissues, and organs controlled by negative feedback systems.	5.9	16.2
Integration - Cells, tissues, organs, and organ systems interact with one another, and are dependent on the function of one another, to sustain life.	5.9	10.8
Physiological adaptation - Adjusting to changes in the internal and external environment and across the lifespan.	4.1	5.4
None of the core concepts	33.0	21.6

Bloom's Taxonomy	**MCQ (%)**	**SAQ (%)**
Knowledge	79	54
Comprehension	15	35
Application	4	8
Analysis	2	2
Synthesis	1	0
Evaluation	1	0

CONCLUSION AND FUTURE WORK

The current work contributes to the NBAC's objective of developing an assessment framework that can be used in the biosciences to map current assessments, guide assessment development and to compare assessments. We have shown that a good degree of consistency is possible when mapping is conducted by experienced educators. Further testing is required before the framework is ready for use across different disciplines.

Firstly, the framework should be trialled by a wider number of educators and for a greater number of assessments. Such trialling will show if mapping remains reliable under more rugged field conditions.

Secondly, current work on refining and 'unpacking' the physiology core concepts should lead to greater consistency of mapping of these. The NBAC will prioritise trialling the completed core concepts within the Framework.

Finally, the NBAC will consider reducing the number of dimensions in the Framework. In particular, based on the outcome of trialling, it will consider if both of the PISA dimensions and the biomedical degree learning outcomes dimensions are necessary.

REFERENCES

Australian Council for Educational Research (n.d.) Australian Medical Assessment Collaboration (AMAC) https://www.acer.org/au/amac accessed 20 June 2021.

Australian Medical Assessment Collaboration (2012). Assessment framework MAC (2012) http://www.acer.edu.au/documents/AMAC_Framework_May_2012.pdf Retrieved 20 June 2021.

Conklin, J. A. (2005). A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy Educational Objectives, *Educational Horizons, 83*(3), 153-159.

Boud, D., & Dochy, F. (2010). Assessment 2020. Seven propositions for assessment reform in higher education. Sydney, Australia: Australian Learning and Teaching Council.

Buntine, M., Price, W., Separovic, F., Brown, T., & Thwaites, R. (2011). Learning and teaching academic standards: Chemistry academic standards statement.

Hackling, M.W. (2012). Assessment of and for learning in science as part of *The art of science teaching* by Venville and Dawson, Aus: Allen & Unwin.

Hattie, J.A.C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement,* London: Routledge.

Jones, S., Yates, B., Kelder, J.A. (2011). Learning and Teaching Academic Standards: Science Standards Statement, *June 2011*, ALTC report.

Macaulay, J. O., & Poronnik, P. (2014). CUBEnet-Collaborating for Innovation in Biomedical Science Education. *Australian Biochemist*, *45*(3), 5-8.

Male, S. (2021). Getting grades right. *Campus Morning Mail, March 28.*

Michael, J, Cliff, W, McFarland, J, Modell, H, Wright, A. (2017) The Core Concepts of Physiology: A New Paradigm for Teaching Physiology. New York: Springer Nature.

OECD (n.d.) AHELO Main Study, https://www.oecd.org/education/skills-beyond-school/ahelo-main-study.htm, retrieved 20 June 2021

OECD PISA Science Framework (2018) https://www.oecd-ilibrary.org/docserver/f30da688-en.pdf?expires=1623374000&id=id&accname=guest&checksum=3FEB35765D4CE9D684210FDC4EAD79C7 retrieved 11 June 2021

Race, P. (2007). Designing assessment and feedback to enhance learning as part of *The Lecturer's Toolkit,* UK: Routledge.

Rice, J., Low, D., Bachor, H., et al. (2012). Learning and Teaching Academic Standards: Physics Standards Statement, *December 2012*, Accreditation Regulations – Australian Institute of Physics.

Sadler, D. R. (1987). Specifying and Promulgating Achievement Standards, *Oxford Review of Education*, *13*, 2, 191 – 209.

The Higher Education Academy (2012). *A Marked Improvement: Transforming assessment in higher education*. UK: York.