

Unclassified

EDU/IMHE/AHELO/GNE(2012)5

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

12-Mar-2012

English - Or. English

DIRECTORATE FOR EDUCATION
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD

Group of National Experts on the AHELO Feasibility Study

REVISED INTERIM FEASIBILITY REPORT

9th meeting of the AHELO GNE

Paris, 19-20 March 2012

This document was prepared by the ACER Consortium.

The AHELO GNE is invited to TAKE NOTE of this document.

Contact:

Consortium: ahelo@acer.edu.au

OECD Directorate for Education: Diane.Lalancette@oecd.org

JT03317552

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

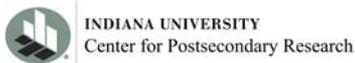
EDU/IMHE/AHELO/GNE(2012)5
Unclassified

English - Or. English



AHELO Feasibility Study Interim Report

March 2012 – Revised version



Statistics Canada Statistique Canada



TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS	4
INTRODUCTION	5
Overview of this report.....	5
Key insights to date.....	5
AHELO FOCUS, RATIONALES AND CONTEXTS.....	7
Overview of the AHELO Feasibility Study	7
Broad rationales for the study	8
Salient contexts shaping the development.....	9
ASSESSMENT DESIGN AND DEVELOPMENT	11
Assessment design.....	11
Development overview.....	12
FORMATIVE FEASIBILITY EVALUATION	13
Evaluation structure.....	13
Interim evaluation of instrumentation	14
Interim evaluation of implementation	19
PROSPECTS FOR AHELO	23
A sound assessment model.....	24
A sustainable business model.....	25
Instrument creation and validation	26
Quality-assured operations.....	27
Engaging data and reports	29
Global positioning	29
REFERENCES	31

ACRONYMS AND ABBREVIATIONS

ACER	Australian Council for Educational Research
AHELO	Assessment of Higher Education Learning Outcomes
ATAV	AHELO Translation, Adaptation and Verification
CAE	Council for Aid to Education
CHEPS	Center for Higher Education Policy Studies
EUGENE	Academic Network European and Global Engineering Education
FCI	Faculty Context Instrument
GNE	Group of National Experts
IC	Institutional Coordinator
ICI	Institution Context Instrument
IMHE	Programme on Institutional Management in Higher Education
NC	National Centre
NIER	National Institute for Educational Policy Research
NPM	National Project Manager
OECD	Organisation for Economic Cooperation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
QAA	UK Quality Assurance Agency
SCG	Stakeholders Consultative Group
SCI	Student Context Instrument
TAG	Technical Advisory Group

INTRODUCTION

Overview of this report

1. This report presents interim evaluative insights arising from design, development and evaluation of the Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study. The report draws together materials from a wide range of sources, and from work that is currently underway. It is designed to provide high level information for the OECD IMHE Governing Board and Education Policy Committee.
2. This report draws heavily on a much longer July Interim Feasibility Report (AHELO Consortium, 2011). The July report incorporated feedback from hundreds of stakeholders, document research and conceptual design, and empirical testing to date. This report took stock of progress at that time, and made formative suggestions for the remainder of the study. This Revised Interim Feasibility Report also references numerous operational documents produced since June 2011.
3. The principal authors of this Interim Feasibility Report are Drs Hamish Coates and Sarah Richardson at the Australian Council for Educational Research (ACER). Dr Coates is the AHELO International Project Director and Dr Richardson manages the international work. ACER is leading a consortium of international agencies to conduct the AHELO Feasibility Study (the ‘AHELO Consortium’). The report has been reviewed and revised in light of feedback from OECD Secretariat, AHELO Consortium, and the TAG. It has then been submitted and discussed with the AHELO GNE in November, which led to further revisions integrated in this revised version.

Key insights to date

4. The AHELO Feasibility Study is a landmark international activity for higher education, requiring considerable creativity and innovation. Overall, as this interim report suggests, significant ground has been made in establishing feasibility. Major outcomes have been achieved—such as international agreement regarding discipline frameworks and instruments—and many challenges have been overcome. This is an important finding. Good progress has been made in little more than a year, and in developing instruments and preparing for fieldwork there is certainly no evidence to suggest that an international assessment of higher education learning outcomes is not possible. More boldly and positively, and without discounting the significant work and challenges to be confronted during implementation, there are genuine and sound indications that much of AHELO is feasible.
5. Together, insights in this interim report provide assurance that the enormous amount of energy invested in the study—by countries, the OECD Secretariat, and contractors, both independently and in collaboration—has already been transformed into significant achievements. Yet the study is only part-way through and has only yielded preliminary results. With the growing significance and evaluation of higher education around the world, the rationales for assessing higher education learning outcomes are stronger than ever.

6. Drawing on available evidence, this report offers an interim response to the eight instrumentation-focused questions posed in the study's evaluative framework—the AHELO Analysis Plan (AHELO Consortium, 2010). In summary this shows that:

1. The initial decision in Generic Skills to adapt an existing instrument for international use—the United States Collegiate Learning Assessment—posed a number of challenges with respect to language translation, cultural adaptation and other implementation issues. While results from phase two implementation are not yet available, a consensus is emerging that, contrary to initial expectations, new instrument development may have both cost and technical advantages over adaptation of existing, country-specific instruments. Currently, therefore:
 1. During instrument adaptation and initial testing it became clear that the lack of an explicit assessment framework for Generic Skills based on international input and consensus is problematic. It is not yet possible to determine whether international consensus has been reached on a Generic Skills Assessment Framework as development has only recently commenced.
 2. While work is well underway it is too soon to determine whether the Generic Skills Assessment provides valid and reliable measurement given the absence of content specification, lack of technical information on constructed response tasks and inclusion part-way through the study of multiple choice questions.
2. International development and validation of the Economics and Engineering Assessment Frameworks has affirmed that it is possible to define discipline-specific learning outcomes internationally.
3. The Economics and Engineering tests, which have been endorsed by domain and national experts and include both constructed response and multiple choice items suggest that it is feasible to develop valid discipline assessments to international standards, noting that further evaluation is required during implementation to confirm that the Economics test is pitched 'above content'.
4. A Contextual Dimension Framework and Student, Faculty and Institution Context Instruments have been developed and validated with stakeholders which, for the purposes of the AHELO Feasibility Study, reflect an international consensus about the important contexts that shape higher education learning outcomes.

7. The implementation phase of AHELO is just commencing but as with any study of this nature much design and development is incorporated into the production of instruments. Feasibility insights thus far include that:

- a. The AHELO Assessment Design sets out a vision and strategy for success which, to date, has been sustained and scaled. The only major revision, in response to feedback, has involved clarifying and augmenting the Generic Skills strand with multiple choice items and an assessment framework.
- b. Work done so far in Economics, Engineering and Contextual Dimension and Project Management strands has worked well practically and has scaled well. National management arrangements have been effective, even despite a range of political, economic and natural problems. Strategies for engaging ministries, institutions, students and stakeholders have proven effective, with much remaining to be done in the implementation phase.
- c. In terms of cross-national, cross-cultural and cross-linguistic generalisability: Evidence is available to confirm that the test instruments, including constructed responses and multiple choice items, for the

Generic Skills strand, as well as for the two discipline-strands, Economics and Engineering, have been translated, adapted and independently verified using transparent processes to international standards. The three context instruments have been translated for all languages, and adaptation and verification is complete in all countries.

- d. The support from systems and institutions has not only been high, but has outstripped predictions. An increasing number of countries have chosen to participate in AHELO as the study has progressed. Similarly, in the majority of participating countries, National Project Managers reported high levels of interest and have not reported difficulty recruiting institutions.
- e. At this stage it is too early to evaluate whether: the study has been implemented in a methodologically rigorous fashion; it has been possible to engage faculty and a random sample of student respondents in the study; the instruments have been delivered successfully; operations have been managed successfully by countries and institutions; people could be trained to score tasks in cross-linguistically and cross-culturally generalisable ways; or statistical reports of assessment results are of value to systems and institutions.

AHELO FOCUS, RATIONALES AND CONTEXTS

Overview of the AHELO Feasibility Study

8. AHELO is an international assessment of bachelor-type degree final-year students' capacity to use, apply and act on their knowledge and reasoning. The broad purpose is to design, develop and evaluate a robust approach to measuring learning outcomes in higher education in ways that are valid across cultures and languages, and across the diversity of institutional settings and missions. If successful, AHELO would be the first international assessment of higher education learning outcomes.
9. The aim of the AHELO Feasibility Study is to assess whether valid and reliable cross-linguistic, cross-cultural and cross-institutional comparisons of higher education learning outcomes are scientifically possible, and whether their implementation is feasible.
10. There are two overarching objectives: the first is to assess whether it is scientifically possible to produce valid and reliable cross-linguistic, cross-cultural and cross-institutional comparisons of higher education learning outcomes; the second is to determine whether it is feasible to implement such comparisons on an international basis. The main research questions are:
 - a. Is it scientifically possible to produce cross-linguistic, cross-cultural and cross-institutional valid comparisons of higher education learning outcomes?
 - b. Is it feasible to implement a valid cross-linguistic, cross-cultural and cross-institutional assessment of higher education learning outcomes?
11. Seventeen countries are participating in the Feasibility Study. Country participation is distributed across three strands of testing. In the instrumentation phase of the study nine countries have been involved in the Generic Skills strand, seven in the Economics strand, and eight in the Engineering strand. All countries participate in the Contextual Dimension, which incorporates student-, faculty- and institution-

level instruments. The tests and questionnaires are translated and adapted into 12 languages and cultures. International reports with a focus on the feasibility aspects of an AHELO will be prepared, along with reports for each participating higher education institution for which the focus will be on internal use of the information collected.

12. Under the oversight of the OECD Secretariat and AHELO GNE, work is conducted by the ACER-led Consortium. Over 100 people are operationally involved in the work, with up to 300 participating through various expert groups, stakeholder groups, and national management centres (National Centres). Up to 250 leaders, 10,000 faculty and 40,000 students will be involved in the testing. Globally, there are many thousands of organisations and individual stakeholders engaged.

13. Despite a difficult international funding climate the AHELO Feasibility Study has been supported by participating countries, and through contributions from: Lumina Foundation for Education (United States), Compagnia di San Paolo (Italy), Hewlett Foundation (United States), Calouste Gulbenkian Foundation (Portugal), Riksbankens Jubileumsfund (Sweden), the Spencer Foundation (United States) as well as the Higher Education Funding Council for England (HEFCE) and the Higher Education Authority of Ireland (HEA).

14. AHELO is designed to engage systems, institutions, faculty and students in evidence-based change. The assessment is based on collaboration between ministries, disciplines and technical experts, faculty and learners, institutional leaders, and a wide range of global stakeholders. With sound design and development AHELO has the potential, without loss of quality, to be scaled to very large numbers of systems, institutions and students.

Broad rationales for the study

15. The decision to conduct AHELO goes to the heart of OECD's mission. Higher education is a central contributor to the success and sustainability of the knowledge economy. Human capital development now rests fundamentally on higher education. This dynamic relationship leads policymakers and a wider public to devote considerable attention to higher education. Substantially driven by rapid expansion in student numbers, investment in higher education is significant and growing, both from public and private sources. The reliance of industry on well-prepared graduates is escalating as unskilled and semi-skilled jobs diminish in the labour market. In the era of globalisation, higher education has experienced decades of accelerating internationalisation, including mobility of students and academics. This suite of characteristics has prompted greater attention to evidence of the quality of learning and teaching, and the relevance of higher education outcomes.

16. At the same time, efforts to improve the quality of teaching and to enhance students' learning outcomes suffer from a considerable and surprising information gap. There is no reliable information which enables comparative judgments to be made about the capabilities of students in different institutions and in different countries, or about the quality of teaching. In the absence of such data on core higher education activities of learning and teaching, the standing of a higher education institution is based largely on reputation and research performance. Uni-dimensional international rankings derived from inputs or research-driven outputs distort decision-making by individuals, institutions and governments. AHELO's objective is to create a richer source of information through designing and testing measures which give due weight to teaching practices and learning outcomes.

17. This objective generated considerable interest within institutional, political and scientific circles. However, some actors harboured doubts about whether it was scientifically and operationally feasible to measure learning outcomes across institutions of very different types, and across countries with different cultures and languages. Even within national contexts there are only a few initiatives on which to build.

18. Three formative meetings with international specialists were convened in 2007 and 2008 to determine how best to address this matter. These meetings concluded that it would be desirable, in public policy terms, to assess and compare central components of higher education outcomes. They acknowledged that such assessment and comparison was theoretically possible. However, they reasoned that a feasibility study was necessary to test this proposition before undertaking any more systematic assessment. The feasibility study would have to test both the science of the assessment and the practicality of implementation.

19. In response to this expert advice, the OECD devised and is conducting the AHELO Feasibility Study which runs until late 2012. Its broad objective is to demonstrate the feasibility of measuring institutions' performance from the perspective of student learning. Specifically, the study was conceived as a means of collecting objective data on final-year learners' Generic Skills, and on Economics and Engineering learning outcomes. So far, the instrumentation phase of the Feasibility Study has involved developing and validating tests with 42 students (Generic Skills), 391 students (Economics) and 301 students (Engineering) across participating countries. In the implementation phase it is expected that around 270 institutions, 40,500 students and 10,000 teachers (hereafter 'faculty') will be involved. Context data will also be collected from students, faculty, and institutional and national representatives.

20. Concurrent with design and development, the study involves formative evaluation of the scientific and practical feasibility of a global assessment of higher education learning outcomes. This report provides an overview of study design and management. Primarily, the report takes stock of the lessons learned during the development of the feasibility study and what these mean for an eventual future, full-scale AHELO. This is an interim report: it focuses essentially on the first phase of the Feasibility Study, which was concerned with developing frameworks, instruments and processes for implementation, and conducting the qualitative validation of these instruments.

Salient contexts shaping the development

21. As captured by Coates and Richardson (2011), AHELO is being developed in an era in which there are pressures driving more performance assessment in higher education. Institutions are adapting to increasingly global and competitive operating contexts. There is a multiplicity of rationales for this, not least interest in better understanding the transparency, effectiveness, diversity, productivity and accountability of an expanding sector. Within this broad frame, 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—most particularly the interest in knowing more about highly skilled individuals and their dramatically increased movement around the world. The following analysis is more modest in intent, and concentrates on large-scale developments within higher education.

22. One of the main rationales driving AHELO is the growing need for data to support continuous quality improvement. Growing beyond an elite or mass scale, many higher education institutions have grown to the point whereby new management and monitoring systems are required for teaching and learning. Institutions are also operating in an increasingly complex external environment. Coupled with other metrics, generalisable data on assessment can assist leaders, teachers and learners understand and position their work.

23. At the same time, AHELO reflects a 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 collegial approaches to defining, assessing and monitoring academic standards are no longer sufficient on their own for yielding generalisable data on what graduates have

learned and can do (Coates, 2010). This is not surprising, for these approaches were never designed for this purpose. 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.

24. 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, from 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 (Van Vught & Westerheijden, 1993). This process has naturally led to questions about international comparison, and directly and indirectly quality assurance processes have provided foundations for a considerable amount of benchmarking. 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.

25. Of course much quality-relevant data does 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 measuring outcomes correlates, however the data collected is subjective in nature and focused on educational processes rather than learning outcomes. International assessment collaborations, such as the United Kingdom’s Medical Schools Council Assessment Alliance (MSC-AA, 2011) or the Australian Medical Assessment Collaboration (ACER, University of Queensland & Monash University, 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.

26. 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 numerous 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 conflation 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.

27. 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 devolved 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).

28. Transparency initiatives such as U-Multirank and Tuning, along with other large-scale initiatives, are important contexts 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 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 remains scarce in many systems.

29. At the same time, a growing number of institutions are seeking ways to self-position strategically and operationally within a more borderless and complex higher education environment. Such work is hindered by existing global research-focused rankings of little value to institutions which do not fit the research-focused mould. Many teaching-focused and developing institutions seek assessment instruments and comparative data to help them determine how their programs and students compare relative to international standards and performance.

30. 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 program content. Explorations of institutional and discipline-specific performance in student learning are growing from a low base. Yet greater insights into comparative learning across countries is of increasing importance given the progressive global mobility of students and graduates, and in the context of finding viable policy solutions to sustaining universal university education. Despite all 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. AHELO has the potential to fill this gap, building on fertile ground created through various projects undertaken over recent decades.

ASSESSMENT DESIGN AND DEVELOPMENT

Assessment design

31. AHELO is an innovative study and significant planning and development was used to design the study’s technical architecture. Drafted in response to the 2009 Call for Tender and then finalised through subsequent consultation, the Assessment Design (AHELO Consortium, 2010) advances the integrated approach for designing and implementing the AHELO Feasibility Study. This approach has been designed to maximise the synergies across the different strands of the AHELO Feasibility Study, streamline communications, reduce resource demands on participating countries, and generate economies of scale.

32. In summary, key elements of the design include:

- a vision and strategy for the assessment;
- need to use the Collegiate Learning Assessment in the Generic Skills strand;
- development of frameworks and instruments for the Economics, Engineering and Contextual Dimension strands;
- leadership and management arrangements, and evaluation strategies;
- methods for translation, adaptation and verification;
- delivery platforms specifications;
- strategies for population specification and sampling management;
- fieldwork procedures and analytic methods;
- approach to implementing the Collegiate Learning Assessment; and
- data products and written reports.

Development overview

33. To frame the evaluative remarks that follow it is useful to sketch of the design and management of the AHELO Feasibility Study. The approach to managing each of the testing strands is described briefly, as is the approach to international management and providing national training and support. The AHELO Assessment Design (AHELO Consortium, 2010) gives more detail.

34. On the recommendation of the 2008 expert groups, a decision was made by the AHELO GNE in 2009 to use the United States Collegiate Learning Assessment as the AHELO Generic Skills Assessment. Work has been managed by the Council for Aid to Education (CAE). A Generic Skills Expert Group was not formed to review framework and test materials, although general oversight has been provided by the AHELO TAG since July 2010. CAE contracted directly with OECD for instrumentation and pre-implementation work. ACER will contract with CAE for implementation.

35. Development of AHELO's Economics assessment was undertaken by the Educational Testing Service (ETS) as part of the AHELO Consortium selected following an international call for tender. Framework and test instrument development were overseen by an international Economics Expert Group.

36. Development of the framework and assessment materials for the Engineering Assessment was undertaken by the AHELO Consortium—specifically ACER, Japan's National Institute for Educational Policy Research (NIER), and the University of Florence. Several international consultants contributed to the development. Framework and test instrument development were overseen by an international Engineering Expert Group.

37. Development of the framework and instruments for the Contextual Dimension was undertaken by the AHELO Consortium—specifically the Centre for Higher Education Policy Studies (CHEPS) at the University of Twente, ACER, and the Indiana University Centre for Postsecondary Research (CPR). The AHELO TAG provided oversight for this work.

38. An international consortium of agencies provides international management for the AHELO Feasibility Study under the overall directorship of Associate Professor Hamish Coates (ACER). ACER has established a transparent and efficient centre to lead international operations and support National Centres. ACER takes overall responsibility for study design and development, analysis and evaluation, coordination of expert groups and national managers, and liaison with the OECD Secretariat, TAG and AHELO GNE. IEA DPC is assisting with fieldwork quality control.

39. Adaptation and translation of the constructed response tasks used in the Generic Skills strand was conducted by participating countries in collaboration with CAE, while the verification of the translation was conducted by Comms Multilingual. Adaptation and translation of the Economics and Engineering tests was conducted by participating countries in collaboration with cApStAn, which also was responsible for the verification process. The adaptation, translation and verification of the Generic Skills multiple choice questions and of the three Contextual Dimension instruments were conducted by cApStAn.

40. All testing in 2012 will be conducted via computer-based platforms, either via the internet or USB. The constructed response tasks used in the Generic Skills strand will be delivered using a platform managed by CAE. Overall management of the testing session and delivery of all other materials (Generic Skills multiple choice questions, Economics and Engineering tests, and three Contextual Dimension instruments) will be via the SoNET testing system.

41. Statistics Canada has led design and will lead implementation of population specification and sampling work, with support from ACER. ACER will manage testing procedures and operations, psychometric and statistical analysis, international, and technical and institutional reporting and evaluation.

42. At the time of writing: five of nine Generic Skills constructed response tasks have been uploaded to the test system ready to use; all nine Generic Skills multiple choice questions are ready for testing; all nine Economics assessments are ready for testing; all but the Abu Dhabi and Russian Engineering assessments are ready for testing; and all but the Russian context instruments are ready for testing. In addition to contributions to instrument development and validation, countries also provided feedback on a large number of manuals and documents which the AHELO Consortium have developed to support and guide their national activities. Detailed plans have been developed for the first international training of NPMs and nominated lead scorers held in late November 2011. This first training encompassed scoring, sampling, the use of the AHELO test system and national management. A second training of NPMs and lead scorers is scheduled in mid-March 2012. The intent of this training is to provide NPMs with a thorough overview of the procedures to implement the AHELO feasibility study, as well as guidance on how to train Institutional Coordinators and Test Administrators. As for the training for lead scorers, the intent is to provide them with detailed instructions on how to score constructed response, train scorers and monitor scoring in their countries.

FORMATIVE FEASIBILITY EVALUATION

Evaluation structure

43. AHELO's viability hinges on proof that it is feasible to develop instruments that measure learning outcomes on an international scale, and that it is feasible to implement these assessments using methods that are scalable, efficient and secure.

44. In broad terms, the main determinant of the success of the study is proof that the assessments developed can be applied in diverse settings with appropriate adaptations and yet provide valid, reliable and bias-free measures of student learning outcomes. Scientific feasibility depends on whether it is possible to develop assessments that are perceived as valid in diverse institutional, cultural and linguistic contexts. The study also needs to gauge whether test items perform as expected and test results meet pre-defined psychometric standards of validity and reliability. Practical feasibility is being evaluated on the

basis of how effective the strategies to secure institutional and student cooperation have been, and to what extent the implementation of the assessments has brought benefits to participating institutions and demonstrated its value for teaching improvement.

45. The AHELO Analysis Plan, drafted by the Consortium in 2010 and ratified by the GNE, organises these considerations and provides the study's evaluative architecture. The Analysis Plan specifies 19 research questions, outlines means for evaluating the cross-linguistic, cross-cultural and cross-institutional validity of the instruments being used in the various assessments and context surveys, and charts the sources of evidence to assess the various dimensions of the study.

46. AHELO is a landmark international initiative, and although only a feasibility study expectations are high. This raises the bar in terms of the criteria used to judge success and the standard of evidence required. In important respects, the responsibility is on the study and its stakeholders to prove feasibility. It is particularly important in the feasibility evaluation, therefore, to be rigorous and candid about the work, materials and outcomes.

Interim evaluation of instrumentation

47. Question 1: Was the Generic Skills Assessment Framework reflective of an international consensus about the areas that are important to assess?

- a. Following a decision by OECD and the AHELO GNE, the 2009 Call for Tender specified that the CAE's Collegiate Learning Assessment was to be trailed in the Generic Skills strand. Hence the Assessment Design and initial work plans did not specify development of a Generic Skills Assessment Framework. Nonetheless, in the course of the study, numerous stakeholders sought clarification on what the test was measuring, and requested production of a framework in line with standard international practice.
- b. Accordingly, the July Interim Feasibility Report (AHELO Consortium, 2011) concluded that a Generic Skills Assessment Framework which is international in scope should be developed. Development of this document commenced in August 2011 and is in progress. As per standard practice this framework will specify the (generic) domain to be assessed, the balance given to elements in the domain, and evidence of the degree to which the framework accounts for institutional, language and cultural differences. This work is being led by ACER in consultation with CAE, and with oversight by the AHELO TAG.
- c. A consultation plan is being developed to progress this framework. Consultation so far has been limited and it is too soon to determine whether an international consensus has been or can be achieved. Based on prior feedback and review of the literature from different cultures it seems likely that it will be possible to find commonality among international conceptualisations of 'generic skills', and how these might be assessed.

48. Question 2: Was the instrumentation developed on the basis of the Generic Skills Assessment Framework faithful to the spirit and intent of the framework?

- a. Evaluation of whether the Collegiate Learning Assessment provides valid and reliable measurement of a Generic Skills Assessment Framework has thus far been limited due to the absence of a framework and of data collected during the first phase of the study on which such assessment might be made. Hence it is too soon to draw an interim conclusion.
- b. Small-scale focus groups have been conducted with 42 students internationally—between 6 to 14 students in each country at between 2 and 5 institutions. The focus groups have indicated that some

students like the look of the tasks, however stakeholders have expressed concerns about the adaptability of pre-existing content into other cultures. Researchers, stakeholders and countries participating in the Generic Skills strand have noted that the contextual relevance of assessment tasks would be increased if they were developed internationally rather than adapted from a single country context.

- c. Detailed scoring rubrics, exemplars and processes will be required to affirm the capacity of the Collegiate Learning Assessment—an assessment requiring complex verbal responses—to be generalised across diverse linguistic, cultural, disciplinary and epistemological contexts. Scoring was not conducted as part of the phase one focus groups and it is too soon to determine whether such scoring is feasible scientifically or practically. Rubrics are currently being reduced—from four to three data elements—to lessen emphasis on writing. The feasibility of scoring constructed response tasks will be possible to assess in 2012 following conclusion of implementation.
- d. To provide an empirical anchor for the generic skills assessment—for scoring, quality control, international equation, bias analyses, and scaling—the July Interim Feasibility Report recommended the incorporation of objectively scored multiple choice questions into the assessment. To facilitate this given time constraint ACER has supplied existing pre-validated multiple choice questions to OECD for the Feasibility Study. NPMs were consulted on and approved these questions in early October. The questions have been adapted culturally, translated, and independently verified. Given modifications necessary since initial design, it is too soon to confirm the properties of the Generic Skills test.

49. Question 3: Was the provisional Economics Assessment Framework reflective of an international consensus about the important learning outcomes in Economics?

- a. Developing a valid provisional Economics Assessment Framework is important for the AHELO Feasibility Study as it suggests that it is possible to measure discipline-specific skills in a social science domain. Clearly, this has direct implications for the assessment of Economics, and also more broadly for the potential development of assessments in other social sciences.
- b. The work has been led by the Educational Testing Service (ETS) and overseen by an independent international Economics Expert Group. Experts and NPMs in all participating countries and a number of other stakeholders have provided input into the framework. Drawing significantly from the AHELO-Tuning document (Tuning Association, 2009a) and the United Kingdom QAA Subject Benchmark Statement for Economics (2007) work begun in July 2010 and progressed until April 2011 when the Economics Assessment Framework (ETS, 2011) was delivered to OECD.
- c. The Economics Assessment Framework defines the domain to be tested and specifies the expected learning outcomes for students in the target population. It provides an overview of contemporary perspectives on the assessment of Economics learning outcomes. It defines five broad learning outcomes, and particularises discrete competencies underpinning each of these. The document provides an overview of the instrumentation required to measure the competencies, with discussion of issues such as time, language level, item type, scoring, assessment delivery and administration, and reporting.
- d. Production of the framework suggests that the provisional Economics Assessment Framework is indeed reflective of an international consensus about the important learning outcomes in Economics. Framework production also helped identify several challenges of broader relevance to developing outcome specifications in higher education such as the weighting of curriculum contents, how curricular contents are distributed across institutions and institutional types, the level of generality/specificity at which testing should be focused, and the separability/chunking of contents.

50. Question 4: Was the instrumentation developed on the basis of the Economics Assessment Framework faithful to the spirit and intent of the framework?

- a. Initial validation of the AHELO Economics Assessment shows it has the potential to operationalise the Economics Assessment Framework well and provide valid, reliable and efficient measurement of target constructs.
- b. Content validity of AHELO's Economics Assessment flows from its derivation in the Economics Assessment Framework. The iterative and multimethod development process is described in detail in the Economics Assessment Development Report (ETS & ACER, 2011), delivered to OECD in June 2011. Validity also flows, given that items for the Economics Assessment were sourced from a psychometrically validated bank of materials used by ETS for the GRE Subject Test in Economics. The creation of bespoke constructed-response tasks for AHELO helped ensure that the Economics Assessment provided the coverage required.
- c. To a certain extent construct validity has been designed into the test by drawing on materials from a validated item bank. Given the adaptation and internationalisation of these materials it is necessary to look carefully at construct validity. At this stage 391 students in participating countries have taken the Economics Assessment, providing a small amount of international data for psychometric analysis and review. While preliminary, this work has affirmed the construct validity and reliability of the assessment at an international level.
- d. Qualitative feedback—from faculty, national managers and experts—has affirmed that different content and test characteristics are seen to have different levels of relevance by different stakeholders. For instance, in some instances people working in institutions or programs focusing on a more applied economics curriculum may seek slightly different forms of assessment than people working with more academic programs. In a full-scale study going beyond the production of a provisional framework a more expansive matrix design could be deployed to embrace such diversity.
- e. As anticipated from the test design, analysis of scored test response data collected during focus groups with 391 students indicates that the Economics Assessment has appropriate levels of reliability and construct validity at the international level. After completing the test students were asked to provide feedback on the material via a brief questionnaire. This feedback provided test developers on the extent to which the test materials: made students apply knowledge and skills in real-world ways; required students to apply capability gained in their program; challenged students to think; were easy to understand; were relevant to future professional practice; stimulated students' interest; had good linkage with other questions; were relevant to the program of study; assessed an appropriate range of knowledge and skills; were relevant to the content being assessed; were sufficient for them to complete this task; and covered topics relevant to the program. Quantitative and qualitative student insights, along with feedback from faculty involved in administering the focus groups, were synthesised with expert feedback to validate the tasks.
- f. The Economics Assessment scoring rubrics were tested by national managers as part of the focus group work. This helped identify areas in need of clarification and improvement. Changes are being made along with finalisation of scoring resources in October 2011, which will form the basis of international training in late November 2011. The feasibility of scoring constructed response tasks will not be possible to assess fully until mid 2012 after conclusion of the implementation phase.

51. Question 5: Was the provisional Engineering Assessment Framework reflective of an international consensus about the important learning outcomes in Engineering?

- a. The field of Engineering was selected for the AHELO Feasibility Study to test whether it was possible to produce an assessment in a scientific professional discipline. Significant development and validation work was conducted between July 2010 and April 2011, building on the AHELO-Tuning document (Tuning Association, 2009b), the Tertiary Engineering Capability Assessment Concept Design (Coates & Radloff, 2008), several symposium in Europe and Asia, and a provisional Engineering Assessment Framework (ACER, NIER & University of Florence, 2011), reflective of an international consensus about important learning outcomes, was delivered to OECD in May.
 - b. A useful index of the Engineering Assessment Framework's feasibility is its reception by academic and professional communities. It is important that consensus was reached by the development team which comprises experts from Australia (ACER and University of Melbourne), Japan (NIER), and several European countries (via the University of Florence EUGENE network). The team's international composition means that many complexities and difficulties were resolved during framework conceptualisation and development. The Engineering Expert Group is highly influential given the group's depth and breadth of experience. Feedback from broader stakeholders consulted throughout the development has also been positive including, for instance, from the European Society for Engineering Education (SEFI), ABET, Japanese Institution of Professional Engineers, Japan Society of Civil Engineers, Engineers Australia, Australasian Association for Engineering Education, International Federation of Engineering Education Societies, and the many organisations and associations involved with the EUGENE network.
 - c. Framework production involved review of research and accreditation documentation, consultation with educators and industry representatives, and drafting and review of framework documentation with reference to substantive, technical and practical considerations. Mapping of items against the definitional structures in the framework provides a means of validating its focus and scope.
 - d. An interesting epistemological issue arose during the development of the Engineering Assessment Framework, which carries broader implications for instrument design and development. The Engineering Expert Group invested considerable discussion in the relative merits of testing discrete competencies using multiple choice items, compared with taking more holistic snapshots of the capacity respondents have to analyse, synthesise and apply their knowledge in more extended real-world problem settings. Ultimately, a decision was made to strike a balance between these epistemological perspectives, and to deploy an item mix that provided coverage of each.
52. Question 6: Was the instrumentation developed on the basis of the Engineering Assessment Framework faithful to the spirit and intent of the framework?
- a. Initial validation of the AHELO Engineering Assessment shows it has the potential to operationalise the Engineering Assessment Framework well and provide valid, reliable and efficient measurement of target constructs.
 - b. The content validity of AHELO's Engineering test flows from its derivation in the Engineering Assessment Framework. The development process is described in detail in the Engineering Assessment Development Report (ACER, NIER & University of Florence, 2011). Content validity has also been addressed through the development of AHELO-specific, problem-based authentic constructed response modules and multiple choice questions. The tasks were crafted to map onto relevant facets of the domain. Items drawn from the licensing examinations developed by the Institution of Professional Engineers Japan and the Japan Society of Civil Engineers had already been tested against the broad conceptual domain. Further review and selection of these by the Engineering Expert Group in their face-to-face October 2010 meeting in Singapore provided an opportunity to validate the international relevance of these items.

- c. Focus group testing has been conducted in participating countries, involving 301 students internationally at around 55 institutions. Scored test data captured from the qualitative testing yielded psychometric evidence of the Engineering Assessment's construct validity and reliability. As well, student feedback provided insights on the extent to which the test materials: made students apply knowledge and skills in real-world ways; required students to apply capability gained in their program; challenged students to think; were easy to understand; were relevant to future professional practice; stimulated students' interest; had good linkage with other questions; were relevant to the program of study; assessed an appropriate range of knowledge and skills; were relevant to the content being assessed; were sufficient for them to complete the tasks; and covered topics relevant to the program. Quantitative and qualitative student insights, along with feedback from faculty involved in administering the focus groups, were synthesised with expert feedback to further validate the tasks.
- d. The Engineering Assessment scoring rubrics have been tested as part of the focus group work, and have been found to function consistently across cultural and linguistic settings. Even so, the empirical work has shown that by themselves constructed response modules are not sufficient to provide valid and reliable measurement of students' ability even in a well-defined discipline. Constructed response tasks do not yield enough data to generate reliable estimates or map underpinning constructs, and require subjective scoring which is difficult to calibrate internationally. Revisions to scoring resources and procedures are being made alongside revisions of the final source version of the instrument. The feasibility of scoring constructed response tasks will not be possible to assess fully until mid 2012 after implementation.

53. Question 7: Was the provisional Contextual Dimension Assessment Framework reflective of an international consensus about the important contexts that shape higher education learning outcomes?

- a. The development of the Contextual Dimension for the AHELO Feasibility Study is an important facet of the study's architecture. In broad terms, contextual information is required to understand and manage quality and implementation, to understand important differences across settings that are needed to adjust or interpret learning results, and to point to institutional policies, practices and other factors related to student outcomes.
- b. The Contextual Dimension Assessment Framework (ACER, CHEPS & Indiana University, 2011) is based on foundation work undertaken in 2008 and 2009 (Tuning Association, 2009a, 2009b). The framework is informed by the processes and practices adopted in contextual dimensions used around the world. These include the cross-national assessments of PISA, TIMSS and PIAAC; system level data sets such as Education at a Glance, IPEDS and HEIMS; cross-institutional collections such as NSSE, JCIRP and REFLEX; and national and regional classifications such as U-Map and U-Multirank.
- c. Development of the Contextual Dimension Assessment Framework was undertaken through research and consultation, and by seeking the expert opinion of a range of groups and individuals from across the world (notably GNE, TAG, NPMs, SCG, Economics Expert Group, Engineering Expert Group, and the AHELO Consortium). The framework, reflective of an international consensus about important learning outcomes, was delivered to OECD in August 2011.

54. Question 8: Was the instrumentation developed on the basis of the Contextual Dimension Assessment Framework faithful to the spirit and intent of the framework?

- a. Widespread consultation on the AHELO Contextual Dimension instruments suggests it has the potential to operationalise the Contextual Dimension Assessment Framework well and provide valid, reliable and efficient measurement of target constructs.

- b. Three survey instruments have been developed in this process to underpin the Framework: Student Context Instrument (SCI), Faculty Context Instrument (FCI) and Institution Context Instrument (ICI). In addition, a range of indicators have been specified for collection at the national level to provide additional context.
- c. Each of the three instruments has been reviewed by stakeholders. In brief, this process has included: a two day face-to-face workshop with the AHELO TAG; review of both the framework and indicator mapping by the AHELO GNE and SCG; and NPM review of the framework and draft questionnaire items.
- d. In addition to the above, qualitative focus groups have been organised in Australia, Japan, Netherlands and United States to gather information from students and staff. Feedback from these consultations along with subsequent revision has confirmed that the instrumentation has been faithful to the intent of the framework.
- e. Technically, the instruments have been mapped against relevant standards, data specifications, sampling requirements, test system requirements, fieldwork coordination requirements, coding needs, psychometric and statistical needs, and reporting requirements.
- f. Through verification of translations and adaptation of nationally-specific terminology, individual countries have further enhanced the utility of the Contextual Dimension instruments.

Interim evaluation of implementation

55. Question 9: Was the AHELO Assessment Design valid and feasible?

- a. The AHELO Assessment Design was written in 2009 to advance the AHELO Consortium's integrated approach for designing and implementing the AHELO Feasibility Study. It has been updated in minor way as new countries have joined the study. Presented in a modular format, the Assessment Design provided an overview from the perspective of the start of the project, acknowledging that discoveries would be made during the project which had not been documented in the plan.
- b. Evaluating the validity and feasibility of the Assessment Design requires examination of a number of factors such as its scope, the extent to which it has undergone changes, and the suitability of its design to provide a basis for full-scale extension. Interim insights can be reported, but it is not possible to conclude such evaluation until the conclusion of the study.
- c. The design was influenced by a vision of success which, so far, has largely been achieved. Within the field of international comparative assessment in higher education, innovative methodologies and new technical standards have been established, a broad range of stakeholders around the world have been engaged and the idea of AHELO—that assessing higher education learning outcomes is an essential checkpoint in the educational process—has found fertile ground internationally. The growing participation of systems and institutions in the study is evidence for this.
- d. At the same time, the Feasibility Study design has provided a basis for critique and development. Given technical review and stakeholder feedback it has been necessary to revise the design of the Generic Skills strand to include a content specification (Assessment Framework) and multiple choice questions. Much emphasis has been placed on incorporating generic skills with disciplinary assessments, potentially via the assessment of 'disciplinary generic skills' (as per the Engineering Assessment) or through the combination of generic and discipline-based assessments (as per the strategy of Colombia, Egypt, Mexico and the Slovak Republic which are involved in several strands). Many facets of the assessment design—for instance, the scoring of complex constructed response

tasks, use of complex sampling techniques, the use of computer-based testing—are yet to be deployed or evaluated.

- e. Evaluating and projecting the future of an innovative international assessment of higher education learning outcomes requires the involvement of an independent and highly regarded Technical Advisory Group separate from policy and operational contexts that oversees and provides guidance on all aspects and all strands of AHELO. Thus far, the significant contribution of a group of eminent higher education scholars has been evident in terms of their willingness to make strategic decisions that maximise the value of an international assessment of higher education learning outcomes.

56. Question 11: Was the study feasible from a practical perspective?

- a. AHELO will only be successful if it is practical. Practical feasibility encompasses a range of measures in relation to cost, timing, and communication processes, the use of technology, and acceptance by institutions, faculty and students. All of these must be scalable if AHELO is to expand to a greater number of institutions, disciplines or countries. At this stage it is possible to make preliminary remarks based on instrumentation and pre-implementation.
- b. Cost/benefit analysis is a central facet of any feasibility study. Subsequent review reports should provide detailed discussion of financial decisions and tradeoffs made during the course of the Feasibility Study. Only limited analysis of this matter has been conducted to date. In its April 2011 meeting the TAG (2011) noted that the funding allocation for the Generic Skills strand (excluding multiple choice questions and framework development) was slightly more than for the Economics, Engineering, Contextual Dimension and Project Management strands combined. The concern of the TAG arose principally from the value for money achieved in each strand. In the Generic Skills strand, development has involved two pre-existing constructed responses tasks without a framework. In the Economics, Engineering and Contextual Dimension strands, development has involved the creation of two internationally validated frameworks, multiple purpose-developed testing instruments, and three context instruments. In the Project Management strand, activities have included overall international project leadership, management of three expert groups, and project evaluation and reporting. The approaches used in each strand, and associated costs, have significant and direct implications for feasibility and AHELO's future.
- c. National operations are another crucial facet of practicality. Countries have made significant investments in setting up national infrastructure for AHELO, both in contracting NPMs and in establishing offices, attending international conference calls and meetings, and sourcing people to undertake translation, adaptation and verification. NPMs have coordinated a number of general meetings with institutions and other stakeholders, including policy level discussions. They have also coordinated scholarly publications and consultations. While it has been necessary in many countries to develop and overcome several practical challenges NPMs in participating countries have been efficient in their conduct of AHELO activities. They have cooperated well with international teams to undertake all the necessary tasks related to their roles. Working from feedback from NPMs in several countries, ICs involved thus far appear to have been equally diligent and efficient. The provision of training to ICs by NPMs, and the availability of detailed documentation on their roles from the AHELO Consortium, has undoubtedly facilitated the engagement and efficiency of ICs. Overall, countries have made considerable investments in the AHELO Feasibility Study. NPMs and ICs have been effective and efficient and have made their best efforts to adhere to timelines even given political revolutions (Egypt) and natural disasters (Japan).
- d. A large number of diverse stakeholders are involved in the AHELO Feasibility Study. To achieve the smooth flow of work, a clear, consistent and transparent approach to communication has been

essential. This outcome has been achieved through effective use of information technology to communicate and exchange materials. Country activities have been facilitated through the availability of a large number of documents to support national endeavours, daily support, and frequent customised training. Face-to-face meetings of NPMs have been scheduled in October 2010, and March and November 2011.

- e. For the instrumentation work, feedback was sought from around 100 institutions engaged in the discipline strands and a dozen institutions in the Generic Skills strand. It is not yet possible to report in a comprehensive way on feedback from institutions about the AHELO Feasibility Study as systematic input will not be sought until 2012. Feedback from NPMs in all participating countries indicates that institutions express a great deal of interest in the study and are enthusiastic to become involved.

57. Question 12: Was the study successfully generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally?

- a. Clearly, an international study such as AHELO can only be considered valid to the extent to which resources and processes have been generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally. Within the framework of a feasibility study, one has to demonstrate that the study design includes processes that provide for adequate capacity building within national teams and for quality control to ensure that equivalence standards are met.
- b. Higher education is an increasingly borderless activity but most institutions and systems still operate within a national context. Although AHELO is not designed like most international education assessments to yield national estimates, it remains critically important to determine the extent to which nations have engaged with the study. Such engagement is a difficult and complex matter to appraise, but much can be read from the scope and scale of ministerial involvement, and from key decisions that have been made. The scope of country participation in the AHELO Feasibility Study has increased since its inception until it was capped to contain the scope and nature of the Feasibility Study. As well, recent entrants into the study have tended to take up two and mostly three testing strands, rather than participate in only a single strand. As well as assisting with participation the 'national' (or more specifically 'systemic') context is important as it is the frame within which instrument development, cultural adaptation and linguistic translation are framed. Sound outcomes from phase one testify that nations/systems engaged in instrument development, including the procurement of expert input, submission of items, and relativisation of materials.
- c. As institutions are both the main units of analysis and the reporting level in AHELO, it is essential that the study be generalisable in this way. It is imperative that AHELO's assessments reflect and encourage institutional diversity rather than compress or even suppress it. Data developed for monitoring and continuous improvement must inspire and enable institutions to innovate within their own strategic and operating contexts. At this stage of the AHELO Feasibility study it is not possible to fully determine whether assessment resources or processes are able to be generalised between institutions.
- d. Establishing cross-cultural and cross-linguistic generalisability was sought by translation, adaptation and verification processes. Two different approaches to translation and adaptation were used: a process led by CAE for the Generic Skills constructed response tasks; and a process led by cApStAn for the Generic Skills multiple choice questions, Contextual Dimension questionnaires, and Economics and Engineering test materials.
- e. In October 2011 translated source versions of the Generic Skills constructed response tasks have been produced for Finland, Mexico and Norway, and evidence is being compiled on the nature and

effectiveness of this work. Adaptation and translation has yet to begin or is commencing in Colombia, Egypt and the Slovak Republic, while adaptation and verification has yet to be conducted with the United States. Follow-up work is underway with Korea and Kuwait to resolve problems identified in earlier work. Due to a change in CAE personnel the methodology and procedures used for double translation and reconciliation are being clarified. Independent verification was not conducted prior to focus groups, but subsequent to focus groups CAE provided all translations to a commercial language service provider for re-translation and *post hoc* review. The firm provided few details about the process and it is unclear what specific adaptation and translation methodology and procedures were adopted, and whether these were consistent with the CAE assessment design advanced in early 2010. Countries reported concerns about the overly-direct way in which the translations had been conducted. No worksheets were provided showing a detailed breakdown of the location of errors in the assessment. Work continues to prepare the constructed response tasks for deployment in 2012, and to document processes and work undertaken.

- f. With the exception of Canada (Engineering) which has recently joined the study and Egypt (Economics and Engineering) in which there are known delays, detailed transaction-level evidence is available to confirm that the Economics and Engineering tests have been translated, adapted and independently verified using transparent processes to international standards exemplified by OECD PISA, OECD PIAAC and several IEA studies. This includes consultation, adaptation, double translation, reconciliation, independent verification and source version confirmation in Australia, Belgium, Colombia, Italy, Japan, Mexico, Netherlands, Russia and Slovakia. Phase one source versions of the instruments were deployed to between 20 and 150 students at a range of institutions in each national context. Final revisions are being made to instrumentation within the online platform in preparation for 2012 implementation.
 - g. With regards to the Generic Skills multiple choice questions, as at end October 2011: NPMs had approved the source version of items and translation was complete for all participating countries. Once translation is complete, countries will review and adapt questions, and the revised translations will then be verified.
 - h. With regards to items from the three Contextual Dimension instruments, as at end October 2011: NPMs had reviewed the Contextual Dimension Framework; NPMs had provided country-specific items; NPMs had been consulted on international items; translation was complete in all countries; adaptation was completed in Australia, Belgium, Canada, Colombia, Finland, Italy, Japan, Korea, Kuwait, Mexico, Norway and Slovak Republic and near completion in others; and verification was complete in Australia, Belgium, Canada, Colombia, Finland, Italy, Kuwait, Mexico and Slovak Republic, and underway in Japan, Korea, Norway and United States.
 - i. From a general feasibility perspective based on quality, delivery, transparency and cost there is much to recommend the approach used by cApStAn. While it required time for some countries to establish national teams, national processes were effective once these were trained. In light of the experiences with Economics and Engineering materials revised workflows were devised for the Generic Skills multiple choice questions and Contextual Dimension instruments which reduced workload and increased productivity for the same quality outcomes and marginal increase in cost.
58. Question 13: Was it possible to engage systems and institutions in the study?
- a. Engaging national higher education systems and institutions in the AHELO Feasibility Study is vital to its success. Any indications that countries and institutions were not interested in participating would indicate that the intent of the study was regarded negatively by the international higher education community and would render its continuance meaningless.

- b. The inverse, however, appears to be the case. The support from systems and institutions has not only been high, but has outstripped predictions on many measures. The enthusiasm and engagement of higher education systems in countries around the world, and from institutions within those systems, indicates the great interest with which the AHELO Feasibility Study is viewed and the broad acceptance of the significance it has for higher education internationally.
- c. An increasing number of systems have chosen to participate in AHELO as the study has progressed. The total number of strand replications has risen from seven in 2008 to 23 in 2011. This alone indicates that there has been great success in engaging governments. A number of other countries have also expressed interest in participating or observing the study.
- d. Similarly, NPMs in all countries taking part in the Economics and Engineering strands reported high levels of interest among institutions and did not have difficulty in recruitment. Indeed, the level of interest from institutions was such that in both Russia and Italy, significantly more than the required number of institutions was recruited—totals of 17 and 23 respectively.

59. At this stage it is too early to respond to the following questions specified in the AHELO Analysis Plan:

- a. Question 10: Was the study implemented in a methodologically rigorous fashion?
- b. Question 14: Was it possible to engage faculty respondents in the study?
- c. Question 15: Was it possible to engage a random sample of student respondents in the study?
- d. Question 16: Was the instrument delivered successfully?
- e. Question 17: Were survey operations managed successfully by countries and institutions?
- f. Question 18: Was it possible to train people in different countries to score tasks in cross-linguistically and cross-culturally generalisable ways?
- g. Question 19: Were statistical reports of assessment results of value to systems and institutions?

PROSPECTS FOR AHELO

60. The design, development and evaluative work conducted thus far provides insights that can be used to explore a high-level, illustrative and future-oriented perspective of a feasible international assessment of higher education learning outcomes. This is not a formal nor comprehensive picture or proposition. Rather, it is a suggestive examination of key parameters and possibilities. Hence the design offers a suggestive vision that contextualises ongoing developmental and evaluative work. It is not intended to be prescriptive.

A sound assessment model

61. An essential component of AHELO is that a large number of stakeholders share a growing interest in collaborating to better understand and improve the outcomes from higher education. Systems and institutions engage with AHELO to yield collective insights into education. Captured from learners, these insights can be aggregated into relevant reports for all levels of policy and practice. Policy, practitioner and research communities can use such insights for monitoring quality and driving change.

62. Producing valid and reliable assessment resources and processes is central to AHELO. It is essential that this include consultative and technically rigorous production of assessment frameworks and instruments, operationalisation of assessment materials, deployment using quality assured and efficient implementation methods, and production of informative data products and reports.

63. Using rigorous assessment materials and practices to generate valid and reliable insights into student learning outcomes is the goal of AHELO. In addition to assessment results these insights come from reports, discussions and meetings. Insights can be used for both continuous quality improvement, and for external monitoring activities. Data on learning outcomes is new in many higher education contexts, and national and institutional capacity is being developed to ensure that results are interpreted in ways most likely to prompt changes in practice. Linking evidence with demonstrable change is one of the most important and challenging tasks for AHELO, but is essential for ensuring the efficacy and perceived utility of the project.

64. The highest standards are expected for AHELO, and there is strong global interest in AHELO's approach and outcomes. Over the last 50 years, technical standards have been developed through numerous international education surveys, and it is essential that AHELO meets these standards. To that end, AHELO must build on technical approaches tested globally over many decades in international studies—mostly with schools, and adapt and position these within the unique operating context of higher education. Robust and efficient international procedures must structure and support national contextualisation and implementation.

65. Confidentiality and security are intrinsic to high stakes testing, but experience in the AHELO Feasibility Study has affirmed that transparency is also vital. Higher education gives great weight to scholarly peer review of material and methods (for instance: task properties, translation, sampling, psychometrics and scoring), and outcomes (for instance: the validity and reliability of results, and the usefulness of reports). The need for initiatives that are designed to increase the transparency of policy and practice to themselves be transparent goes to the need for consultative governance arrangements, the application of standard technical procedures, effective project leadership, documentation of operational work, oversight of key areas by expert advisors, ongoing training and support, and the application of strict financial controls.

66. Clear communication management is essential to any effective international assessment. The success of international activities such as AHELO rests largely on the effectiveness of structures and processes for exchanging ideas and resources. Such communication must sustain training, advising and support, reporting, and daily operational work. Face-to-face meetings are invaluable, but delivery of a study such as AHELO would not be possible without effective use of collaborative information technologies.

67. Achievements to date have demonstrated that international management systems can be established and maintained to facilitate communication and transparency within a higher education operating context. At the same time, a systematic approach has ensured that access to secure materials can

be closely controlled. The success of AHELO rests on these systems and their further expansion and development into the future.

A sustainable business model

68. The AHELO Feasibility Study has operated within a context of significant financial constraints. Insufficient funds were available to enable the study to proceed as originally conceived and designed, leading to significant scope reductions. The commencement of work on different strands had to be phased according to the availability of funds. These constraints have shaped the way in which AHELO has progressed, and have forced those involved in project governance, leadership and management to give careful and creative consideration to viable governance and management structures for AHELO. Paradoxically, while the funding situation made conducting the study much more difficult, uncertain financial contexts increased the need for performance data in many national contexts.

69. At the broadest level, the OECD is well placed to lead and implement AHELO given its institutional framework for co-operative and international comparative work, as well as its credibility and demonstrated expertise in developing large-scale international assessments. Within the OECD, IMHE provides a platform for the engagement of higher education institutions in AHELO alongside governments, in a way which will ensure that the approaches adopted take account of institutional needs and concerns. IMHE is open to all recognised higher education institutions in OECD countries, as well as associations of higher education institutions and government representatives. In addition to Secretariat functions, OECD and IMHE support the following governing bodies: IMHE Governing Board, Education Policy Committee, and GNE.

70. Specialist agencies provide operational and technical management of AHELO. To ensure quality, productivity, transparency and the efficient use of resources it is helpful that a suite of international agencies from various regions are involved, with one lead agency taking responsibility for delivery of the work. Specialist oversight and input comes from expert groups, including the AHELO TAG, and internationally convened groups for each substantive domain to be assessed. Studies like AHELO involve consultation with many stakeholders, and formal input is expressed via the SCG.

71. International design and coordination is just one element of AHELO. It is equally imperative that participating countries and institutions are able to provide appropriate infrastructure for managing the national activities of such assessments. Much has been learned about the different possible kinds of national and institutional arrangements. Participating countries engage primarily with AHELO in the context of a strategic oversight offered by OECD's Education Policy Committee via the GNE, and operationally via an AHELO National Centre. The National Centre is led by a NPM. Each participating higher education institution is represented by an IC who reports directly to the country's NPM.

72. Potentially, there are several models for engaging systems and institutions in AHELO, including that OECD engage with:

- a. ministries, then ministries recruit institutions;
- b. both ministries and institutions; or
- c. ministries, institutions and students.

73. The first approach—involving no direct contact between OECD and institutions—has been used for the AHELO Feasibility Study. This follows the natural architecture of OECD assessments, but has left unmet demand from institutions in non-participating countries which have wanted to engage on their own terms outside national arrangements. This demand is evident, and has accelerated as the Feasibility Study

has progressed, with serious interest being expressed by institutions in non-participating countries. Given such demand, the autonomy and transnational positioning of many higher education institutions, the way in which many international assessment activities operate, the institutional focus of reporting data on higher education in many countries it would be fruitful to explore whether arrangements could be made to assist institutions engage directly with AHELO, and what form these arrangements might take. The third approach—further expanding to involve students directly—is more ambitious. Yet given increasing mobility and devolution of quality deliberations to students, it may be necessary to consider how they could be more actively engaged in the assessment process.

Instrument creation and validation

74. At the heart of AHELO is a suite of online tests which measure later-year students' knowledge and capacity to solve real-world problems. Achievements in the discipline strands show that new assessments can be developed well. The assessments can triangulate existing tests and confer even more multidimensional perspectives on educational outcomes. The expansion of AHELO beyond Generic Skills, Economics and Engineering is an obvious way forward and example illustrative fields include History, Medicine or Mathematics. Building on the work of the Tuning Process (Tuning, 2011), another direction would be the move towards assessment in multidisciplinary fields, such as Cognitive Science, Information Sciences or Life Sciences. This would reflect the multidisciplinary demanded in many professional fields and subsequent increase in multidisciplinary approaches to higher education learning and teaching, enabling sub-fields to be bridged and significant efficiencies in. To be sure, there may well be fields in which it is impossible or unhelpful to generate relatively uniform objective assessments.

75. The AHELO Feasibility Study has stimulated discussion about the extent to which it is possible to define and measure a set of context-free generic skills. The assessments in the Feasibility Study comprise three design alternatives: a discipline-neutral Generic Skills strand; a test that encompasses discipline-specific (Engineering) generic skills; and a test more focused on discipline-specific competency (Economics). The Feasibility Study will enable some evaluation of the extent to which generic reasoning is usefully separable from disciplinary context.

76. As clarified by experience in the AHELO Feasibility Study, it is essential that all tests are developed with formative input from practitioners, discipline experts and technical specialists. Tests can and should deploy a range of item types, must be developed to the highest technical standards, and must be transparent to all users with respect to underlying constructs and item properties.

77. Experience in AHELO also indicates the value of having assessments which are underpinned by rigorous assessment frameworks. Frameworks define the domain being measured, conceptually organise the domain, and detail the form and characteristics of assessment tasks. Frameworks synthesise curriculum materials, outcome specifications, and regulatory frameworks. Expert groups consisting of faculty, industry and regulatory experts review and validate the frameworks. In the absence of frameworks, it is difficult to define and position what assessment instruments are seeking to measure and report.

78. In principle, AHELO's test items require students to examine, analyse and synthesise prompt materials, and to draw on discipline and/or generic knowledge to construct responses that solve real-world problems. Evidence from the AHELO Feasibility Study shows that tests can employ a range of item types, with formats balanced to ensure coverage, authenticity, cultural sensitivity, efficiency and rigor. Tests are delivered online over a secure internet connection, facilitating advanced authentication and response data verification.

Quality-assured operations

79. Like any study of its scale, AHELO involves a range of complex technical and operational activities. As the AHELO Feasibility Study suggests, these can be distilled into three phases: preparation, assessment and reporting.

80. Preparation consists of countries establishing a national management centre that coordinates national activities and is responsible for liaison with international project managers. The National Centre adapts, translates and validates test materials (either directly or through subcontracting arrangements), and coordinates institutional implementation. Supported by specialised training, international project managers and a library of support materials, each country's National Centre promotes AHELO and engages institutions. The National Centre helps institutions prepare for testing.

81. Assessment involves the collection and reporting of data. This involves sampling and recruiting students into the study. Testing is conducted by institutions under standard conditions and underpinned by robust quality assurance regimes. Tests are delivered online in supervised environments. Faculty are involved in scoring open-ended responses.

82. Informative reports are prepared by international project managers, and distributed to systems and then institutions. Reports are designed for monitoring and continuous improvement, and include high level results and breakdowns by key context characteristics. Reports map out levels of proficiency against psychometrically validated variables. Protocols are used to ensure institutional and individual confidentiality.

83. To ensure quality control, all national activities need to be supported through systematic training and support as well as consistent documentation of processes and responsibilities. The importance of transparent and regular communication in the management of national activities cannot be understated and is a critical component of the success of AHELO.

84. It is essential that AHELO stakeholders have confidence that data and reports are fit for use for the intended purposes. To ensure this, strict quality control is an integral part of all assessment activities and is critical to help ensure that AHELO provides data that are comparable across institutions and subgroups. This quality control has three basic components:

- development of AHELO Technical Standards
- monitoring the quality of resources and processes, and
- detection of variations to design during implementation.

85. Due to its original nature, the AHELO Technical Standards are being developed as part of the study. These AHELO Technical Standards will be clearly documented, ratified by the GNE, and embedded in leadership, management and training operations. They will convey that:

- instruments that have known and acceptable reliability;
- multifaceted forms of embedded and independent quality monitoring;
- high-quality manuals, advisory structures and security procedures;
- tests with good content, construct, criterion and face validity;
- analysis of relevant multilevel contexts;
- transparent translation, adaptation and verification procedures;

- clear population definitions and scientific sampling approaches;
- well-regulated training and field operations;
- stringent response rate requirements, and coordinated coding operations;
- data cleaning and verification to international standards;
- complex weighting and variance estimation techniques;
- psychometric scaling, and analysis of item validity and performance; and
- production of reports and data products to international standards.

86. The verification of translated assessment materials is a particularly important element of quality control. Ensuring that translations have not altered the constructs to be tested or made the assessment tasks more or less difficult for students is essential in ensuring the validity and reliability of assessments. The use of both domain and linguistic experts has been vital to the success of verification procedures in AHELO and ensures that translation quality is optimised.

87. Monitoring the quality of resources and processes plays out in a variety of ways, including:

- collaborative international consortia with overlapping expertise;
- the implementation of well-tested methodologies;
- a phased management approach;
- transparent communication with key stakeholders;
- ongoing monitoring, training and support activities;
- embedded quality review processes designed to detect and revise deficiencies;
- the development of an explicit quality assurance plan;
- strategies for engaging and managing the participation of key stakeholders;
- the involvement of international experts in design and development; and
- routine reporting to the OECD and GNE.

88. Complexity, uncertainty and the unexpected is the norm in large-scale studies, and an integral part of any effective quality control regime is having effective procedures in place to detect any deviations to the design. While not fully implemented in the AHELO Feasibility Study, key future procedures must include:

- effective communication structures that enable problem identification;
- independent quality monitors who audit materials and processes;
- the availability of experts who can assist with problem diagnosis;
- resources (notably, time and expertise) to develop effective solutions;
- governance and management structures that have the authority to implement workarounds; and
- diligent record-keeping to create and audit trail and support with subsequent review and reporting.

Engaging data and reports

89. 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 scale, in which an assessment cycle might span years, it is important to create various opportunities for engagement. Effective engagement is decisive for the study's impact on reform.

90. The AHELO Feasibility Study is producing a suite of international reports, along with reports for institutions. Data products will also be produced. Reports will be framed by guidelines developed as part of the Feasibility Study.

91. The design of international, institutional, policy, technical and scholarly reports in AHELO is driven by aspirations to:

- provide leaders with information on how education can be steered to preserve and enhance their country's highly skilled human capital;
- deliver valid, reliable and relevant data on learning outcomes within a common framework that systems and institutions can use for benchmarking and continuous improvement;
- allow systems to use learning outcomes data for transparency initiatives, to understand and promote diversification, to evaluate accountability systems, and to monitor the performance and contribution of higher education;
- create new communities of faculty engaged in designing and evaluating assessment materials, and using evaluation data to transform policy and practice;
- assist institutions measure performance and guide growth through informative benchmarking reports, to showcase their achievements within the framework of their distinctive missions and goals, and to demonstrate the quality of provision to employers, regulators, and other stakeholders;
- advance traditional definitions of 'quality', giving sharper focus to learning outcomes and graduate capability within a comparative global context;
- provide valid and meaningful data that is benchmarked internationally, thus provoking powerful and innovative thinking about improving teaching and the student experience;
- engage students in the global world of knowledge through reports that provide international perspectives on the standard and scope of achievement; and
- deploy robust and effective methods, validated test instruments, and informative approaches to reporting.

Global positioning

92. Learning lies at the heart of higher education, and outcomes data is relevant to a wide range of stakeholders. In themselves, AHELO's processes provide a real means of building assessment capacity,

and in leading evidence-based change. Experience in the feasibility study has affirmed sentiments expressed at the outset—that AHELO has the potential to provide feedback to learners, systems, institutions, faculty and researchers. Linking AHELO into existing networks, and exploring emerging contexts, is an important part of the study’s short-term evolution.

93. The AHELO Feasibility Study has yielded important new insights into core facets of higher education. At the same time, it has 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 tuning/alignment, ranking and classification exercises, determine the broader significance of AHELO, build national capacity within funding, regulatory and quality agencies, boost institutional capacity to manage and understand outcomes data, assist students and their parents make better sense of information on higher education, and support emerging research communities that underpin applied work with methodological and scholarly inquiry.

REFERENCES

- ACER, CHEPS & Indiana University (2011). *Contextual Dimension Assessment Framework*. Paris: OECD.
- ACER, NIER & University of Florence (2011). *Engineering Assessment Development Report*. Paris: OECD.
- ACER, University of Queensland & Monash University (2011). *Australian Medical Assessment Collaboration*. Accessed 1 September 2011 from: www.acer.edu.au/amac
- AHELO Consortium (2010). *AHELO Feasibility Study Assessment Design*. Paris: OECD.
- AHELO Consortium (2011). *July Interim Feasibility Study Report*. Paris: OECD.
- Center for World Class Universities (2010). *Academic Ranking of World Universities 2010*. <http://www.arwu.org/ARWU2010.jsp>. Accessed 3 March 2011.
- CHERPA-Network (2010). *Design and Testing the Feasibility of a Multi-dimensional Global University Ranking: U-Multirank Interim Progress Report*. Design Phase of the Project.
- Coates, H. & Radloff, A. (2008). *Tertiary Engineering Capability Assessment: Concept Design*. Canberra: Group of Eight.
- Coates, H. & Richardson, S. (2011). An international assessment of bachelor degree graduates' learning outcomes. *Higher Education Management and Policy*, 23(3), 51-69.
- Coates, H. (2010). Defining and monitoring academic standards in Australian higher education. *Higher Education Management and Policy*, 22(1), 1-17.
- ETS & ACER (2011). *Economics Assessment Development Report*. Paris: OECD.
- ETS (2011). *Economics Assessment Framework*. Paris: OECD.
- MSC-AA (2011). *Medical Schools Council Assessment Alliance*. Available from <http://www.medschools.ac.uk/MSA-AAA/Pages/default.aspx>
- Nusche, D. (2007). *Assessment of Learning Outcomes in Higher Education: A comparative review of selected practices*. Accessed 3 June 2008 from: www.oecd.org/dataoecd/14/8/40257354.pdf
- Technical Advisory Group (TAG). (2011). *OECD AHELO Technical Advisory Group April 2011 TAG Report*. Paris: OECD.

TSL Education (2010). *World University Rankings 2011*. <http://www.timeshighereducation.co.uk/world-university-rankings/>. Accessed 10 June 2011.

Tuning Association (2009a). *A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in the Science of Economics*. Accessed 1 March 2011 from: www.oecd.org/edu/ahelo/callfortenders

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

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

US News and World Report (2011). *Best Colleges*. Accessed 28 September 2011 from: www.usnews.com/education

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