

Unclassified

EDU/IMHE/AHELO/GNE(2011)19/ANN6/FINAL

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

20-Apr-2012

English - Or. English

DIRECTORATE FOR EDUCATION
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD

Group of National Experts on the AHELO Feasibility Study

ENGINEERING ASSESSEMENT DEVELOPMENT REPORT

AHELO Feasibility Study

8th meeting of the AHELO GNE
Paris, 28-29 November 2011

This document was prepared by the ACER Consortium.

It is a finalised version of the document presented at the 8th meeting of the AHELO GNE and is declassified with the GNE and Consortium's approval.

It is only available in PDF format.

Contact:

Consortium: ahelo@acer.edu.au

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

JT03320213

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(2011)19/ANN6/FINAL
Unclassified

English - Or. English



Engineering Assessment Development Report

January 2012



CONTENTS

| | |
|---|----|
| CONTENTS..... | 2 |
| INTRODUCTION | 3 |
| Overview | 3 |
| Assessment scope and format..... | 3 |
| Assessment development timeline | 5 |
| ENGINEERING ASSESSMENT FRAMEWORK | 7 |
| Overview | 7 |
| The competencies | 7 |
| DEVELOPMENT OF CONSTRUCTED RESPONSE TASKS | 10 |
| Overview | 10 |
| Initial preparation | 10 |
| Panelling of draft assessment materials..... | 11 |
| Development of scoring guide | 12 |
| Framework mapping..... | 13 |
| Focus groups..... | 13 |
| Expert group meeting | 14 |
| PHASE 1 Final versions..... | 14 |
| Obstacles and resolutions | 14 |
| DEVELOPMENT OF MULTIPLE CHOICE ITEMS | 15 |
| Overview | 15 |
| Expert Group meeting | 15 |
| Panelling..... | 15 |
| Framework mapping..... | 15 |
| Development of scoring guides..... | 16 |
| Obstacles and resolutions | 16 |
| PHASE 1: QUALITATIVE TESTING | 17 |
| Translation and adaptation | 17 |
| Focus groups..... | 17 |
| REVISION OF ENGINEERING ASSESSMENT | 19 |
| PHASE 1 feedback from countries..... | 19 |
| Review and revision process | 19 |
| Revised assessment instrument | 20 |
| Online testing system | 21 |
| Revised scoring guide | 21 |
| Obstacles and resolutions | 21 |
| PHASE 2: AHELO IMPLEMENTATION..... | 23 |
| Overview | 23 |
| APPENDIX A: ENGINEERING EXPERT GROUP | 24 |
| APPENDIX A: ENGINEERING EXPERT GROUP | 24 |

INTRODUCTION

Overview

1. One of the disciplines selected for the OECD's Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study focuses on the development of an assessment in civil engineering (hereafter, the 'Engineering Assessment'). This assessment is designed to measure the skills and knowledge of final-year first-cycle (bachelor degree) university students and to be suitable for translation and use in a range of countries and languages.
2. Development of the Engineering Assessment Framework and Engineering Assessment took place between July and December 2010, while revision of the Engineering Assessment post Phase 1 testing took place between July 2011 and January 2012. This report details their evolution. The Engineering Assessment Framework defines the domain to be tested and specifies the expected learning outcomes for students in the target group. The Engineering Assessment comprises both constructed response tasks and multiple-choice items and is designed for completion in 90 minutes.
3. The development of both the Engineering Assessment Framework and Engineering Assessment was undertaken by a consortium of organisations in Australia, Japan and Italy, led by the Australian Council for Educational Research (ACER), and incorporated the expertise of engineering educators and specialists from around the world.
4. Sufficient materials were developed to enable the rotation of tasks during Phase 1 (development and qualitative validation) of AHELO with final selections of materials to be informed by input from university students and faculty in participating countries in preparation for Phase 2 (implementation).

Assessment scope and format

5. Civil engineering is being tested in Module C of AHELO. The duration of the Engineering Assessment is 90 minutes. The Assessment includes a broad sample of items covering a range of difficulty that will enable the strengths and weaknesses of populations and key subgroups to be determined with respect to the components of engineering competency.
6. Two types of assessment tasks have been developed. The first type of task has a constructed response format. Constructed response tasks have been designed to comprehensively assess four key competencies defined in the Engineering Assessment Framework: engineering Generic Skills, Analysis, Design, and Practice.
7. The constructed response tasks introduce an authentic engineering scenario structure, design, situation or problem in a specific context and present students with a set of items related to that context. The scenario is introduced through the use of a range of stimuli including photographs, diagrams, tables and charts and students then respond to a number of items requiring both short answers and longer responses.
8. In Phase 1 of AHELO, four constructed response tasks were included and these were rotated through focus groups, to enable data to be gathered from students to inform the selection of two tasks for Phase 2. For Phase 2, three constructed response tasks were selected.
9. The second type of task has a multiple choice format. Multiple choice items have been designed to assess the fifth key competency defined in the Engineering Assessment Framework – Basic and Engineering Sciences. They have been included to provide a fast and efficient way to collect data on students' engineering knowledge, understanding and skills and to compliment the

constructed response tasks. They cover a wide range of basic engineering knowledge, along with specific above content competencies. The multiple choice items are also included as a means of verifying the robustness of competencies assessed by the constructed response tasks. Since application of basic engineering and scientific principles requires their mastery, results from multiple choice items should indicate whether students have in fact developed the fundamentals that underlie competencies required to analyse and synthesise solutions to complex engineering problems.

10. To respond to multiple choice items, students need to select one correct response out of four possibilities. In total, 30 multiple choice items have been selected, (from the initial 40 that were developed), grouped in six sets of five. They are designed so that students can respond to 25 items in 50 minutes, and will be rotated in Phase 2 of AHELO, with students completing five of the five-item sets.
11. Both constructed response tasks and multiple choice items have been designed to be suitable for either paper-based or computer delivery. In Phase 1 of the AHELO Feasibility Study, students were presented with a paper version of the test and were required to hand-write their responses to all items on the assessment form. This approach was taken as the assessment materials were still in a developmental stage. In Phase 2 of AHELO, students will respond to both constructed-response tasks and multiple choice items through computer delivery.
12. There are a number of rotations for Phase 2 of the Engineering Assessment, as indicated in **Error! Reference source not found..** Data can be obtained from students' responses to all the materials, however, across a given institution.
- 13.

Table 1: Phase 2 Engineering Assessment rotations

| Rotation | Constructed response task | Multiple choice set | | | | |
|----------|---------------------------|---------------------|--------|--------|--------|--------|
| 1 | CRTM1 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 |
| 2 | CRTM1 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 |
| 3 | CRTM1 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 |
| 4 | CRTM1 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 |
| 5 | CRTM1 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 |
| 6 | CRTM1 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 |
| 7 | CRTM2 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 |
| 8 | CRTM2 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 |
| 9 | CRTM2 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 |
| 10 | CRTM2 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 |
| 11 | CRTM2 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 |
| 12 | CRTM2 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 |
| 13 | CRTM3 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 |
| 14 | CRTM3 | ENGMC2 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 |
| 15 | CRTM3 | ENGMC3 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 |
| 16 | CRTM3 | ENGMC4 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 |
| 17 | CRTM3 | ENGMC5 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 |
| 18 | CRTM3 | ENGMC6 | ENGMC1 | ENGMC2 | ENGMC3 | ENGMC4 |

Assessment development team

14. Experience from the development of other international assessment materials demonstrated the value of involving diverse international teams in the creation of the Engineering Assessment for AHELO. This was to ensure that a conceptually rigorous assessment was created which is applicable in a range of cultural and linguistic contexts. Consequently, an international team worked together to develop the Engineering Assessment.

15. Development of the Engineering Assessment was led by the Australian Council for Educational Research (ACER), working in close consultation with Japan's National Institute for Educational Policy Research (NIER), and the University of Florence's School of Engineering in Italy (typically referred to in this report as the 'development team'). The development team worked closely with Associate Professor Roger Hadgraft of the School of Engineering at the University of Melbourne, a civil engineer with more than 15 years involvement in improving engineering education and past President of the Australasian Association for Engineering Education (AAEE). Professor Kikuo Kishimoto of the Tokyo Institute of Technology was instrumental in gaining support from the Institution of Professional Engineers Japan and the Japan Society of Civil Engineers, allowing the item development team to reference their licensing examinations, and to build upon the expertise accumulated by these institutions. Professor Claudio Borri of the School of Engineering at the University of Florence played a key role in framework and item development and was able to draw on the European and Global Engineering Education Academic Network (EUGENE), representing 76 international partners from 32 countries.
16. An Engineering Expert Group drawn from participating countries and key international organisations was also heavily involved with item validation and revision and in supporting the development team. The Engineering Expert Group was chaired by Emeritus Professor Robin King and included members from Australia, Germany, Japan, Sweden and the United States. A full list of members is found in Appendix A.

Assessment development timeline

17. Development of the Engineering Assessment Framework and Engineering Assessment took place between July and December 2010. The framework was initially drafted in July and August 2010 and was then circulated to the development team and Engineering Expert Group for feedback, with revisions made based on advice received. The constructed response tasks were created between July and August 2010 and subjected to a rigorous development and validation process. Multiple choice items were translated from Japanese in August 2010 and the English versions were circulated to members of the consortium and the Expert Group in September 2010.
18. A face-to-face meeting of the Engineering Expert Group was held in Singapore in October 2010. The purpose of this meeting was to:
 - a. provide a final review of the Engineering Assessment Framework;
 - b. select a short-list of constructed response tasks for use in Phase 1 of AHELO;
 - c. recommend revisions to selected constructed response tasks;
 - d. select those multiple choice items most suitable for use in Phase 1 of AHELO; and
 - e. recommend revisions to selected multiple choice items.
19. After the meeting of the Engineering Expert Group, all recommended revisions were made to the assessment framework and assessment materials, with these circulated to members of the Expert Group for final approval. The Engineering Assessment Framework and Engineering Assessment were also provided to a small group of members of the EUGENE network for feedback. Once approved, the Engineering Assessment Framework and Engineering Assessment tasks were provided to participating countries for review and feedback.
20. Overall, the timeline for the development of the Engineering Assessment was as indicated in Table 2.

Table 2: Engineering Assessment development timeline

| Activity | Period |
|---|-------------------------------|
| Initial framework development | January 2010 – March 2010 |
| Framework development by ACER-led team | July 2010 – August 2010 |
| Item development | July 2010 – November 2010 |
| Expert group meeting | October 2010 |
| Finalisation of assessment materials | October 2010 – December 2010 |
| Distribution to countries for feedback | October 2010 – December 2010 |
| Distribution of final source versions (Phase 1) | December 2010 |
| Translation into national languages | January 2011 – March 2011 |
| Qualitative testing in participating countries | March 2011 – May 2011 |
| Review and revision of assessment | July 2011 – September 2011 |
| Distribution of final source versions (Phase 2) | October 2011 |
| Translation into national languages | October 2011 – November 2011 |
| Finalisation in online testing system | December 2011 |
| Revision and distribution of scoring guide | September 2011 – January 2012 |
| Quantitative testing in participating countries | March 2012 – May 2012 |

ENGINEERING ASSESSMENT FRAMEWORK

Overview

21. AHELO's Engineering Assessment Framework was the guiding document during instrument development. Materials were developed in direct consultation with the key competencies, which are explicated in more detail in the framework itself.
22. The Engineering Assessment Framework was based on the AHELO-Tuning document (OECD 2009a), the AHELO Engineering Assessment workshop held at ACER in Melbourne in January 2010, the TECA document (Coates & Radloff, 2008), and broader AHELO technical materials. It was informed by the processes and practices adopted in the PISA literacy surveys (e.g. OECD 2009b), and the combined expertise of consortium staff. Subsequent drafts incorporated review comments from consortium members and Engineering Expert Group members.
23. The Engineering Assessment Framework defines the domain to be tested, specifically: first-cycle engineering competency is the demonstrated capacity to solve problems by applying basic engineering and scientific principles, engineering processes and generic skills. It includes the willingness to engage with such problems in order to improve the quality of life, address social needs, and improve the competitiveness and commercial success of society.

The competencies

24. An assessment instrument must tap into the different aspects of a test taker's proficiencies. Engineering competency entails applying relevant skills and knowledge in solving problems of interest to an engineer. Recognising that Engineering problems occur in a diverse array of situations, a representative sample of engaging contexts for items has been chosen to exercise the constituent components of Engineering competency.
25. The contexts in which students need to demonstrate their skills and knowledge include both those specific to civil engineering and those more generally applicable across a number of fields of engineering. Competencies are activated when a student tackles an assessment item which focuses on a specific aspect of engineering and competencies can be demonstrated in the student responses.
26. An overview of the key components is given in Figure 1: AHELO Engineering Assessment Framework: Key Components. This provides a summary of how the domain is organised, showing the elements of importance for the assessment of first-cycle engineering competency.

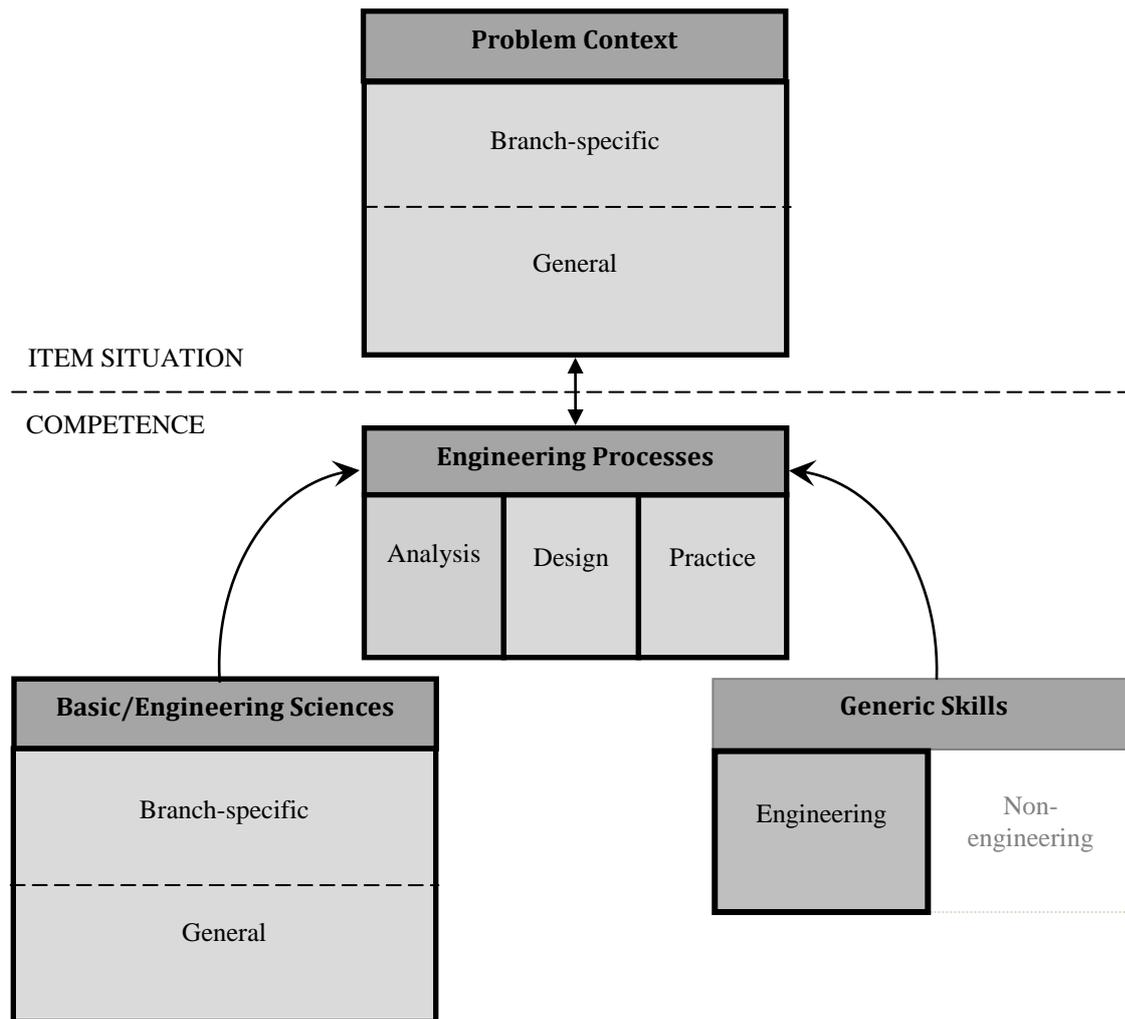


Figure 1: AHELO Engineering Assessment Framework: Key Components

27. Each of the key components can be broken down into a number of sub-components, as indicated below.

- a. Engineering Generic Skills
 - i. The ability to use diverse methods to communicate effectively with the engineering community and with society at large;
 - ii. The ability to demonstrate awareness of the wider multidisciplinary context of engineering.
- b. Basic and Engineering Sciences
 - i. The ability to demonstrate knowledge and understanding of the scientific and mathematical principles underlying their branch of engineering;
 - ii. The ability to demonstrate a systematic understanding of the key aspects and concepts of their branch of engineering;
 - iii. The ability to demonstrate comprehensive knowledge of their branch of engineering including emerging issues. For civil engineering in this assessment, this comprises the following five specialised areas: materials and construction (BES3(i)), structural engineering (BES3(ii)), geotechnical engineering (BES3(iii)), hydraulic engineering (BES3(iv)), and urban and rural planning (BES3(v)).
- c. Engineering Analysis

- i. The ability to apply knowledge and understanding to identify, formulate and solve engineering problems using established methods;
 - ii. The ability to apply knowledge and understanding to analyse engineering products, processes and methods;
 - iii. The ability to select and apply relevant analytic and modelling methods;
 - iv. The ability to conduct searches of literature, and to use databases and other sources of information;
 - v. The ability to design and conduct appropriate experiments, interpret the data and draw conclusions;
 - vi. The ability to demonstrate workshop and laboratory skills.
- d. Engineering Design
- i. The ability to apply knowledge and understanding to develop designs to meet defined and specified requirements;
 - ii. The ability to demonstrate an understanding of design methodologies, and an ability to use them.
- e. Engineering Practice
- i. The ability to select and use appropriate materials, equipment and tools;
 - ii. The ability to combine theory and practice to solve engineering problems;
 - iii. The ability to demonstrate understanding of applicable techniques and methods, and their limitations;
 - iv. The ability to demonstrate understanding of the non-technical implications of engineering practice and commitment to professional ethics, responsibilities and norms of engineering practice;
 - v. The ability to demonstrate understanding of the health, safety and legal issues and responsibilities of engineering practice, the impact of engineering solutions in a global, economic, societal and environmental context;
 - vi. The ability to demonstrate knowledge of project management and business practices, such as risk and change management, and be aware of their limitations.
28. As far as is practicable, each item in the Engineering Assessment focuses on a single component of competency. Accordingly, some items assess understanding of mathematics and science, some of engineering sciences, others assess methods of engineering analysis, and others the ability to realise engineering designs. In others, the capacity to apply engineering knowledge and understanding to realistic problems is assessed.
29. Table 3: Distribution of score points in the AHELO Engineering Assessment indicates the recommended distribution of score points across the domain components. The Basic and Engineering Sciences competency is primarily assessed in the two multiple choice clusters. The three process components – Analysis, Design and Practice – are assessed in the constructed response tasks. Each process has been given equal weight. A small number of score points have been allocated to Engineering Generic Skills. Given the relatively limited number of items, the figures in the table, and the percentages, are approximate only. The rows represent the different available response formats: multiple choice items (MC) and constructed response tasks (CR).

Table 3: Distribution of score points in the AHELO Engineering Assessment

| | Engineering Generic Skills | Basic/ Engineering Sciences | Analysis | Design | Practice | Weighting |
|------------------|-----------------------------------|------------------------------------|-----------------|---------------|-----------------|------------------|
| Weighting | 10% | 45% | 15% | 15% | 15% | 100% |

DEVELOPMENT OF CONSTRUCTED RESPONSE TASKS

Overview

30. The constructed response tasks (CR) have been designed to be coherent sets of items which centre on authentic engineering contexts. The contexts offer students an engineering structure, design, situation or problem. The CRs have been developed with an aim to engage the students with interesting, innovative, real-world situations that arise in the profession of civil engineering. A wide range of civil engineering contexts are offered.
31. Twelve discrete CRs were initially developed, each undergoing a comprehensive process of collaborative and iterative development to ensure that assessment materials were valid and would be reliable both at the time of development and into the future.
32. Two significant quality control features of this approach are that it provides opportunities for assessment materials to be reviewed and revised at every stage of development and that substantive specialists and project stakeholders can make meaningful contributions to the review and refinement of the assessment instrument.
33. The choice of CRs was informed by extensive research into civil engineering contexts. Each CR was developed and then reviewed by all consortium partners, with feedback incorporated during revision. The CRs were developed to ensure a complete mapping to the framework. That is, they were designed so that all competencies in the framework could be covered evenly if several combinations of CRs were selected. The CRs aimed to comprehensively assess engineering generic skills, engineering analysis, engineering design, and engineering practice.

Initial preparation

34. Wide-ranging research into civil engineering contexts was conducted by ACER. This research resulted in the development of the 12 initial CRs. Methods for this investigation were varied and included surveys of:
 - a. civil engineering textbooks;
 - b. civil engineering exam papers;
 - c. civil engineering concept inventories;
 - d. international civil engineering professional assessments (such as certification examinations); and
 - e. civil engineering projects worldwide.
35. Possible contexts and materials were also submitted to ACER by consortium partners and stakeholders, following the AHELO Engineering Assessment Workshop held at ACER in Melbourne in January 2010. ACER developed the materials in consultation with engineering expert Associate Professor Roger Hadgraft (The University of Melbourne).
36. Care was taken to ensure a range of contexts were employed as one means of controlling for students' interests and prior knowledge. Real world situations are often extremely complex and a balance needed to be struck when constructing items between authenticity of a context and practicality of assessment.
37. It was essential to develop the CRs in line with the 'above content' thrust of the framework and the entire AHELO ethos. The aim was not to see whether final-year students could sit another

university engineering exam. Rather, the aim was to assess their ability to think like an engineer; to display the non-technical competencies that practising engineers must possess.

38. To ensure divergent validity, it was important to ensure that the CRs were not easily accessible to anyone with generic competencies. That is, an individual with high-level critical thinking skills may be able to engage with above content material if it is developed without a sufficient basis founded on discipline specific content. Thus, all the CRs had to contain engineering subject matter and technical information, even if technical skills were not assessed in the CRs.
39. Language difficulty was set at an appropriate level for final-year first-cycle engineering students. Photographs and diagrams are used where appropriate to avoid excessively long passages of text.
40. The 12 CRs developed by assessment developers at ACER, were: 'wind', 'dams', 'freeway', 'Falkirk Wheel', 'bridges', 'flood prevention', 'storm barriers', 'Toshka Project', 'train technology', 'concrete', 'timber' and 'islands'.
41. Each CR underwent a comprehensive process of iterative development as detailed in the following sections.

Panelling of draft assessment materials

42. Once assessment developers felt that a CR was complete, they submitted it to the process of 'panelling'. Panelling is a team approach to reviewing assessment materials. It is a rigorous, robust quality control mechanism employed in the development of assessment materials worldwide. The practice is based on the recognition of the importance of exposing material to multiple viewpoints.
43. Members of the panel for AHELO included the test developer, the project director and additional assessment experts. Prior to a panelling session, members of the panel received draft CRs and attempted the items as though they were test-takers. During the panel session, participants compared their answers to the questions, and raised issues about the questions and the material. It was a robust process of debate which aimed to ensure that items performed their intended function, were unambiguous and could withstand a range of challenges.
44. ACER has a large number of specialist assessment developers who are able to bring a broad range of perspectives and expert opinion to this review process. Up to six individuals participated in each panelling session to ensure that a range of viewpoints could be accessed during the panel.
45. The following questions were used during panelling sessions for AHELO Engineering Assessment materials:
 - a. Content validity
 - i. How does the material relate to the assessment specifications?
 - ii. Do the questions test the assessment framework?
 - iii. Do the questions relate to the essence of the stimulus or do they focus on trivial side issues?
 - iv. How will this material stand up to public scrutiny (including project stakeholders and the wider community)?
 - b. Clarity and context
 - i. Is it coherent? Unambiguous? Clear?
 - ii. Is the material interesting? Is it worthwhile? Of some importance?
 - iii. Is it self-contained? Or does it assume other prior knowledge, and if so is this appropriate?
 - iv. Is the reading load as low as possible?

- v. Are there any ‘tricks’ in the question that should be removed?
 - vi. When a unit includes more than one item, are there dependencies between the items? For example: Does one item give a clue to the next one? Would a different order of items within a unit make a difference? If a response to one item is incorrect, does this affect possible responses for other items in the unit?
- c. Format
- i. Is the proposed format the most suitable one for the item?
 - ii. Is the key (the correct answer to a multiple choice question) indisputably correct?
 - iii. Are the distractors (the incorrect options to a multiple choice question) plausible but indisputably incorrect?
- d. Test takers
- i. How will the test-takers perceive this material? To answer this, panel members must imagine the cognitive, cultural and response format demands of the items from the point of view of test-takers.
 - ii. Is it at the right level, both in terms of the expected ability level, age or school year-level(s) of the test-takers?
 - iii. Does the material breach any ethical, cultural or other sensitivities?
 - iv. Is it likely to be biased, i.e. is it likely to be easier or harder for certain subgroups in the assessment population for reasons other than differences in the ability being measured?
 - v. Is it clear what would constitute an answer to the question? That is, will test-takers know exactly what they are being asked to produce (as distinguished from knowing how)?
- e. Scoring
- i. Is the proposed scoring consistent with the underlying ability described by the assessment domain? Would test-takers possessing more of the underlying ability always score better than test-takers with less?
 - ii. Is the chosen marking scheme the most suitable one for the purpose?
 - iii. Are there other kinds of answer that have not been catered for in the marking guide (e.g. any that do not fall within the ‘correct’ answer category description, but seem to be correct)?
 - iv. Are there different approaches to arriving at the same answer? Do these different approaches represent equivalent or different levels of proficiency?
 - v. Should partial credit be given if part of the answer is achieved?
 - vi. Are the scoring criteria practicable for the markers, or are they excessively cumbersome? Are the different levels of performance clearly distinguishable?
46. After panelling, the CRs were modified in response to the panel recommendations. The revised CRs were then shown to at least one expert assessment developer who had not previously seen the materials to check that the revision had not created new unforeseen technical problems with the material.
47. The CRs were then reviewed by the development team and Engineering Expert Group. Feedback was incorporated into each one as they were revised.

Development of scoring guide

48. The CRs used in AHELO Engineering Assessment require students to produce a range of responses, including completing short engineering designs, describing analytic processes, evaluating and making use of complex data to make recommendations, or suggesting solutions to

engineering problems. This approach is used in order for students to be able to demonstrate innovative thinking, something that would not have been possible with a multiple choice format.

49. Comprehensive scoring rubrics were constructed to evaluate the responses of students to each of the items within each CR. The rubrics were constructed on the basis of the components of engineering competency outlined in the framework, with criteria for achievement based on the specific constituent learning outcomes of each component. For example, for engineering analysis one criterion is “selects and applies relevant analytic and modelling methods”. The rubrics allows the recognition of different levels of attainment in the test-takers’ work for each relevant criterion.
50. As such, the constructed response nature of the tasks does not mean that candidates have absolute freedom to answer in any fashion they desire. Instead, the scoring rubrics identify the key ideas which must be conveyed for a student to receive different score points. Acknowledging that there are multiple ways of expressing the same idea, example responses are given for every item.
51. Scorers are required to make a judgment as to whether a given response matches the key ideas given in the scoring rubrics sufficiently for a score point to be awarded. This approach is used across a wide range of international assessment instruments, and with sufficient marker training and the establishment of marker support networks, runs efficiently and effectively.
52. In this instance, the scoring rubrics were developed concurrently through the process of iterative development used to develop the CRs. Items were designed to measure varying levels of proficiency. Less demanding items were designed to measure the kind of competence which is generally associated with reproduction. Higher levels of proficiency are measured by items that assess the extent to which individuals make connections between different aspects of knowledge and skill. Higher-order reflective forms of reasoning are assessed by the most demanding items. Across the assessment, items tap into different levels of student performance.

Framework mapping

53. The CRs were developed to ensure a complete mapping to the framework. That is, they were designed so that all competencies in the framework could be covered evenly if several combinations of CRs were selected.
54. The CRs comprehensively assess all competencies in the framework, excluding Basic and Engineering Sciences which are instead covered by the multiple choice items. Thus, the CRs assess the competencies of Engineering Generic Skills, Engineering Analysis, Engineering Design, and Engineering Practice.
55. Wherever possible, items within the CRs were developed to map to a single competency. It was later deemed, however, that several items mapped to multiple competencies in the framework. A detailed mapping is provided in a separate document.

Focus groups

56. All 12 CRs were piloted at the University of Melbourne in September 2010 during focus groups with final year undergraduate civil engineering students. In each of the four focus groups, students were asked to complete three CRs within 60 minutes. A further 60 minutes was then dedicated to discussion and feedback. This process provided a chance to gain some initial data from current students. The aim was not to produce comprehensive psychometric data but rather to gauge the way in which students actually engaged with the CRs and thus to inform assessment development. In addition to providing useful feedback during discussion, each participating student was asked to fill in a questionnaire which asked certain questions regarding the CRs.

57. The students participating in these focus groups provided an overwhelmingly positive response in the feedback questionnaire and during the group discussions. They found the CRs to be authentic, practical, clear, interesting and engaging.

Expert group meeting

58. All twelve CRs were uploaded to the AHELO Exchange, providing the Expert Group with the opportunity to review the CRs and to provide feedback. This feedback was used to further refine the CRs, before the Expert Group meeting.
59. The twelve CRs were presented at the Expert Group meeting in Singapore on 22 and 23 October 2010. The Expert Group selected their preferred four CRs from the twelve initially developed. These four were deemed to be suitable for possible inclusion in the final assessment form. All four required some further development and revision.
60. Once modifications had been made to these four CRs, the Module C Consortium and the Expert Group were given a final opportunity to comment on the CRs. Overall, the Consortium and the Expert Group were content with these final versions and approved them for use in Phase 1 of AHELO.

PHASE 1 Final versions

61. A final four CRs were selected for possible inclusion in the Engineering Assessment. To produce a comprehensive mapping to the framework, it was recommended that the final assessment (for Phase 2) include either the first two or the second two CRs.

Obstacles and resolutions

62. Given the need for CRs to assess the above-content learning of final-year civil engineering students worldwide, problems which arose from cultural specificities have been carefully removed from the draft instrument. That is, the CRs have been written in such a way that students from certain countries will not be disadvantaged due to cultural factors.
63. Due to variations in civil engineering programs across different countries it was important to uncover areas of essential commonality in different programs so as not to favour certain candidates over others. The emphasis on above content assessment, however, means that the CRs were developed with an eye to assessing the engineering competencies and capacities of the student, rather than their content knowledge.
64. There was concern that students in some countries would find certain contexts difficult due to familiarity issues, and that this would affect their results. For instance, there was concern that evaporation rates from irrigation channels in a desert would be too unfamiliar for Swedish students and this led to the omission of the Toshka Project CR. Similarly, the Freeways CR was omitted due to perceptions of cultural relativity.
65. A further issue related to time. After the focus groups and expert feedback, it was deemed that 20 minutes was not sufficient for a student to complete each CR. Many students struggled to complete the CRs in the designated time during the initial focus groups. A decision was therefore made to increase the time allocation to 30 minutes per CR.
66. As the CRs themselves present authentic engineering contexts, issues may arise with copyright. It is important to ensure that materials used in the CRs are available to be used without restrictions. It may be necessary to gain permission to use data and images, especially when these are derived from official documents or reports.

DEVELOPMENT OF MULTIPLE CHOICE ITEMS

Overview

67. The multiple choice (MC) items, were designed as a set of items which prompt students to demonstrate their competency in Basic and Engineering Sciences. They have been included to provide a fast and efficient way to collect data on students' engineering knowledge, understanding and skills. They complement the CRs in providing an instrument which covers a wide range of basic engineering knowledge, along with specific above content competencies. They also verify the robustness of competencies assessed by the CRs. Since application of basic engineering and scientific principles requires their mastery, results from multiple choice items should indicate whether students have in fact developed the fundamentals that underlie competencies required to analyse and synthesise solutions to complex engineering problems.
68. Development of the multiple choice items began with licensing examinations developed by the Institution of Professional Engineers Japan and the Japan Society of Civil Engineers. An extensive list of translated items from this source was presented by NIER to the Engineering Expert Group for their review. Forty items were selected and revisions and further developments were advised, with Engineering Expert Group members approving final versions. Four clusters of MC items were developed. Each cluster consists of 10 MC items and students will be presented with two clusters – 20 MC items – which together require 30 minutes to complete.

Expert Group meeting

69. ACER produced a preliminary item analysis of the NIER material for discussion at the Expert Group meeting on 22 and 23 October 2010. ACER staff analysed all material submitted by NIER and selected possible items for inclusion.
70. The Expert Group selected several other items for inclusion at the Expert Group meeting. It was decided that approximately 40 MC items be included in the clusters.

Panelling

71. As with the CRs, ACER's specialist assessment developers submitted each MC cluster to a process of rigorous panelling (see above for a detailed description of this process).
72. The MC clusters were modified in response to the panel recommendations. As part of this process, the revised MC items were shown to at least one expert assessment developer who had not previously seen the materials to check that the revision had not created new unforeseen technical problems with the material.
73. The materials were then reviewed by the development team and the Expert Group. Feedback was incorporated into the MC clusters as they were revised. The Expert Group then approved the four MC clusters.

Framework mapping

74. The 40 MC items were mapped to the Basic and Engineering Science component of the framework. This competency was divided into the following specialised areas with the 40 MC items to be spread evenly across each area: general strand; materials and construction; structural engineering; geotechnical engineering; hydraulic engineering; urban and rural planning.

Development of scoring guides

75. Multiple choice items can provide a valid, reliable and efficient means of collecting data on students' engineering knowledge, understanding and skills. In the case of Basic and Engineering Sciences – which are assessed using items with a multiple choice format – the rubric is very simple. Thus, the scoring guide simply states the correct answer for credit to be given.

Obstacles and resolutions

76. A large proportion of the multiple choice items selected for inclusion in the MC clusters required revision to ensure that they were internationalisable. That is, it was important to maintain a-cultural contexts and content in the MC clusters, as with the CRs. If the content of a MC item was deemed to be country specific, it was removed from the pool of items. For instance, one item assessed student's knowledge of the specific engineering codes in Japan. An item such as this is not within the scope of the AHELO project.
77. There was a large range of difficulties in the MC items for possible inclusion. It was decided to spread the difficulties across the clusters. Some items were deemed to be too difficult, however, as they did not focus on basic engineering sciences, but contained a more specialised focus for a specific type of engineering student.
78. Furthermore, some items would have taken a student too much time to complete, which was not practical due to the time constraints of the AHELO assessment.
79. Several MCQs which focused on non-technical aspects of civil engineering were also removed from the pool for two reasons. First, these items mapped to areas of the framework which were covered by the CRs. Second, these items were rather easy and it was deemed that any individual with high-level critical thinking skills would have selected the correct answer.

PHASE 1: QUALITATIVE TESTING

Translation and adaptation

80. The process of translating the Engineering Assessment was managed by cApStAn – international experts in linguistic quality control. cApStAn prepared the AHELO Translation, Adaptation and Verification monitoring manual (ATAV), including guidelines for all individual items. The ATAV was sent to National Project Managers (NPMs) in January 2011.
81. During a teleconference in February 2011, cApStAn trained NPMs on how to implement translation and adaptation in their countries.
82. Translation of the Engineering Assessment took place between February and May 2011. Two independent translators in each participating country were sent the translation and adaptation guidelines. Both translators translated the Engineering Assessment, without communicating with each other. One of these translators was also asked to translate the coding guide and focus group manual.
83. Reconciliation took place between February and May 2011. Reconciliation involved a committee consisting of at least one senior translator/editor, one domain expert and one lecturer in the domain. The translators involved were not the same people as those who translated the assessment.
84. The committee of reconcilers examined the two translations of the Engineering Assessment and took the best elements of each in order to create the reconciled version. The committee also documented national adaptations from the international source in the ATAV monitoring workbook. Their records will be released together with the source version.
85. Verification took place between March and May 2011. After the reconciliation had been completed, NPMs sent the reconciled version of the instrument to cApStAn. Verification involved the reconciled translation being checked by a verifier and domain expert contracted by cApStAn. Verification feedback was also documented in the ATAV workbook.
86. At the conclusion of the verification process, NPMs decided whether to accept or reject the proposed corrections and suggestions. Once the finalised version had been determined, NPMs arranged for the materials to be piloted in focus groups.

Focus groups

87. In the AHELO Feasibility Study, focus groups were used to assist with the development of the Engineering Assessment in Phase 1. The feedback from focus groups was used to revise the Engineering Assessment to make it ready for more widespread testing of students in Phase 2 of the study.
88. The focus groups gathered data on candidates' perceptions of the instrument. The focus groups provided feedback on the Engineering Assessment, and informed the broader question of feasibility for AHELO.
89. Focus groups took place in all of the countries which participated in the feasibility study. In each country, focus groups were conducted at between five and 10 institutions with participants who volunteered to take part. During the focus groups, participants worked on printed versions of the Engineering Assessment which included a number of tasks. After the focus groups were

complete, the tasks were reviewed by the translator and assessment developers based on the information that was collected from participants.

90. During the focus groups, the four CRs and four MCQ clusters were circulated among participants. In each focus group, participants completed one (1) constructed response task (30 minutes) and two (2) sets of multiple choice questions (30 minutes).
91. The final assessment form of the Engineering Assessment will consist of two CRs and six MC clusters. During Phase 1, the four CRs selected by the Expert Group were trialled. After the focus groups were conducted, and at the completion of Phase 1, the best two CRs were retained and the other two were discarded. The Phase 2 booklet combinations were shown earlier in **Error! Reference source not found.**

REVISION OF ENGINEERING ASSESSMENT

PHASE 1 feedback from countries

92. This section discusses the feedback from the Phase 1 implementation of AHELO relevant to the CR tasks and MC items.
93. After implementing Phase 1, countries provided multiple forms of feedback to ACER for analysis and evaluation. The type of feedback varied, and included:
- reviews from in country experts;
 - comments from domain verifiers in the ATAVs;
 - feedback from students via the focus group feedback forms;
 - data on student performance.

Psychometric analysis

94. Further to these resources, ACER's psychometrics division performed a Rasch analysis on the data gathered from student responses. This analysis provided an item by item report of the assessment instrument, providing test developers with item statistics and Item Characteristic Curves (ICCs). Unfortunately, as numbers were rather low for each item (less than 200 students per item), the psychometric analysis was deemed indicative only and not a robust picture of the psychometric attributes of the items.
95. An example of the item statistics presented to the test development team is shown in figure 2 below. These statistics represent one of the MC items that performed well in Phase 1 and was thus selected for inclusion in Phase 2.

```
=====
item:2 (MC1Q02)
Cases for this item      66   Item-Rest Cor.  0.24   Item-Total Cor.  0.48
Item Threshold(s):     0.81   Weighted MNSQ  0.96
Item Delta(s):         0.81
-----
```

| Label | Score | Count | % of tot | Pt Bis | t (p) | PV1Avg:1 | PV1 SD:1 |
|-------|-------|-------|----------|--------|--------------|----------|----------|
| 1 | 1.00 | 23 | 34.85 | 0.24 | 1.95 (.055) | 0.46 | 0.58 |
| 2 | 0.00 | 20 | 30.30 | -0.05 | -0.43 (.668) | 0.11 | 0.48 |
| 3 | 0.00 | 9 | 13.64 | -0.16 | -1.34 (.186) | 0.05 | 0.53 |
| 4 | 0.00 | 6 | 9.09 | -0.08 | -0.61 (.546) | 0.02 | 0.23 |
| 9 | 0.00 | 8 | 12.12 | -0.03 | -0.24 (.809) | 0.04 | 0.50 |

```
=====
```

Figure 2: Example Item Statistics from Phase 1

96. An example of the ICCs presented to the test development team is shown in figure 3 below. These statistics represent the same MC item that performed well in Phase 1 and was thus selected for inclusion in Phase 2.

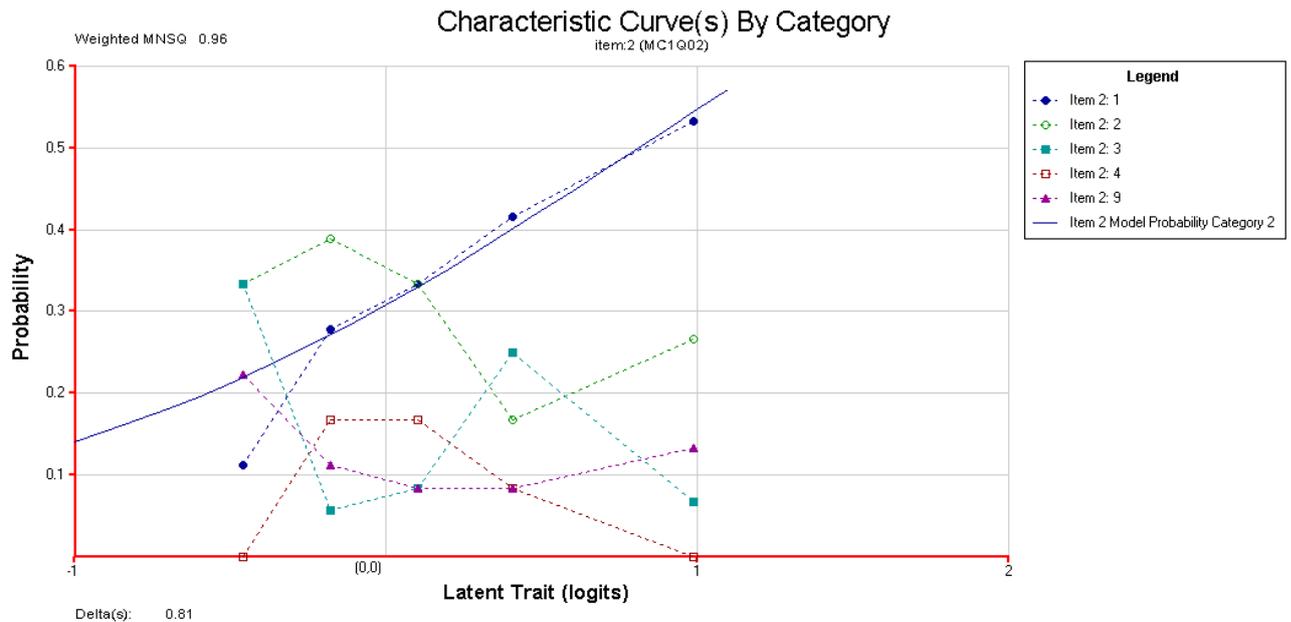


Figure 3: Example ICC from Phase 1

Review and revision process

97. In preparation for Phase 2 implementation, the assessment development team analysed the feedback received and psychometric analyses for each CR task and MC item. The comprehensive review process was performed between July and September 2011.
98. Some of the feedback from countries offered explicit suggestions for minor revisions to the assessment instrument. Other suggestions were broader, highlighting a concern regarding the wording of an item or more complex conceptual matters. It was recommended to the assessment development team that some items be removed altogether without revision due to irreconcilable problems with the item.
99. The ‘Bridges’ CR task was removed for Phase 2. The task was deemed too difficult for the target student cohort. Significant modifications would have been required and would have changed the unit to be unrecognisable from its original form.
100. The ‘Concrete’ CR task received notable modifications, and the latter part of the task incorporated some unproblematic components of the ‘Bridges’ CR task. The other two CR tasks required only minor modifications to some of the items.
101. Of the initial 40 MC items, 10 MC items were removed for Phase 2. Some of these decisions rested upon the psychometric data which indicated that the items were either too difficult or inherently problematic. The other decisions were based upon the issues noted by countries and submitted in various forms of feedback.

Revised assessment instrument

102. The revised assessment instrument was distributed to the expert group via the AHELO Exchange for final validation before being released as the Phase 2 source version.

103. This revised source version consisted of three CRs and 30 MC items. This material is to be rotated (see Table 2 earlier) so that each student attempts one CR and 25 MC items in the time allocation of 90 minutes.

104. After being released, the revised source version was subjected to the same rigorous translation and adaptation process as with the Phase 1 source version (detailed above). Once again, cApStAn facilitated this process in line with their quality assurance protocols.

Online testing system

105. For Phase 2, the test is to be administered in an online environment. cApStAn worked closely with ACER and SONET (the software engineers and coordinators of the online testing system) to upload the assessment instrument into a workable online testing system.

106. The online testing system was based on SONET's existing online testing environment, but was built specifically for the AHELO project to allow for the required functionality. After the translation and adaptation process, all national versions of the assessment instrument were prepared in the online testing system. ACER and NPMs were responsible for checking the online test forms for consistency against the translated source versions. Once the national versions were ready, they were released to NPMs to administer the test.

Revised scoring guide

107. The review process included substantial revision to the scoring guide for the CR tasks. Some of the most detailed feedback from countries was in reference to the scoring rubrics for each item in the CR tasks. There were many suggestions for correct responses not captured by the Phase 1 scoring rubrics. In line with these suggestions, the test development team performed a comprehensive analysis of the scoring rubrics. The rubrics were updated and expanded to be more encompassing to a range of valid responses, and to explicate and elucidate scoring procedures to potential contentious responses to items.

108. The revised scoring guide was presented to countries at the NPM meeting in Paris in November 2011. This session produced much fruitful discussion on the revised scoring rubrics and led to further modifications based on feedback from country experts. A final revised scoring guide was distributed to the expert group in January 2012 for final validation, before being subjected to the thorough translation, adaptation and verification process through cApStAn.

Obstacles and resolutions

109. One of the major obstacles post Phase 1 was that the student numbers were not as large as initially hoped. It is difficult to make valid conclusions from the psychometric analysis of data from student responses with small numbers. Fortunately, the student numbers were large enough to at least provide indications of how the items were functioning in the Rasch analysis. Thus, it was essential when reviewing items to interpret the statistical information carefully. As such, the data was used as a tool in the decision-making process and was not considered equivalent to what the final statistics might be. For future implementation, it would be beneficial to be able to trial items with larger cohorts of students.

110. Removing a small amount of the materials post Phase 1 was always expected. This is why four CRs and 40 MC items were initially trialled in Phase 1. Removing some of the items did not cause a problem in moving towards Phase 2.

111. A decision was made to increase the time allocation to 30 minutes per CR task for Phase 1. However, feedback still indicated those students were struggling to complete the task in this time period. Both the in-country experts and the students themselves (on the feedback forms) noted

this. Thus, for Phase 2, the number of tasks was reduced. In the online testing system, students are given 90 minutes to complete one CR task (40 minutes) and 25 MC items (50 minutes—average of 2 minutes per item).

112. As noted earlier, copyright issues are bound to arise when presenting authentic contexts. Once again, it was important to ensure that materials used in the CRs are available to be used without restrictions. Images may need to be photographed, or diagrams may need to be drawn specifically for the use in the instrument.

PHASE 2: AHELO IMPLEMENTATION

Overview

113. The ultimate objective post-Phase 1 is for the AHELO project to enter Phase 2. Phase 2 will see a full-scale online implementation of the Engineering Assessment. It will not be possible to conclude this Engineering Assessment Development Report until this quantitative phase is complete.

APPENDIX A: ENGINEERING EXPERT GROUP

The AHELO Feasibility Study's Engineering Expert Group (EEG) was drawn from participating countries and key international organisations. The EEG has supported the development of the Engineering Assessment Framework and Engineering Assessment. The work of the EEG builds on that of the Tuning-AHELO Expert Group (TAEG). Table 4 lists members of these groups, affiliations, and periods of service.

Table 4: Engineering Expert Group membership

| Expert Group member | Group | Affiliation | Service |
|--|-----------|---|---------|
| Professor Robin King (Chair) | EEG | University of Technology Sydney, Australia | 2009-12 |
| Professor Giuliano Augusti | TAEG, EEG | Universita 'La Sapienza' , Italy | 2008-12 |
| Professor Michael Hoffman | EEG | University of Ulm, Germany | 2009-12 |
| Professor Kikuo Kishimoto | TAEG, EEG | Tokyo Institute of Technology, Japan | 2008-12 |
| Professor Johan Malmqvist | EEG | Chalmers University of Technology, Sweden | 2008-12 |
| Professor Jim Melsa | EEG | Iowa State University, United States | 2008-12 |
| Professor Lueny Morell | EEG | Hewlett Packard, United States | 2009-12 |
| Professor Nobotoshi Masuda | EEG | Tokyo City University, Japan | 2009-12 |
| Professor Alfredo Soeiro | TAEG | University of Porto, Portugal | 2008-09 |
| Professor Andrew Downing | TAEG | University of South Australia, Australia | 2008-09 |
| Professor Edwin Jones Jr | TAEG | Iowa State University, United States | 2008-09 |
| Professor Eric van der Geer-Rutten-Rijswijk | TAEG | Eindhoven University of Technology, Netherlands | 2008-09 |
| Professor Francesco Maffioli | TAEG | Politecnico di Milano, Italy | 2008-09 |
| Professor Iacint Manoliu | TAEG | Technical University of Civil Engineering, Romania | 2008-09 |
| Iring Wasser | TAEG | ASIIN, Germany | 2008-09 |
| Professor Jim Birch | TAEG | Engineering Council UK, United Kingdom | 2008-09 |
| Professor Mats Hanson | TAEG | KTH- Royal Institute of Technology, Sweden | 2008-09 |
| Professor Patricia Daniels | TAEG | Seattle University, United States | 2008-09 |
| Professor Paulino Alonso | TAEG | President of the Council of Engineering Deans of Chile, Chile | 2008-09 |
| Professor Peter Wieringa | TAEG | Delft University of Technology, Netherlands | 2008-09 |
| Professor Philippe Wauters | TAEG | FEANI Secretary General, Belgium | 2008-09 |

| Expert Group member | Group | Affiliation | Service |
|----------------------------|--------------|--|----------------|
| Professor Pierre Compte | TAEG | Commission des titres d'ingénieur, France | 2008-09 |
| Professor Yinghe He | TAEG | James Cook University, Australia | 2008-09 |
| Professor Rob Best | TAEG | London South Bank University, United Kingdom | 2008-09 |
| Professor Roger Hadgraft | TAEG | University of Melbourne, Australia | 2008-09 |