DIRECTORATE FOR EDUCATION
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD

Group of National Experts on the AHELO Feasibility Study

INITIAL INSIGHTS AND PROPOSALS FOR PHASE 2
6th meeting of the AHELO GNE

Paris, 28-29 March 2011

This document was prepared by the ACER Consortium.

The AHELO GNE is invited to:
• TAKE NOTE and DISCUSS the initial insights from Phase 1
• TAKE NOTE and DISCUSS the consortium proposal for Phase 2
Design and plans

March 2011
ASSESSING HIGHER EDUCATION LEARNING OUTCOMES
ON AN INTERNATIONAL SCALE

Introduction

1. This document takes stock of AHELO, positioning progress to date and providing plans for completing the AHELO Feasibility Study. The document proceeds in four sections, with these focussing on:

- general contexts and rationales for assessing higher education learning outcomes on an international scale;
- emerging design characteristics of a feasible international assessment;
- an assessment of the feasibility of developments to date; and
- work plans for Modules E, A, B, C and D to conclude the AHELO Feasibility Study.

2. To set the scene, the balance of this first section steps back from the momentum of current operational and technical progress and reviews relevant contexts and AHELO’s unique value proposition. In March 2011, near the middle of the study, these broad insights frame the designs presented later in the document.

3. The second section of this document reviews what would appear to be key features of a feasible assessment design. This design builds on foundations set through the AHELO Feasibility Study to offer an illustrative conceptualisation of key characteristics of a successful international assessment of higher education learning outcomes. Discussion is not prescriptive, but offers a suggestive vision that contextualises ongoing developmental and evaluative work.

4. Working from these foundations, the third section reviews what evidence there is that instrumentation and pre-implementation work conducted to date sustains these broader rationales and conceptualisations of what AHELO could achieve. So far, in short, is there evidence that AHELO is feasible? The early results are very promising, and suggest that AHELO is indeed looking both feasible and achievable.

5. The final section of this document details the implementation work that is required to conclude the AHELO Feasibility Study in accordance with the 2009 AHELO Call For Tender and AHELO’s Assessment Design.

Pressing trends in global higher education

6. Higher education is growing in significance and scale. More people than ever are undertaking qualifications that will change their lives, and their societies. Innovation, knowledge and human capital are at the core of progress in our economies and societies, and higher education is fundamental in preparing students to become effective workers and citizens, and in fostering social mobility and equity. Systems and institutions have an intrinsic interest in understanding what graduates know and can do.

7. At the same time higher education is faced with huge cost pressures. After decades of rapid expansion higher education has grown and many nations are moving from mass to universal scale systems.
Institutions compete globally for students, costs of provision are increasing, and particularly in recent years students and systems have had less money to invest.

8. Internationalisation is an inherent part of contemporary higher education. While long part of institutional research agendas, international opportunities and pressures now touch many facets of educational operations as well. This lends particular significance to international perspectives that speak to the core business of teaching and learning.

9. Such changes place an increasing emphasis on the need for sophisticated forms of management, evaluation and quality monitoring. Systems need to do more with less, which further affirms the value of evidence-based decision making. Institutions need data to establish positions in increasingly competitive markets, to identify how to attract, educate, retain and graduate students, and to meet the demands of accreditation and funding agencies.

10. Policymakers and the public increasingly pay attention to the outcomes of higher education. Nevertheless, efforts to improve the quality of teaching and enhance learning outcomes suffer from a serious and significant information gap. We know very little about what and how students learn in higher education. The reputations and rankings of higher education institutions are based largely on research performance. This distorts leadership and management as institutions engage in a costly and damaging race to a ‘unidimensional top’, often to the detriment of teaching and learning. Meanwhile, almost one in every three students enrolling in higher education will drop out without a degree. Students from disadvantaged backgrounds suffer more than anyone else, thereby fostering inequity.

AHELO: A clear and unique contribution

11. AHELO plays a unique and increasingly significant role within this global dynamic. AHELO is an online assessment of final-year students’ capacity to use, apply and act on their knowledge and reasoning. AHELO builds on robust methods and promotes a clear vision for future higher education moves beyond existing rankings to provide new insights into core facets of higher education. It carries a real potential of yielding information on the outcomes of higher education which help to counter serious challenges posed by growth, quality, cost and internationalisation.

12. AHELO is an innovative endeavour driven by significant rationales:
   - Providing leaders with information on how education can be steered to preserve and enhance their country’s highly skilled human capital.
   - Delivering valid, reliable and relevant data on learning outcomes within a common framework that systems and institutions can use for benchmarking and continuous improvement.
   - Allowing systems to use learning outcomes data for transparency initiatives, to understand and promote diversification, to evaluate accountability systems, and for monitoring the performance and contribution of higher education.
   - Creating new communities of faculty engaged around the definition and production of assessment materials, and the transformation of results into changing policy and practice.
   - Assisting institutions to measure performance and guide growth through informative benchmarking reports, and to demonstrate the quality of provision to employers, regulators, and other stakeholders.
   - Advancing traditional definitions of ‘quality’, giving sharper focus to learning outcomes and graduate capability within a comparative global context.
• Providing valid and meaningful data that is benchmarked internationally to provoke powerful and innovative thinking about improving teaching and the student experience.

• Engaging students in the global world of knowledge through participation in AHELO. Reports provide international perspectives into the standard and scope of achievement.

• Deploying robust and effective methods, validated test instruments, and informative approaches to reporting.

13. AHELO is designed to engage systems, institutions, teachers and students in evidence-based change. The assessment is based on collaboration between ministries, discipline and technical experts, faculty and learners, institutional leaders, and a wide range of global stakeholders. As demonstrated by early results from developmental work undertaken since 2010 AHELO has the potential, without loss of quality, to be scaled to very large numbers of systems, institutions and students.

14. There are significant risks associated with failure to progress a high-quality international AHELO. Not least, flawed, simplistic international ranking systems will continue to proliferate. The absence of data from a respected international agency will take away one important reason national leaders have to create education policies that improve their post secondary education systems.

15. The need for valid data is growing and AHELO’s value proposition is clear. AHELO responds to the growing need to ensure that higher education delivers the right set of skills which allow students and graduates to thrive. Measures of learning outcomes are the key to diagnosis and reform in higher education institutions worldwide.
EMERGING DESIGN ELEMENTS OF A FEASIBLE AHELO

Introduction

16. The AHELO Assessment Design sketches key characteristics of a feasible AHELO. In March 2011, close to the mid-point of the AHELO Feasibility Study, it is both helpful and possible to give greater clarity and substance to the design of a successful international assessment.

17. Working from the AHELO Assessment Design and the developmental and evaluative work conducted thus far, this section explores a high-level, illustrative and future-looking perspective of a feasible international assessment of higher education learning outcomes. This is not a formal nor comprehensive picture or proposition, but rather a suggestive examination of key parameters and possibilities. Hence the design is not intended to be prescriptive, but to offer a suggestive vision that contextualises ongoing developmental and evaluative work. What, in short, do achievements to date portend about the shape and approach of an international assessment of higher education learning outcomes?

A sustainable and transparent model

18. 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 and change. Figure 1 sketches the various cycles of engagement and change.

19. The AHELO Feasibility Study is providing an unparalleled opportunity to develop and test a feasible model for AHELO. The sustainable model that is emerging reflects a balance between quality, efficiency and cost. It takes account of national and international contexts and relativities.

20. The highest standards are expected for AHELO, and there is strong global interest in AHELO’s approach and outcomes. Over the last 50 years, quality standards have been developed through dozens of international education studies, and AHELO must meet and surpass these expectations. To that end, AHELO must build on technical approaches tested globally over many decades in international studies,
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.

21. The AHELO Feasibility Study has affirmed that the main cost drivers in a study of this kind are scale, known complexities, unknown challenges, capacity development, system development, implementation costs, and quality assurance. While most direct and indirect costs for any study of this kind occur within countries, much can be done to reduce in-country costs through efficient international design. Conversely, small cost savings at the international level can reduce collective synergies and lead to substantial in-country cost increases for all participants.

22. Confidentiality and security are intrinsic to high stakes testing, but interest in the AHELO Feasibility Study has affirmed that transparency is also vital. Higher education is an industry that gives great weight to scholarly peer review of material and methods (for instance: task properties, 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.

**Multidimensional instrumentation**

23. At the heart of AHELO is a suite of online objective tests that measure later-year students’ knowledge and capacity to solve real-world problems. Current student assessments focus on generic, engineering and economics skills. All students complete a brief context instrument, as do faculty and institutions (see Figure 2).

![Figure 2: AHELO Feasibility Study instrumentation](image)

24. Higher education exists within discipline contexts, and to make institutional change work AHELO results need to speak to people who teach, learn and lead in disciplines. Achievements to date show that new assessments could well be developed to triangulate existing tests and confer even more multidimensional perspectives on educational outcomes. Example illustrative fields include history, medicine or mathematics (see Figure 3).
25. 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.

How AHELO works

26. 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 (see Figure 4).

27. 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, 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 centre helps institutions prepare for testing.

28. 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.

29. 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.

**Test creation and validation**

30. As clarified by experience in the AHELO Feasibility Study, it is important that all tests are developed with formative input from practitioners, discipline experts and technical specialists. Tests can deploy a range of item types and should be developed to the highest international standards.

31. Expert Groups are established for each assessment. These groups consist of leading international experts – faculty, industry professionals, and people from regulatory agencies – who oversee a structured development process (see Figure 5).

32. Assessment frameworks are produced that 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 review and validate the frameworks.

33. Items are created with formative input from faculty and discipline experts. Item generation workshops yield draft materials which are subjected to rigorous technical review. Remaining items are mapped against the assessment framework and across the range of expected student performance. Final items are translated and adapted in-country, prior to independent verification of translation.

34. To ensure their validity and the transparency of the development process, items must be subjected to rigorous validation and review. This includes focus groups and cognitive testing, and psychometric examination of data from pilot tests. Items are then loaded into the delivery system.

*Figure 5: Instrument development process*
Evidence-based contributions

35. 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.

36. The real value and contribution of AHELO derives not just from reading results, but from international communities formed to construct tasks, undertake assessments, and use results to guide individual, institutional and system growth. As experience to date has shown, AHELO has the potential to provide feedback to learners, systems, institutions, faculty and researchers (see Figure 6).

Figure 6: Potential horizons for change
CURRENT INSIGHTS INTO AHELO’S FEASIBILITY

Introduction

37. The feasibility of the design sketched above hinges on proof that it is feasible to develop instruments that measure learning outcomes at an international scale, and that it is feasible to implement these assessments using methods that are scalable, efficient and secure. The AHELO Feasibility Study is part-way through testing the science and practicality of the assessment. This section provides preliminary insights into the evaluation of feasibility to date.

Evaluation structure

38. In broad terms, the main criterion for the success of the AHELO Feasibility Study is evidence that the various instruments considered can be applied in diverse settings with appropriate adaptations and yet provide valid, reliable and free-of-bias measures of student learning outcomes. The scientific feasibility of AHELO depends on whether it is possible to develop assessment instruments that will be 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. As the next section of this document shows, practical feasibility will be 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 feasibility study assessments has brought benefits to participating institutions and demonstrated its value for teaching improvement.

39. The AHELO Analysis Plan organises these considerations and provides the evaluative architecture for the feasibility study [EDU/IMHE/AHELO/GNE(2010)18]. The Analysis Plan specifies research questions posed by the various strands, outlines the best means of assessing 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.

40. AHELO was first announced in 2006 and several planning and policy meetings were held between 2007 and 2009. Operational and technical work began in early 2010 with planning meetings in New York City and Washington DC. Much has been progressed and achieved over the last 15 months, and it is timely and appropriate to document the emerging evaluative observations. This initial assessment helps take stock of contributions and challenges, create a constructive future, and guide implementation planning.

41. To recap, the aim of the AHELO Feasibility Study is to assess whether reliable cross-linguistic, cross-cultural and cross-institutional comparisons of higher education learning outcomes are scientifically possible and whether their implementation is feasible.

42. Two main research questions underpin the study’s scientific and practical dimensions:

1. Is it scientifically possible to produce valid cross-linguistic, cross-cultural and cross-institutional comparisons of higher education learning outcomes?
2. Is it feasible to implement a valid cross-linguistic, cross-cultural and cross-institutional assessment of higher education learning outcomes?

43. If successful, AHELO would be the first international assessment of higher education learning outcomes. There are thus a large number of subsidiary questions that need to be considered to assess whether such an endeavour is scientifically or practically feasible.

44. The AHELO Analysis Plan presents these subsidiary questions in an operationalisable framework. The framework specifies 19 subsidiary research questions, a series of underpinning ‘feasibility indicators’, and diverse sources of evidence.

45. The main sources (along with abbreviations used in Figure 7) include:

- AHELO Consortium (AC);
- Expert Group (EG);
- Faculty Respondents (FR);
- Group of National Experts (GNE);
- Institutional Coordinators (IC);
- National Project Managers (NPM);
- Organisation for Economic Cooperation and Development (OECD);
- Stakeholders Consultative Group (SCG);
- Student Respondents (SR);
- Test Administrators (TA); and
- Technical Advisory Group (TAG).

Emerging insights

46. Looking broadly, excellent progress has been made on the development, adaptation and translation of assessment frameworks and instruments measuring Generic Skills, Economics and Engineering. Work has begun on the Contextual Dimension, as has qualitative testing in the discipline strands and pre-implementation preparations for Generic Skills. Large numbers of international experts have contributed through the study’s advisory groups. Moreover, a growing number of countries have decided to participate and even more are expressing interest. National Project Managers have reported strong engagement from institutions. Overall, there is significant and growing momentum, interest and goodwill among experts, countries and institutions.

47. The AHELO Feasibility Study is a landmark international activity for higher education, requiring considerable creativity and innovation. At this stage – perhaps around half-way through the feasibility study – a much more modest attempt is made to distil emerging high-level insights that help take stock and guide the implementation phase of the study.

48. Figure 7 provides a status update on feasibility assessment. For both instrumentation and implementation phases, it details the research questions posed in the AHELO Analysis Plan and sources of evidence. Where applicable, the codes indicate the status of the evaluation and feasibility assessment (see key opposite).
49. It should be noted that the assignment of sources of evidence to research questions is based on the ‘feasibility indicators’ specified in the AHELO Analysis Plan, not along governance or management lines. It must be stressed that feasibility assessments are made within the scope of the study, which itself has been limited in numerous major ways compared with other national/international assessments.
<table>
<thead>
<tr>
<th>Phase</th>
<th>Research question</th>
<th>Source of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Was the Generic Skills assessment framework reflective of an international consensus about the areas that are important to assess?</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>2. Was the instrumentation developed on the basis of the Generic Skills framework faithful to the spirit and intent of the framework?</td>
<td>F F P</td>
</tr>
<tr>
<td></td>
<td>3. Was the provisional Economics assessment framework reflective of an international consensus about the important learning outcomes in Economics?</td>
<td>F F P</td>
</tr>
<tr>
<td></td>
<td>4. Was the instrumentation developed on the basis of the Economics framework faithful to the spirit and intent of the framework?</td>
<td>F F P</td>
</tr>
<tr>
<td></td>
<td>5. Was the provisional Engineering assessment framework reflective of an international consensus about the important learning outcomes in Engineering?</td>
<td>F F P</td>
</tr>
<tr>
<td></td>
<td>6. Was the instrumentation developed on the basis of the Engineering framework faithful to the spirit and intent of the framework?</td>
<td>F F P</td>
</tr>
<tr>
<td></td>
<td>7. Was the provisional Contextual Dimension conceptual framework reflective of an international consensus about the important contexts that shape higher education learning outcomes?</td>
<td>P P P</td>
</tr>
<tr>
<td></td>
<td>8. Was the instrumentation developed on the basis of the Contextual Dimension framework faithful to the spirit and intent of the framework?</td>
<td>P P P</td>
</tr>
<tr>
<td></td>
<td>9. Was the AHELO Assessment Design valid and feasible?</td>
<td>F F F F F F</td>
</tr>
<tr>
<td></td>
<td>10. Was the study implemented in a methodologically rigorous fashion?</td>
<td>P P Y Y Y</td>
</tr>
<tr>
<td></td>
<td>11. Was the study feasible from a practical perspective?</td>
<td>P P Y P P Y</td>
</tr>
<tr>
<td></td>
<td>12. Was the study successfully generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally?</td>
<td>Y Y Y Y Y</td>
</tr>
<tr>
<td></td>
<td>13. Was it possible to engage systems and institutions in the study?</td>
<td>P P Y P P Y</td>
</tr>
<tr>
<td></td>
<td>14. Was it possible to engage faculty respondents in the study?</td>
<td>Y Y Y Y Y</td>
</tr>
<tr>
<td></td>
<td>15. Was it possible to engage a random sample of student respondents in the study?</td>
<td>Y Y Y Y Y</td>
</tr>
<tr>
<td></td>
<td>16. Was the instrument delivered successfully?</td>
<td>Y Y Y Y Y</td>
</tr>
<tr>
<td></td>
<td>17. Were survey operations successfully managed by countries and institutions?</td>
<td>P P Y P Y</td>
</tr>
<tr>
<td></td>
<td>18. Was it possible to train people in different countries to score tasks in cross-linguistically and cross-culturally generalisable ways?</td>
<td>P Y P Y Y</td>
</tr>
<tr>
<td></td>
<td>19. Were statistical reports of assessment results of value to systems and institutions?</td>
<td>Y Y Y Y Y Y</td>
</tr>
</tbody>
</table>

**Figure 7: Formative insights on feasibility**
What Figure 7 shows is that, as much as has been possible given workflows to date, significant advances have been made in establishing the feasibility of AHELO. Major outcomes have been achieved – such as international agreement regarding discipline frameworks - and many challenges have been overcome. At this stage no facet of the study has been stamped ‘N’ for ‘NOT FEASIBLE’. This is an important finding.

Table 1 and Table 2 present various evidence of feasibility in terms of the feasibility indicators given in the AHELO Analysis Plan. For parsimony, only questions, indicators and evidence of current relevance are listed. For example, no information is given about differential item functioning, as this requires analysis of data from test implementation. Questions pertaining to the Contextual Dimension are not listed, as this work has only just begun.

### Table 1: Feasibility of instrumentation – emerging insights

<table>
<thead>
<tr>
<th>Research question</th>
<th>Feasibility indicators (underlined) and evidence (dot points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the Generic Skills assessment framework reflective of an international consensus about the areas that are important to assess?</td>
<td>Achieved consensus among experts on the domain to be assessed</td>
</tr>
<tr>
<td></td>
<td>⇒ Selection of CLA to measure facets of Generic Skills domain</td>
</tr>
<tr>
<td></td>
<td>Achieved consensus among experts on the balance given to elements in the domain</td>
</tr>
<tr>
<td></td>
<td>⇒ Selection of CLA to measure facets of Generic Skills domain</td>
</tr>
<tr>
<td></td>
<td>Degree to which framework accounts for institutional, language and cultural differences</td>
</tr>
<tr>
<td></td>
<td>⇒ Extensive translation/adaptation processes developed and changes suggested</td>
</tr>
<tr>
<td>2. Was the instrumentation developed on the basis of the Generic Skills framework faithful to the spirit and intent of the framework?</td>
<td>Construct, content and face validity of instrument considered appropriate</td>
</tr>
<tr>
<td></td>
<td>⇒ Adaptation of two (of nine) USA-derived performance tasks for international contexts</td>
</tr>
<tr>
<td></td>
<td>⇒ Development of trial performance tasks to familiarise students with this type of assessment</td>
</tr>
<tr>
<td></td>
<td>Instrument has appropriate levels of reliability</td>
</tr>
<tr>
<td></td>
<td>⇒ Successful replication (via translation/adaptation) of instrumentation for different national contexts, giving support to the generalisability of the instrument</td>
</tr>
<tr>
<td>3. Was the provisional Economics assessment framework reflective of an international consensus about the important learning outcomes in Economics?</td>
<td>Achieved consensus among experts on the domain to be assessed</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group formed with members from Italy, Japan, Mexico, Netherlands, Russian Federation, and USA</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group reviewed initial framework, advised changes and approved final version</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group specified six detailed learning outcomes based on those identified by the AHELO Tuning Expert Group</td>
</tr>
<tr>
<td></td>
<td>Achieved consensus among experts on the balance given to elements in the domain</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group discussed coverage of framework and balance within it</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group reached agreement on their coverage in final draft of framework</td>
</tr>
<tr>
<td></td>
<td>Degree to which framework accounts for institutional, language and cultural differences</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group discussed differences in way Economics taught in different countries and reached agreement on applicable areas</td>
</tr>
<tr>
<td></td>
<td>⇒ Key areas reflected in framework and in choice of assessment items</td>
</tr>
<tr>
<td></td>
<td>⇒ Framework and assessment items approved by Expert Group</td>
</tr>
<tr>
<td></td>
<td>Acceptance by a broader group of stakeholders and the community</td>
</tr>
<tr>
<td></td>
<td>⇒ Pending – NPMs have final versions of assessment materials and framework and are preparing these for translation</td>
</tr>
<tr>
<td></td>
<td>⇒ Until qualitative data has been collected there is little information on which to assess the views of the broader economics education community and economics students</td>
</tr>
<tr>
<td>4. Was the instrumentation developed on the basis of the Economics framework faithful to the spirit and intent of the framework?</td>
<td>Construct, content and face validity of instrument considered appropriate</td>
</tr>
<tr>
<td></td>
<td>⇒ Expert Group assessed a pool of constructed-responses and multiple-choice questions and approved two constructed-response questions and 50 multiple-choice questions for use in qualitative testing</td>
</tr>
<tr>
<td></td>
<td>⇒ Items selected reflect the aspects identified in the framework and the differential weighting in importance given to each one by the Expert Group</td>
</tr>
</tbody>
</table>
### Research question | Feasibility indicators (underlined) and evidence (dot points)
--- | ---
5. Was the provisional Engineering assessment framework reflective of an international consensus about the important learning outcomes in Engineering? | Achieved consensus among experts on the domain to be assessed
- Engineering Expert Group formed with members from Australia, Germany, Italy, Japan, Sweden, and USA
- Expert from Mexico as additional participant in face-to-face expert group meeting
- Expert Group reviewed and approved framework
- Feedback from Engineering Experts from the academic network of the Academic Network of European and Global Engineering Education (EUGENE) indicated widespread support for framework

Achieved consensus among experts on the balance given to elements in the domain
- Engineering Expert Group discussed need for framework to cover areas of key importance to domain of Civil Engineering
- Expert Group reached agreement on coverage of key areas in final draft of framework

Degree to which framework accounts for institutional, language and cultural differences
- Expert Group considered differences in Civil Engineering education worldwide and reached agreement on applicable areas
- Experts Group approved coverage of these core areas in framework and choice of assessment items

6. Was the instrumentation developed on the basis of the Engineering framework faithful to the spirit and intent of the framework? | Construct, content and face validity of instrument considered appropriate
- Expert Group assessed 12 constructed-responses and nearly 100 multiple-choice questions
- Expert Group approved 4 constructed-response questions and 40 multiple-choice questions for use in qualitative testing
- Items selected reflect the aspects identified in the framework and the differential weighting in importance given to each one by the Expert Group
- Expert Group assessed instrument to have construct, content and face validity
- Feedback from experts in EUGENE network indicated agreement with materials and coverage
### Table 2: Feasibility of implementation – emerging insights

<table>
<thead>
<tr>
<th>Research question</th>
<th>Feasibility indicators (underlined) and evidence (dot points)</th>
</tr>
</thead>
</table>
| 9. Was the AHELO Assessment Design valid and feasible? | Relevant scope for an international assessment of learning outcomes  
  - AHELO Assessment Design approved by political and technical governance committees  
  - Significant interest from countries evidenced by increase in participants  
  - Agreement on assessment frameworks and instruments among disciplinary experts  
  - Development of support documents and guides for countries  
  Degree of change to design during implementation, analysis and reporting  
  - Essential design is sustained despite changes in study contexts and scope  
  Degree to which design provides basis for full-scale extension  
  - Study scaled up to include additional countries  
  - All practices can be scaled up to include additional countries, institutions and respondents |
| 10. Was the study implemented in a methodologically rigorous fashion? | Count of significant variations between design and operationalisation  
  - Operationalisation has surpassed design considerations, despite cut in study budget and scope  
  Positive feedback on technical design and operationalisation  
  - Review of study design by technical, management and political committees  
  - Design sustained after feedback from hundreds of stakeholders  
  - Design provides flexible core for national implementations  
  Degree to which the methodology advances higher education research  
  - International development and validation of assessment frameworks suggests it is possible to define learning outcomes internationally  
  - Instrument development to international standards in Civil Engineering and Economics  
  - Student capability focus places attention of institutions on teaching role |
| 11. Was the study feasible from a practical perspective? | Study delivered within agreed timeframe and budget  
  - Study delivered within timeframe significantly shorter than original intention  
  - Study delivered within budget significantly less than original intention  
  NPMs deliver within agreed timeframe and budget  
  - Countries and NPMs have invested considerably in the study  
  - Countries and NPMs have worked through a large number of novel challenges  
  - Most NPMs appear on track to conclude translation/adaptation and qualitative testing  
  Extent to which communication processes enable the smooth flow of work  
  - Communication plan facilitates transparent and efficient communication arrangements  
  - Regular updates provided to countries to assist national management  
  - Online resources (WebEx, Skype, email, wiki) used for training and support  
  Extent to which information technology led to management efficiencies  
  - AHELO Exchange (online wiki) set up to enable secure exchange of materials  
  - Efficiencies in sharing, access to and updating of data files  
  - Increased use of many IT platforms for virtual meetings has led to significant cost effectiveness  
  - Versioning of documents controlled and errors minimised |
| 13. Was it possible to engage systems and institutions in the study? | Count of countries involved compared with initial plans  
  - Module A: growth from 5 to 9 countries  
  - Module B: growth from 5 to 7 countries  
  - Module C: growth from 3 to 7 countries  
  - Interest in participating/observing by Chinese Taipei, India, Kingdom of Saudi Arabia, and Tunisia |
Research question | Feasibility indicators (underlined) and evidence (dot points)
--- | ---
Count of institutions per country compared with expected | • Early feedback suggests most countries have engaged 10 or more institutions

17. Were survey operations successfully managed by countries and institutions? | Training of NPMs implemented in timely manner
--- | ---
• Rigorous adaptation, translation and verification guidelines and workbooks (ATAVs) prepared on time
• Detailed resources for qualitative testing and recording data prepared on time
• Training of NPMs in translation and adaptation took place in February 2011
• Online training of NPMs in qualitative testing concluded successfully in February 2011

**An emerging assessment**

52. 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 an AHELO is feasible.

53. Even a little less than half way through, the study has secured significant insights into scientific feasibility, including:
   - acceptance by policymakers and experts of the design of a feasible international assessment of higher education learning outcomes;
   - international consensus on the definition of generic, economics and civil engineering learning outcomes;
   - operationalisation of assessment frameworks in various forms of items for the purpose of measurement;
   - emerging definition of a contextual assessment framework relevant to learning outcomes; and
   - implementation of high-quality adaptation and translation procedures.

54. Several preliminary findings around practical feasibility also emerge:
   - effective coordination of a growing number of countries and more expansive participation across multiple strands of work (from 13 to 23 participant/strand);
   - timely production of high-quality work and deliverables under financial constraints;
   - development of a suite of manuals to assist with scalable and replicable international and national management, technical work, and operations;
   - building communities with key stakeholders and experts, that help position and assure the quality of processes and outcomes;
   - assistance with creating national centres, and with training national project managers and associated technical staff; and
   - centralised communication strategies and control of access to documents.

55. Together, these insights provide an 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 remain stronger than ever.
56. The progress made on AHELO to date shows that firm scientific and practical foundations have been established for phase two of the feasibility study, and potentially therefore for a full-scale AHELO. Key technical and operational parameters, opportunities and constraints have been clarified.
THE FEASIBILITY OF AHELO’S IMPLEMENTATION

Introduction

57. The clear case for AHELO and the risks of not proceeding, coupled with major progress and sound indications of feasibility, affirm the value of testing the feasibility of implementation.

58. Clear fieldwork designs, plans and resources are essential if AHELO is to be tested empirically in a transparent and rigorous fashion that meets international standards. The application of rigorous and internationally scalable procedures that can be used to assure the quality of AHELO is absolutely essential.

59. Information on the nature and rationales of fieldwork methods is provided in the AHELO Assessment Design. Work to be conducted as part of implementation includes:

- Operationalising Economics, Engineering and Context instruments for delivery;
- Engaging institutions and sampling faculty and students;
- Coordinating survey procedures and operations;
- Analysing results; and
- Producing data products and written reports.

60. This involves building international systems, and supporting countries and institutions to implement these systems. These implementation plans specify work which must be undertaken in 2011 and 2012 – beginning as early in 2011 as possible – to keep the AHELO Feasibility Study on track.

61. Note that these tasks are in addition to (not the same as or instead of) ‘pre-implementation’ work undertaken by the Council for Aid to Education (CAE) for the AHELO Generics Skills Strand. The work specified in this document is required as well as the ‘pre-implementation’ work undertaken by CAE both to support the Generic Skills Strand and also to implement, and hence test, the feasibility of AHELO.

Implementation plans

62. Module E implementation involves the overall coordination, analysis and quality control of the AHELO Feasibility Study. The Module E team will ensure the highest possible standards of process and outcomes are achieved from the project and will work closely with those involved in Modules A, B, C and D to ensure clarity, coordination and consistency across the different strands of the project.

63. This work will be led by the Australian Council for Educational Research (ACER), cApStAn (linguistic quality control), with input from the following partners: Educational Testing Service (ETS), IEA DPC and Statistics Canada (recruiting and sampling), SoNET (online platform), and CAE (implementation of Generic Skills Strand). Together, these partners form the ‘Module E team’.

64. The work will be directed by Hamish Coates (ACER) in close collaboration with Roger Benjamin (CAE), and overseen by the AHELO Technical Advisory Group (TAG) (Dr Peter Ewell, Chair). National Project Managers (NPMs) will play an important role, as will directors of Modules A, B, C and D.
The following work will be conducted:

1. **Project Management**: ACER will be responsible for the oversight and coordination of Module E and therefore the whole AHELO Feasibility Study. This role involves:
   a. Working closely with Consortium members and the OECD Secretariat to ensure:
      i. effective and efficient completion of implementation tasks;
      ii. the consistent use of the highest quality standards in the work completed;
      iii. openness and transparency in all dealings with each other, with AHELO Feasibility Study participants, and with the OECD Secretariat; and
      iv. the application of strict financial controls.
   b. More specifically, in its overall management role, the ACER team will:
      i. maintain the AHELO Exchange to enable efficient management and sharing of documentation for all AHELO Feasibility Study participants;
      ii. coordinate the organization of Expert Groups and the Technical Advisory Group (TAG);
      iii. assist countries maintain a National Centre;
      iv. support National Project Managers (NPMs) in each participating country; and
      v. assist NPMs, where required, to interact with Institutional Coordinators (ICs) in each participating institution.

2. **Instrument delivery**: SoNET will be primarily responsible for the instrument delivery platform for Modules B, C and D. The SoNET platform, which is already used for international assessments:
   a. enables multiple language versions to be adapted for each assessment;
   b. supports a wide range of multiple-choice and short-response item types;
   c. supports online marking and marker workflow management;
   d. incorporates existing testing software and consequently greatly reduces development time; and
   e. enables instruments to be administered online, using the internet to connect institutional computers to remote servers.

3. **Engaging institutions and sampling faculty and students**: ACER, IEA DPC and Statistics Canada will be the primary organisations responsible for this part of Module E. This will involve:
   a. Constructing robust population definitions for students, staff and institutions to be involved in the study.
   b. Preparing the AHELO Feasibility Study sampling plan, which will involve:
      i. providing NPMs with further guidance on selecting a convenience sample of higher education institutions (typically ten or more per participating country);
      ii. providing further resources to assist NPMs engage institutions to participate and providing guidance for the establishment of the study in each institution;
      iii. facilitating the selection of an unbiased, probabilistic sample of students in each institution, which may require the adaptation of a general sample design to local conditions; and
      iv. sampling faculty within each institution.
c. Sample monitoring and quality control undertaken in conjunction with NPMs.

d. Specifying sample size and participation rate requirements, for students and faculty.

e. Preparing a sampling manual which will:
   i. clearly describe all different steps required to implement the national sampling plans, select a convenience sample of institutions, and identify students and faculties to be sampled;
   ii. be tailored to each country’s specificities, bearing in mind the population(s) invited to the AHELO Feasibility Study in that country; and
   iii. account for the specific requirements of each strand and allow for adaptation to local (national or institutional) conditions.

f. Prepare estimation and replication weights for each sample of students meeting the agreed-upon participation rates.

4. Survey procedures and operations: ACER and ETS will be the organisation with the responsibility for this aspect of Module E, involving a number of logistical and operational tasks:

a. Test administration preparation, which involves supporting ICs with producing an institution coordinator manual so that the assessments are administered according to the standardised procedures and scripts.

b. Facilitation of coding operations including:
   i. coordinating one or more coding seminars for participating countries;
   ii. creating a coding manual (in close collaboration with teams leading Module B, C and D);
   iii. designing, planning and analysing coder reliability statistics.

c. Collection and verification of data, involving the development and implementation of a number of databases for field operations all of which will be matched and linked using unique multilevel record identifiers so that once a final database is built all cases can be tracked and assigned to the corresponding institution.

d. Oversight of the production of a number of manuals to guide NPMs, ICs and others involved in the study through its operational aspects (where appropriate, these will be tailored to individual countries or institutions).

e. Monitoring quality and survey procedures primarily through the NPMs, who will be assisted in planning and implementing each step of the project. This will involve monitoring:
   i. sampling;
   ii. the preparation of the instruments;
   iii. the coding of responses; and
   iv. data collection and data management.

f. Data verification and processing to ensure that the data adhere to international formats, that data confidentiality is maintained, that data from different instruments can be linked between different survey files, and that the data accurately and consistently reflect the information collected within each country. This will involve:
   i. standardisation of national file structures;
   ii. establishing cleaning rules and procedures; and
iii. computation of weights.

5. **Implementation of Generic Skills Strand:** CAE is responsible for operationalisation and implementation of the Generic Skills assessment. Organisations in the Module E team, lead by ACER, will work with CAE to ensure that the Module A assessment is consistently incorporated into the whole AHELO Feasibility Study. The Module E team will work with CAE in this regard to plan fieldwork preparation and management. Key links in these processes will be made with those being implemented for Modules B, C and D. This task will include:

a. making available two translated performance tasks;

b. distributing fundamental instructional tools for proctor training, and coordinating with NPMs a small-scale piloting of the testing platform to provide practical experience in advance of testing;

c. agreeing on coding criteria for performance tasks and translating into country languages;

d. CAE, in consultation with NPMs, selecting a set of benchmark students’ responses for coder training from each country (evidencing the full range in response quality for each performance task);

e. CAE will work with the NPMs to identify the person who will serve as the Chief Reader in each participating country (coding trainers for each testing language).

f. CAE, in conjunction with the NPM and Chief Reader will conduct a training session in each of the participating countries in a central location in each country identified by and secured by the NPM.

g. CAE will enable each scorer on the internet platform for language-specific scoring of responses.

6. **Analysis of results:** ACER will be the key organization involved in the analysis and documentation of the outcomes of the AHELO Feasibility Study, which involves:

a. establishing a scaling methodology for assessing scaling characteristics of the test data, based on the item response theory (IRT) and also using other classical procedures;

b. Analysis of test items, which follows the cleaning and scaling of the data. Reports presented at NPM and Expert Group meetings will include these results:

i. item statistics (IRT and classical item statistics);

ii. statistically undesirable characteristics in the items;

iii. test targeting and dimensionality;

iv. item-by-institution and item-by-language interactions; and

v. analysis of non-systematic missing data.

c. Analysis of context items which will involve:

i. review of percentages across categories to provide information about the possible skew of items and the amount of missing responses;

ii. exploratory and confirmatory factory analysis to review the dimensionality of items that are designed to measure latent constructs;

iii. deriving scale scores for these constructs using the IRT partial credit model;

iv. providing psychometric reports describing validity, reliability, DIF and scale precision; and
v. a general descriptive analysis in which cognitive outcomes are disaggregated by major student groupings and institute characteristics within country and subject area.

7. **Data products and written reports:** To support data analysis and assist an effective international dissemination of the results from the AHELO Feasibility Study, the partners involved in Module E will provide the following products and services:
   a. AHELO Feasibility Study database and codebooks;
   b. AHELO Feasibility Study compendia;
   c. AHELO Feasibility Study Technical Report;
   d. AHELO Feasibility Study Institution Reports / templates; and
   e. AHELO Feasibility Study Report.

**Schedules, milestones and deliverables**

1. **Schedule:** The schedule for AHELO implementation is shown in Figure 8.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Coordination</td>
<td>Quarterly progress report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual budgetary update</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Support NPMs and Ics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality assurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meetings</td>
<td>AHELO GNE</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical Advisory Group</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economics Expert Group</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Engineering Expert Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generic Skills NPM</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economics NPM</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engineering NPM</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>Develop population definitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Develop sampling plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Support institutional engagement</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undertake/oversee student sampling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oversee leader and faculty sampling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operations</td>
<td>Prepare test administrations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coordinate coding operations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Develop support manuals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Run training workshops</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monitor quality and procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country data capture and verification</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data verification and processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td>All countries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module A</td>
<td>Proctor training for NPM/IC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Final PP to ITS to load on testing platform</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coordinate and support IT during testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide support to NPM during testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Score training during testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scoring of student responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data analysis and reporting by CAE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Draft support and technical materials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Psychometric analyses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Statistical analyses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evaluate feasibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting</td>
<td>Prepare data products</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prepare AHELO Technical Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prepare Institution Report template</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prepare AHELO Feasibility Study Report</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 8: Implementation development schedule*
2. **Deliverables**: ACER and partner organisations will deliver:

   a. the faculty and student sampling plans for the AHELO Feasibility Study, including the sampling manual for NPMs and institutional coordinators;

   b. documentation on survey procedures and quality assurance procedures, including:
      i. test administration procedures, including technical support in case of electronic delivery;
      ii. technical standards;
      iii. coding guides (obtained from the teams leading Modules A, B, C and D); and
      iv. data management manual;

   c. cleaned databases, including results for the assessment instruments and context surveys responses with the associated documented data product; and

   d. analyses, tables and technical documentation to provide support for the OECD Secretariat in preparing the final report from the AHELO Feasibility Study.

   e. and following management deliverables:
      i. a quarterly report on progress against the agreed project timetable; and
      ii. an annual budgetary update including the distribution of resources (physical and human) by task will be prepared and submitted to the OECD Secretariat no later than 30 January covering the previous calendar year.

**Roles and responsibilities**

66. The core staff members (and organizational affiliations) involved in Module E will include:

   - **Australian Council for Educational Research (ACER):**
     - Hamish Coates – AHELO Project Director;
     - Sarah Richardson – project management;
     - Alisdair Daws – software development, IT management;
     - Siek Toon Khoo – technical advice, quality assurance; and
     - Tim Friedman – technical support.

   - **Council for Aid to Education (CAE):**
     - Roger Benjamin – module director;
     - Richard Shavelson – technical leader;
     - Jim Hundley – director of operations;
     - Stephen Klein – technical advisor;
     - Scott Elliot – testing operations coordinator; and
     - Robert Keeley – technical support and planning for scoring.

   - **cApStAn:**
Andrea Ferrari – definition of procedures;
Steve Dept – operations director; and
Raphaël Choppinet – assistant project manager.

• Educational Testing Service (ETS):
  Thomas Van Essen – Director, Module B;
  Claire Melican – content director;
  Rick Morgan – psychometric support and instrument design; and
  Rae Jean Braunmuller Goodman (Professor of Economics, US Naval Academy) – consultation on framework and item development.

• IEA Data Processing and Research Center (IEA DPC):
  Dirk Hastedt – coordination; and
  Falk Brese – survey operations, data coordination.

• SoNET Systems:
  Mike Janic – director of software development; and
  Stephen Birchall – software development manager;

• Statistics Canada:
  Jean Dumais – sample design and quality control; and
  Isabelle Michaud and Sylvie LaRoche – sampling coordination, verification.

67. The AHELO Technical Advisory Group (TAG) includes the following experts:

• Dr Peter Ewell, United States (Chair)
• Professor Vaneeta D’Andrea, United Kingdom
• Professor Paul Holland, United States
• Dr Motohisa Kaneko, Japan
• Professor Lynn Meek, Australia
• Dr Keith Rust, United States
• Professor Frans Van Vught, Netherlands
• Professor Robert Wagenaar, Netherlands