Higher Education Management and Policy
Volume 23/3

An international assessment of bachelor degree graduates' learning outcomes

by
Hamish Coates and Sarah Richardson
Australian Council for Educational Research (ACER), Australia

This paper examines rationales, aspirations, assumptions and methods shaping an international assessment of learning outcomes: the OECD's Assessment of Higher Education Learning Outcomes (AHELO) feasibility study. The first part of the paper is analytical, exploring formative rationales, and shaping contexts and normative perspectives that frame the evaluation. The discussion then turns to review scientific and practical challenges involved in an assessment of the study, which will be tested on an international scale, and to sketch ideas and innovations being created in response. In conclusion, the paper offers reflective suggestions for positioning AHELO in global higher education, should the initiative prove feasible.

Higher Education Management and Policy
Volume 23/3

Une évaluation internationale des résultats dans l'enseignement supérieur au niveau licence

par
Hamish Coates et Sarah Richardson
Conseil australien pour la recherche en éducation (ACER), Australie

Ce document examine les justifications, aspirations, hypothèses et méthodes nécessaires à l'élaboration d'une évaluation internationale des résultats d'apprentissage : l'étude de faisabilité de l'évaluation internationale des performances des étudiants et des universités (AHELO) de l'OCDE. La première partie de ce document est analytique ; elle explore les justifications formatives et façonne les contextes et perspectives normatives qui encadrent l'évaluation. L'examen se penche ensuite sur les défis scientifiques et pratiques liés à une évaluation de l'étude qui sera testée à l'échelle internationale, et esquisse des idées et innovations en cours d'élaboration en réponse. En conclusion, le document propose des suggestions réfléchies sur le positionnement de l'AHELO dans l'enseignement supérieur mondial dans le cas où l'initiative s'avèrerait être réalisable.

Creating evidence to improve learning

This paper discusses work underway to develop and evaluate an international assessment of bachelor degree students' learning outcomes. Taking a research perspective, it examines rationales, aspirations, assumptions and methods shaping the Organisation for Economic Co-operation and Development (OECD)'s Assessment of Higher Education Learning Outcomes (AHELO) feasibility study (OECD, 2011a).* It looks into salient contexts and research foundations and explores a normative vision guiding development and evaluation. The paper then focuses on technical underpinnings and innovations, looking methodologically at solutions being developed and tested in response to challenges and contexts. In conclusion, it reviews next steps required to confirm the theoretical and technical feasibility of this assessment. It provides an introduction to a highly complex and multi-layered study, one with potentially significant implications for higher education worldwide.

The OECD proposed to undertake an international learning outcomes assessment in 2006 (Ischinger, 2006). Between 2007 and 2009 the idea was fleshed out at meetings which affirmed the policy and educational desirability of generating comparative insights into learning outcomes at the international level. Given the foundations and innovation required, participants acknowledged that while such assessment was likely to be possible technically, it was necessary to conduct a feasibility study before launching a full-scale assessment.

The AHELO feasibility study commenced in early 2010 and is scheduled for completion by the end of 2012. The study's broad objective is to determine "whether in an international context it is scientifically and practically feasible to collect objective data on final-year bachelor degree learners' capacity to use, apply and act on their knowledge and reasoning" (OECD, 2010a). Under the overall direction of the OECD and AHELO's Group of National Experts, a consortium of international agencies is developing and validating tests of

* This work has been funded by participating countries and the following foundations and organisations: Lumina Foundation for Education, Compagnia di San Paolo, Swedish National Agency for Education, Higher Education Funding Council for England (HEFCE), Gulbenkian Foundation, Riksbankens Jubileumsfund, Ireland Higher Education Authority (HEA) and Spencer Foundation. The policy steering of the project is carried out by the OECD on behalf of its member countries.

generic, economics and engineering learning outcomes. Some 35 000 students at 230 higher education institutions in 16 countries are collaborating in this venture. The work involves national centres in participating countries, international expert groups and a wide range of higher education stakeholders.

AHELO reflects an innovative fusion between educational measurement and policy research. The following methodological overview provides a helpful prelude for the balance of the paper. Supported by policy makers and institutional leaders, the study began with test developers and assessment experts working together to determine learning outcomes which students in specific fields of study should have achieved towards the end of a bachelor's degree. The developers worked with experts and stakeholders to design assessment instruments that map to these outcomes. As discussed later in this paper, these instruments go through a complex process of qualitative and psychometric validation, linguistic translation, cultural adaptation and independent verification to ensure they are measuring equivalent constructs in different languages. The tests are complemented by student, faculty and institutional questionnaires designed to collect demographic and contextual information. National managers in participating countries are responsible for co-ordinating system-level activities and are provided with ongoing training and support. Students are scheduled to take the tests in the first six months of 2012. Test data will then undergo rigorous psychometric modelling and evaluation to inform conclusions about the validity, effectiveness and relevance of an international assessment of student learning. Results will be reported to institutions and participating systems; lastly, a series of international reports will be prepared.

Shaping contexts and rationales

AHELO is being developed at a time when there is pressure to conduct more performance assessment in higher education. There is a multiplicity of rationales for this, not least the desire to better understand the transparency, effectiveness, diversity, productivity and accountability of an expanding sector. Within this broad framework, it is useful to review the main precursors and rationales that position and prompt the study. Of course, many motivating forces go beyond higher education: one particularly powerful driver is interest in knowing more about highly skilled individuals and the way they move around the world. The following analysis is more modest in intent, and concentrates on large-scale research and policy developments within higher education.

AHELO reflects a fundamental move beyond traditional collegiate approaches to assuring the quality of graduate outcomes. Traditionally, the

definition and assessment of learning outcomes has been an internal matter for universities, nuanced in various ways by interactions with regulatory agencies, professional bodies, other parts of the education system and the economy as a whole. But there are signs that in expanding systems traditional collegiate approaches to defining, assessing and monitoring academic standards no longer suffice to yield generalisable data on what graduates have learned and can do (AUQA, 2009; Coates, 2010; Ewell, 2009; Ikenberry and Kuh, 2009; Salmi, 2009; Spellings, 2006; VSA, 2009). This is not surprising, for these approaches were designed for elite systems rather than those much larger in scale. Developing methods for assessing what students know and can do, and for comparing outcomes across institutions can, in combination with numerous other kinds of formative assessment and evaluation, offer an empirical anchor which can be used by a variety of stakeholders – not least institutions and faculty – to underpin determinations of graduate competence.

The growth of the international "quality movement" over the last few decades reflects the same interest in reinforcing and extending traditional forms of quality assurance. Building on practice within the United States, as of the mid-1980s quality assurance systems emerged around the world that were essentially based on a three-phase model of self-study, external peer review and public report (Vught and Westerheijden, 1993). This process has naturally led to questions about international comparison and, directly and indirectly, quality assurance processes have provided the foundations for a considerable amount of benchmarking. The development of a robust quality culture and community is an important antecedent for AHELO, but the focus on inputs and processes exposed an opportunity – and indeed a need – to develop comparable information about what learners actually achieve. In calling for quality assurance that is more aligned to the needs of society Massaro (2010, p. 25), for instance, argues that "it is now urgent that the metrics be developed to measure standards and outcomes using some valuable examples as starting points".

Of course much quality-relevant data do exist, produced by institutions, systems and transnational networks. The collection of data on student engagement, for instance, has stimulated important discussions in several countries about learning processes and outcomes. Similarly, there has been a proliferation of surveys measuring students' satisfaction with educational services. Such collections can provide useful insights, particularly when linked with outcomes' correlates, however the data collected are subjective in nature and focus on educational processes rather than learning outcomes. Assessment collaborations, such as the United Kingdom's Medical Schools Council Assessment Alliance (MSC-AA, 2011) or the Australian Medical Assessment Collaboration (ACER, 2011), move one step further, delivering

learning outcomes data which can be generalised beyond local contexts. Ultimately, such initiatives are undertaken for formative purposes, flavouring the appropriateness and adequacy of their use in external initiatives. Various national approaches to assessing outcomes do exist (Nusche, 2008), yet few span national boundaries. Hence a particularly important rationale for a study such as AHELO is the production of consistent disciplinary or institutional information on learning outcomes that is international, and potentially global, in scope.

The proliferation of global rankings testifies to the hunger of systems, institutions and individuals for internationally comparable data on what universities achieve. Prominent initiatives include the *Times Higher Education* (TSL Education, 2010), the Shanghai Jiao Tong index of "world-class universities" (CWCU, 2010), and the *US News and World Report* Best Colleges (US News, 2010). With myriad others, these developments have driven a "rankings movement" which has considerably sharpened focus on data-driven cross-institutional comparison. Both the rankings and the discourse that surround them are a direct prompt for AHELO inasmuch as they highlight a need to compensate for the shortcomings of existing metrics. The lack of data on learning, the preoccupation with scientific research, the focus on whole institutions and the compression of institutional types create space and need for comparative data on learning outcomes. By far the most significant attempt to overcome many limitations of prevailing rankings is the U-Multirank project (CHERPA-Network, 2010). U-Multirank is, in many senses, a natural companion project for AHELO, having aspirations to increase the validity, scope, diversity and transparency of information on higher education.

Stemming from policy change within Europe, the Tuning Process (Tuning Association, 2011) is another important prompt for cross-institutional learning assessments. Beginning in 2000, "Tuning" is a process which involves describing and aligning degree outcomes. Working with communities of scholars, Tuning reflects the need for universities to sustain their unique missions within collaboratively determined frames of reference. The work, now expanded into the Americas and other regions of the world, provides important preconditions for AHELO as it spurs conversations about learning outcomes, encourages scholars to consider curricula and qualification comparisons and produces cross-national communities of practice. In 2008 and 2009 Tuning provided a direct input into AHELO through foundation work undertaken to map broad learning outcomes in the selected fields of economics and engineering (Tuning Association, 2009a, 2009b).

Transparency initiatives such as U-Multirank and Tuning, along with other large-scale initiatives, are important stimuli for AHELO and are reflective of more general trends. The shift from elite to mass – and in certain countries, to universal – systems of higher education has multiplied the

number of stakeholders with an interest in higher education outcomes. No longer limited to reproducing the high standing of the elite, higher education has taken on a broader responsibility for educating larger sections of the population. Greater interest and scrutiny – and hence greater transparency – is an inevitable consequence of this growth. While it would appear to be among the most important pieces of information on higher education, public data on what learners know and can do remain scarce in many systems.

This snapshot sketches key trends shaping the collection of generalisable data on students' learning outcomes. As it illustrates, significant foundations exist, but to date work has focused on elements of higher education which are relatively easy to compare such as institutional structures, educational processes and programme content. Explorations of institutional and discipline-specific performance in student learning are growing from a low base. Yet greater insights into comparative learning across different institutional types and missions is of increasing importance in an environment of ever-growing global mobility of students and graduates, and in the context of finding viable policy solutions to sustaining universal higher education. Despite all the foregoing developments and initiatives, there remains a need for rigorous and generalisable measurement of student learning outcomes which is comparable across institutions and across national systems. A feasible AHELO has the potential to fill this gap, building on fertile ground created through various projects undertaken over recent decades.

Aspirations and guiding principles

The collection of information on learning outcomes can be viewed as one means of spurring large-scale reform at the institutional level. Drawing from the AHELO assessment design (OECD, 2010b), the "guiding principles" elaborated here detail aspects of this intent. These principles are deliberately forward looking and aspirational. They are intended to contribute foundational insights into what a feasible assessment of higher education learning outcomes could look like. While perhaps lofty, in the absence of precursor studies they provide a positive normative framework within which innovation and evaluation can proceed.

Basically, the development work rests on an assumption that it is indeed possible to undertake an international assessment of final-year students' capacity to use, apply and act on their knowledge and reasoning. This, in turn, implies that new methodologies and technical standards can be established for higher education research. It means that policy makers, institutional leaders, faculty and students can be engaged, and that they see assessment

processes and outcomes as valuable information on education. It implies that institutions can take steps to convert results into improvement-oriented change, industry and government leaders can see new possibilities for assessing graduate capability and international education can make use of a new data source. Importantly, it implies that learning outcomes data are seen to offer a significant, effective and additional means of understanding higher education.

The international dimension is vital for AHELO and the OECD hopes that comparative data on learning outcomes will provide a powerful force for institutional benchmarking. Simply by asking institutions to participate in an international assessment sends a message that they are part of an interdependent global knowledge community. Providing multidimensional reports to institutions highlights international learning networks and the transnational flow of knowledge. Learners and graduates should also benefit from better information on outcomes that is international in scope. When reported in a sound fashion, information on outcomes has the potential to offer powerful insights that prospective learners can use to inform study decisions. Any benchmarked performance feedback provided to learners and graduates offers insights to them and to prospective employers or graduate institutions that might assist mobility and placement in a whole range of ways.

As a guiding structure for evaluation, it is assumed that measuring later-year students' learning and capacity to perform will become a routine facet of higher education practice. Such metrics offer information that complements conventional assessments of academic achievement, and facilitate progressions into further study and professional practice. At the same time, the information can provide institutions and faculty with a reference point against which to estimate the efficacy of their own goals and practices.

But to add value, and to avoid stifling diversity, which can be dangerous, such assessments need to go beyond testing knowledge. They must test students' capacity to reason in complex and applied ways, and to use skills and competencies effectively in different settings. The assessments need to be sophisticated and to align with the forms of thinking and professional work in which most graduates will engage. They need to employ a wide range of methods, provide for a more balanced view of higher education quality, and tap into capabilities that both educators and professionals recognise as important for educational success – capabilities such as collaboration and teamwork, oral and written communication skills, creative and analytical abilities, and leadership. It is important to take account of the disciplines within which students learn and the trans-disciplinarity of professional life. Developing such assessments requires innovation. It requires conceptualising

new constructs, developing items and systems for capturing high-level reasoning and for reporting in informative and informed ways.

In a sector increasingly driven by competitive forces, it is vital that AHELO stimulates what Vught and Kaiser (2008) refer to as diversity and diversification. With the notable exception of U-Multirank (CHERPA-Network, 2010), existing rankings tend to focus on vertical institutional diversity and are unable to address programmatic diversity (Marginson and van der Wende, 2007). By developing multidimensional approaches that move beyond standardised institutional rankings, information about learning outcomes has the capacity to inform and enhance each institution's distinct mission and autonomy, as well as subsequent efforts to improve performance.

Fuelling evidence-driven continuous improvement is a powerful guiding principle. Collecting and reporting statistical data, regardless of how novel and intriguing, is rarely sufficient to prompt institutional or systemic change. Hence to yield successful returns for institutions, systems and stakeholders, AHELO must embrace broader plans for engaging institutions, faculty and students in evidence-based change. Large-scale institutional or programme-level assessments can be difficult to link with practice unless clear strategies are in place to help leaders, teachers and students access and use information (Coates and Seifert, 2010). Without strategies to link reports in meaningful ways across systems, institutions and individuals it is difficult to energise the people that make change happen. Hence an overarching engagement strategy which links individuals and institutions, disciplines and faculty, leaders and systems must be at AHELO's core. Indeed, participating in an assessment process should be an informative experience in itself.

Challenges and innovations

AHELO is *sui generis*, and in its infancy. To be successful, innovation of such scope needs to explore and overcome major scientific and practical challenges – many of which may be unknown or unexpected. Constructing an assessment that is valid across institutions, cultures and disciplines means accounting for factors such as institutional diversity, differences between national systems of higher education, selectivity – which is inherent to systems and institutions –, as well as variations in the duration and content of programmes. Other aspects which also need to be considered are how to motivate students and institutions to participate, how to ensure a fair assessment of institutions and programmes, and cultural and linguistic diversity. The remainder of this paper is essentially methodological. It explores developmental challenges and design innovations intended to initiate the assessment.

Engaging change: converting the controversial to the common

Engaging a large global industry around new forms of data that relate to core educational business is a significant proposition. A major amount of work is required to understand philosophical, political and historical scepticism to assessment innovation and to inspire stakeholders to see "learning data" as part of the future higher education scenario. While global trends provide an increasing predisposition for AHELO, considerable work is also required to link rationales with organisational and educational realities. This task is complicated by any perception or assertion that generalised scientific assessment hinders – rather than enhances – institutional or faculty practices, autonomy or experiences. Clearly, the existence of information opens possibilities for its misuse, but such risk can be minimised by sound design and regulation of data and reports. Indeed, as discussed below, well-formed reports carry a real potential to help higher education institutions and systems innovate, grow and respond to the several challenges that institutions face.

Since 2006 AHELO's innovators (policy leaders and technical experts) have engaged early adopters (participating ministries, institutions, stakeholders and experts). Country participation has grown from 7 in 2008 to 16 in late 2011, with a rise from 70 to 230 in the potential number of participating institutions. Technical development underway since 2010 has engaged many hundreds of organisations and individuals. To be feasible, AHELO needs to expand further and involve leaders, faculty and students in the data collection and reporting process. Engaging the global higher education system more broadly lies beyond the scope of a feasibility study. But there are early signs that this may be possible, given a notable shift, since 2006, from discourse questioning rationales and methodology to discussions about more tangible matters such as operations, results and timelines. Ultimately, as with existing metrics for admissions and research, it is possible to imagine a culture in which assessment of learning outcomes is integral to the architecture of higher education.

The prospects of such change hinge, to a large extent, on the capacity of systems to lead required reform. While AHELO is led and developed by independent agencies, national centres in participating countries take considerable responsibility for implementation. Various models have been used within countries to establish centres that combine higher education policy expertise with educational measurement capability. Around one-third of the countries participating in the feasibility study has housed national operations within government-affiliated research agencies; a further third is working from university research centres. National centres in remaining countries are located at independent research agencies, quality agencies, or in

peak bodies. National centres are funded and managed in a range of ways, and draw on external experts for translation, content advice and analysis. Together, the centres spotlight new approaches to leading large-scale assessment work within higher education. They highlight the emergence of new institutional architectures, practitioner networks and research communities.

Balancing generalisability and specificity

Managing tensions between specificity and generalisability is one of the greatest challenges in research that crosses cultural, linguistic, disciplinary, programmatic, national, curricula and institutional boundaries, amongst others. To be valid, assessments must be relevant to local contexts of measurement and reporting, and also sufficiently global to enable comparability. Defining and managing or resolving tensions between specificity and generalisability is essential to the success of such research. In many areas it is possible to apply proven methods (OECD, 2009), even though these may be novel or less common in higher education. In other instances, innovative assumptions, perspectives and design solutions are required.

For instance, a considerable amount is known about methods for establishing linguistic and cultural comparability, drawing on experience in studies such as PISA (OECD, 2011b) and TIMSS (IEA, 2010). It has been possible to adopt and adapt approaches proven with school-level research for AHELO. In brief, English source versions of test instruments undergo "dual translation" in participating countries, involving two professional translators working independently to arrive at separate translations. These translations are then "reconciled", a process whereby domain and language experts review translations and retain the strengths of each to arrive at a superior composite version. Except for generic skills, the reconciled translation is then verified independently, a process that involves further scrutiny by domain and linguistic experts to ensure the translated material is optimised for a given national context. Even where translation is not required, the English source version passes through a stringent process of adaptation and verification to ensure that an optimal national version is delivered. The implementation of a detailed multi-stage approach ensures that translated assessments are linguistically and culturally equivalent to source versions, facilitating the reliable measurement of student performance internationally.

Of course, it is critically important to achieve clarity in relation to the outcomes to be assessed. Combining "Tuning" processes with techniques used by measurement scientists to produce assessment frameworks offers a practical and robust solution. In AHELO, test developers draw on current developments (for instance, Tuning Association 2009a, 2009b; Quality

Assurance Agency, 2006) to define expected competencies. As anticipated, given the positioning of these competencies on the qualification hierarchy, they involve more than the basic reproduction of accumulated knowledge, calling on students to mobilise thinking skills to solve real-world problems (OECD, 2003). Assessment developers work closely with advisory groups composed of domain assessment experts from around the world to agree key competencies and summarise these in assessment frameworks. The purpose of the frameworks is to provide a rigorous backbone to guide the development of assessment tasks, and test developers are able to demonstrate that the assessment tasks map closely to the assessment frameworks.

When developers focus on outcomes it is not necessary to harmonise programmes, curricula or pedagogy, but it remains important to account for any implications arising from differences between educational structures, resources and approaches. A specific engineering competence, for instance, must be assessed in ways which reflect its manifestation in the curriculum, institutional culture and mission, as well as the epistemologies that shape teaching. This is a significant point that must be addressed to avoid irrelevance or retreat into scholarly particularity or relativism. This calls for methods and evaluation strategies that balance practical, philosophical and scientific considerations. Much can be proven through standard measurement science. Other uncertainties can be resolved through wide-scale consultation or expert review. Various practicalities imply certain necessary compromises.

A scientific sampling strategy

A well-designed, managed and regulated sampling approach is critical to any international educational assessment. Without controlling who and how many people take the test, it is not possible to assure the statistical power and relevance of estimates – even with *post hoc* statistical adjustment. An international assessment such as AHELO is challenged in several ways by this requirement, for while the technical principles of sampling are sustained these have not been applied at scale in post-compulsory education (with the notable exception of TEDS [IEA, 2011]); uncertain adaptations may be required and various conventions and precedents do not exist.

Developing a pertinent sampling strategy that affords appropriate balance between technical, practical and substantive considerations is an important facet of AHELO. Within the constraints of a feasibility project, the sampling strategy seeks to secure the voluntary participation in the test of a sufficient number of randomly selected individuals or groups. Participating individuals will interact with a stratified random sample of test items which are mapped against the contents and constructs given in the test instrument specification. Underpinned by the kind of sophisticated methodological reasoning that informs school-level assessments (OECD, 2009), the design is

expressed in simple ways that can be adapted to systemic and institutional contexts. A controlled and devolved approach is used whereby national centres and international experts reach agreement on national approaches; national centres work with institutional co-ordinators to draw samples, then international monitors verify and assure the comparability of participant yields. The techniques and their application have been well-rehearsed in school-level assessments, but their relevance in AHELO hinges on whether it is possible to engage targeted students in post-compulsory education. Significant analysis will be required, particularly given the lack of various benchmarks, to determine whether the sampling yields robust and defensible performance estimates.

Quantifying complex cognitive responses

While well-designed multiple choice questions can measure complex real-world thinking, many advanced forms of reasoning are best measured using tasks that require students to construct their response in the form of writing or drawing rather than selecting from a set of pre-defined alternatives. Administering such "constructed response tasks" necessitates using automated or manual scorers. Designing valid and efficient approaches to scoring constructed response tasks is another area requiring methodological innovation. As with sampling, many (if not all) of the scoring methods required for AHELO have been developed and tested in diversified and large-scale contexts. AHELO adds extra dimensions, however, inasmuch as it requires consistency in difficult and complex cognitive domains, and traverses many educational, cultural and disciplinary contexts. Data produced from students' constructed responses will not be usable without effective scoring strategies, operations and quality controls. If the same response is scored differently by different scorers, or even by the same scorer at different points in time, reliability suffers, which in turn threatens validity and results.

With this challenge in mind, scoring methods have been designed to enact desired technical principles in large, devolved and diverse contexts. Each country selects a senior academic with expertise in economics, engineering or generic skills as lead scorer, with scorers undergoing rigorous technical and practical training. This training involves methodological analysis and comparing benchmark student responses from participating countries with rubrics produced by test developers. This process enables detailed discussions of how score points are allocated to differing responses and allows lead scorers to become thoroughly immersed in the intricacies of quantifying student response. At the same time, any linguistic and cultural differences can be highlighted and addressed, facilitating international consistency. Prior to national scoring, lead scorers recruit and train a national scoring team. Scoring is conducted using standardised online tools with lead

scorers closely monitoring other scorers' work to ensure individual and national consistency and randomly auditing scored responses to determine scorer accuracy.

Despite the careful design of scoring procedures in AHELO, it is important to note that methods for scoring tertiary student test responses across national boundaries are at an early stage of technical development. Lessons can be gleaned from large-scale national assessments, but much needs to be established during AHELO's lifetime and innovative solutions will need to be found to problems which arise during scoring activities. Particular challenges lie in scoring constructed response tasks across cultures and languages, where not only syntax, vocabulary and script vary, but there are also differences in accepted approaches to communication and the nature and description of reasoning.

Reports that engage stakeholders and prompt change

Studies like AHELO are not conducted to generate statistical estimates, but to drive productive reform. Perhaps the biggest challenge confronting AHELO is engaging stakeholders in evidence-based change. The real value and contribution of the assessment derives not just from reading reports, but from international communities being formed to design and construct tasks, participate in the assessments, and use results to guide individual, institutional and system growth. In a study with global reach, in which an assessment cycle might span years, it is important to create opportunities for engagement. Effective engagement is decisive for the study's impact on reform.

The study's approach engages stakeholders in each phase of the work. By way of example:

- A stakeholder consultative group – including a broad group of higher education stakeholders (associations of higher education institutions, student and teachers' unions, quality assurance agencies, professional associations and business groups – has been engaged in the work since its inception, shaping design, development and review.
- Even within the constraints of the feasibility study, framework and test development has involved hundreds of experts and faculty in structured conversations about learning.
- Along with the tests, context questionnaires have been developed to engage institutional representatives and faculty in data collection, and help ensure analyses and reports resonate with local contexts.
- Taking part in an international assessment has the potential to be an intrinsically engaging experience for participating countries, institutions, faculty and students.

- While respecting necessary confidentiality and privacy constraints, reports are being designed to be as transparent and informative as possible.
- Participating countries are already establishing support mechanisms to help systems and institutions identify how to interpret reports for the purposes of diagnosis, monitoring and improvement.

By its conclusion, it is hoped that AHELO will have yielded important new insights into core facets of higher education. At the same time, it will have taken only the first tentative steps in a large and growing field of higher education. Significant work will be required to link AHELO with existing harmonisation, ranking and classification exercises, to determine the broader strategic significance of such assessment and to build capacity within funding, regulatory and quality agencies. Efforts will also be needed to boost institutional capacity to manage and understand outcomes data, to assist students and their parents to make better sense of information on higher education and to support emerging research communities that underpin applied work with methodological and scholarly inquiry.

Initial impressions, and strengthening foundations

This paper offers an introduction to efforts underway to build an international assessment of bachelor degree learning outcomes: work that touches major aspects of higher education. It has surveyed contexts driving the study and methodological innovations required for a feasible outcome. Given the breadth and potential significance of this work it has only been possible to provide a brief overview of the highly complex and multi-faceted research involved. By the end of 2012, it is anticipated that data will have been collected from students and faculty at some 230 institutions worldwide, and then subjected – along with myriad other forms of data – to multifaceted analysis and review. Only then will it be possible to evaluate the feasibility of measuring learning outcomes across cultures in ways which are valid, reliable and practical.

The evaluation framework developed at the start of the study (OECD, 2010c) sets forth the qualities required of a feasible international assessment. In terms of instrumentation, it will be necessary for each assessment – engineering, economics and generic skills – to reflect international consensus about the content areas that it is important to assess. From an operational point of view, the instruments should reflect the spirit and intent of the content specification, hence it will be necessary to secure international consensus so that the student, faculty and institutional questionnaires capture important contexts that shape higher education learning outcomes. To be successful in terms of implementation, the assessment will need to be practical and methodologically rigorous, be generalised cross-nationally,

cross-culturally, cross-linguistically and cross-institutionally. It will also need to effectively engage systems, institutions, faculty and a random sample of students, be delivered successfully and in a consistent way across countries, be scored in cross-linguistically and cross-culturally generalisable ways and be reported in ways that engage systems and institutions.

As this paper has indicated, work is proceeding against a backcloth of numerous diverse and stimulating research dynamics. Essential to such work is that a large number of international agencies share a growing interest in collaborating in order to better understand and improve higher education outcomes. It is essential to produce valid and reliable assessment resources and processes. This should include a consultative and technically rigorous production of assessment frameworks and test instruments, the effective operationalisation of assessment materials, deployment using quality-assured and efficient implementation methods, and production of informative data products and reports. Adopting a multidimensional and multidisciplinary stance is vital to drive improvement and diversity. Confidentiality and security are obviously intrinsic to testing, but these must be balanced with transparency. Clearly, in times of financial volatility any initiative with global intent must be affordable and scalable.

Learning is core to higher education and data on learning outcomes is relevant to a wide range of stakeholders. Significant work will be required to link outcomes assessments with existing harmonisation, ranking and classification exercises. It is necessary to build national capacity within funding, regulatory and quality agencies, and to boost capacity within international communities to manage and understand outcomes data. Over time, work will be required to help students and their parents make better sense of this new information on higher education. More research communities will emerge to underpin applied work with methodological and scholarly inquiry. Current research is yielding important new insights into higher education. At the same time, it reflects only first tentative steps in a significant and emerging field.

The authors:

Hamish Coates, PhD (corresponding author)
Australian Council for Educational Research (ACER)
19 Prospect Hill Road
3124 Camberwell VIC
Australia
E-mail: coatesh@acer.edu.au

Sarah Richardson, PhD
Australian Council for Educational Research (ACER)
19 Prospect Hill Road
3124 Camberwell VIC
Australia
E-mail: richardsons@acer.edu.au

Acknowledgement

Hamish Coates is AHELO Project Director, and Sarah Richardson manages the international work. They are extremely grateful to many colleagues who provided advice on aspects of the research and sections of this text.

No specific funding source was called upon to support the research presented.

References

ACER (Australian Council for Educational Research) (2011), *Australian Medical Assessment Collaboration*, *www.acer.edu.au/amac*, accessed 1 July 2011.

AUQA (Australian Universities Quality Agency) (2009), *Setting and Monitoring Academic Standards for Australian Higher Education: A discussion paper*, AUQA, Melbourne.

CWCU (Center for World Class Universities) (2010), *Academic Ranking of World Universities 2010*, *www.arwu.org/ARWU2010.jsp*, accessed 3 March 2011.

SHERPA-Network (2010), *Design and Testing the Feasibility of a Multi-dimensional Global University Ranking*, *www.u-multirank.eu*, accessed 20 September 2011.

Coates, H. (2010), "Defining and monitoring academic standards in Australian higher education", *Higher Education Management and Policy,* Vol. 22, No. 1, OECD Publishing, pp. 1-17.

Coates, H. and T. Seifert (2010), "Linking assessment for Learning, Improvement, and Accountability", *Quality in Higher Education,* Vol. 17, No. 2, pp. 179-195.

Ewell, P. (2009), "Stuck on Student Learning", in *Measuring Up 2008: The National Report Card on Higher Education*, The National Centre for Public Policy and Higher Education, San Jose.

IEA (International Association for the Evaluation of Educational Achievement) (2010), *Trends in International Mathematics and Science Study (TIMSS)*, *www.iea.nl/timss_2011.html,* accessed 1 September 2011.

IEA (2011), *Teacher Education and Development Study in Mathematics (TEDS-M 2008)*, *www.iea.nl/teds-m_2008.html*, accessed 1 September 2011.

Ikenberry, S. and G. Kuh (2009), *More Than You Think, Less Than We Need: Learning Outcomes Assessment in American Higher Education*, Indiana University, Bloomington.

Ischinger, B. (2006), "Higher Education for a Changing World", *OECD Observer,* pp. 255.

Marginson, S. and M. van der Wende (2007), *Globalisation and Higher Education*, OECD Publishing.

Massaro, V. (2010), "Cui bono? The relevance and impact of quality assurance", *Journal of Higher Education Policy and Management*, Vol. 32, No. 1, pp. 17-26.

MSC-AA (Medical Schools Council Assessment Alliance) (2011), *London Medical Schools Council, www.medschools.ac.uk/MSC-AA/Pages/default.aspx*, accessed 1 September 2011.

Nusche, D. (2008), *Assessment of Learning Outcomes in Higher Education: A Comparative Review of Selected Practices*, OECD unclassified document, *www.oecd.org/dataoecd/13/25/40256023.pdf*, accessed 3 June 2008.

OECD (Organisation for Economic Co-operation and Development) (2003), *The Definition and Selection of Key Competencies (DeSeCo): Executive Summary*, *www.oecd.org/dataoecd/47/61/35070367.pdf*, accessed 12 February 2010.

OECD (2009), *PISA 2009 Technical Report, www.oecd.org/document/19/0,3746,en_2649_35845621_48577747_1_1_1_1,00.html*, accessed 1 September 2011.

OECD (2010a), *Roadmap for the OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study*, *www.oecd.org/dataoecd/50/50/41061421.pdf*, accessed 19 August 2011.

OECD (2010b), *AHELO Assessment Design*, *www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2010)17&doclanguage=en*, accessed 7 July 2011.

OECD (2010c), *AHELO Feasibility Study Analysis Plan*, *www.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2010)18&doclanguage=en*, accessed 15 July 2011.

OECD (2011a), *OECD AHELO*, *www.oecd.org/edu/ahelo*, accessed 1 September 2011.

OECD (2011b), *OECD PISA*, *www.pisa.oecd.org*, accessed 1 September 2011.

Quality Assurance Agency (2006), *Subject Benchmark Statement: Engineering*, *www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Subject-benchmark-statement-Engineering-.aspx*, accessed 2 October 2008.

Rogers, E.M. (2003), *Diffusion of Innovations*, Free Press, New York.

Salmi, J. (2009), "The Growing Accountability Agenda in Tertiary Education: Progress or Mixed Blessing?", *Education Working Paper Series*, No. 16, the World Bank, Washington, p. vii.

Spellings, M. (2006), *A Test of Leadership*, US Department of Education, Washington.

TSL Education (2010), *World University Rankings 2011*, *www.timeshighereducation.co.uk/world-university-rankings*, accessed 10 June 2011.

Tuning Association (2009), *A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in the Science of Economics*, *www.oecd.org/edu/ahelo*, accessed 1 September 2011.

Tuning Association (2009), *A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering, www.oecd.org/edu/ahelo*, accessed 1 September 2011.

Tuning Association (2011), *Tuning Educational Structures in Europe*, *www.unideusto.org/tuningeu/home.html*, accessed 25 March 2011.

US News and World Report (2011), *Best Colleges*, *www.usnews.com/education*, accessed 28 September 2011.

Van Vught, F.A. and F. Kaiser (2008), *Mapping Diversity: Developing a European Classification of Higher Education*, CHEPS, Eschede.

Van Vught, F.A. and D.F. Westerheijden (1993), *Quality Management and Quality Assurance in European Higher Education: Methods and Mechanisms*, Commission of the European Communities, Luxembourg.

VSA (Voluntary System of Accountability) (2009), *Voluntary System of Accountability*, *www.voluntarysystem.org*, accessed 1 February 2009.