

**DIRECTORATE FOR EDUCATION
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD**

Group of National Experts on the AHELO Feasibility Study

AHELO ASSESSMENT DESIGN

Paris, 25-26 October 2010

This document was prepared by the Consortium.

*The AHELO GNE is invited to
NOTE and DISCUSS this document*

Contact:
Consortium: ahelo@acer.edu.au
OECD Directorate for Education: Diane.Lalancette@oecd.org

JT03289933

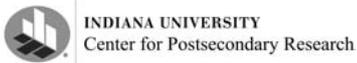
Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format





AHELO Assessment Design

September 2010



Statistics
Canada

Statistique
Canada



WESTAT
AN EMPLOYEE-OWNED
RESEARCH CORPORATION

TABLE OF CONTENTS

PREFACE: CONSORTIUM OVERVIEW	9
Introduction to this work plan	9
General features.....	11
An ethic of collaboration	11
The Consortium’s management plan.....	12
The Consortium’s guiding principles	13
New understanding	13
A picture of success	13
Internationalisation.....	13
Assessing global graduate capability	14
New perspectives on higher education.....	14
New methods	15
High-quality, efficient and meaningful methodologies.....	15
Advanced approaches to quality assurance.....	15
Enhancing the rigour of higher education research.....	15
A plan for engaging systems, institutions, faculty and students	16
Practical contexts and constraints	16
MODULE A: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE GENERIC SKILLS STRAND	18
Introduction	18
Development of the CLA Performance Tasks	18
Module management	20
References	20
MODULE B: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE ECONOMICS STRAND	21
Introduction	21
Approach to framework development.....	21
The purposes of assessment frameworks.....	21
Framework development approach.....	22
Economics assessment instrument development.....	23
General approach.....	23
Measuring ‘above content’	24
Overview	24
Innovative multiple choice items	25
Situational judgment tasks in Economics	25
Constructed response tasks	25
Scoring of constructed response items.....	26
Item validation.....	27
Pretesting.....	27
Instrument design and reliability.....	28

Documentation.....	28
Module management.....	29
References.....	29
MODULE C: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE ENGINEERING STRAND	30
Introduction.....	30
Developing the Engineering assessment framework.....	31
The purposes of assessment frameworks.....	31
Guidelines for developing and reviewing the framework	32
Framework development process	33
Audit of existing resources	33
Framework development.....	34
Test specification.....	34
A broad conceptual model for Engineering capability.....	34
Alternative assessment designs.....	34
Engineering assessment instrument development.....	35
Overview	35
Item submission process.....	35
Item types	36
Maximising the validity of items.....	37
Conceptual framework.....	37
National diversity.....	37
Cognitive rigour.....	37
Finalisation of the items	37
Coding guides and coder training.....	38
Refinement of coding guides	38
Coder training	38
Module management.....	38
References.....	39
MODULE D: DEVELOPMENT OF THE SURVEY INSTRUMENTS FOR THE CONTEXTUAL DIMENSION	41
Introduction.....	41
Context surveys framework development.....	42
Principles guiding framework development.....	42
Background reviews	43
Review of research.....	43
Audit existing resources and practices.....	43
Synthesis of existing practice.....	44
A general structure for the framework	44
Consultations and review.....	46
Documentation of the framework.....	46
Development of the context surveys	46
Overview	46
Consultative and rigorous item development	47
Defining design specifications.....	47
Mapping and drafting items	48
Item validation and review	49
Qualitative testing	49
Quantitative testing	50

Final review and documentation	51
Module management	52
References	52
MODULE E: PROJECT MANAGEMENT, SURVEY OPERATIONS AND SUPPORT FOR THE ANALYSIS OF THE AHELO FEASIBILITY STUDY RESULTS.....	56
Introduction	56
Project management	57
Project leadership	57
Lead staff and areas of responsibility	57
Management process	59
Expert Groups.....	60
National Centres	60
National Project Managers	61
Institutional Coordinators	62
Meeting arrangements	63
Assessment design and analysis plan	63
AHELO Assessment Design.....	63
AHELO Feasibility Study analysis plan.....	64
Translation, adaptation and verification	65
Approach to translation	65
Quality assurance.....	66
Quality control.....	67
Adaptations.....	68
Instrument delivery	69
Engaging institutions and sampling students	70
Population definitions.....	70
Preparation of the AHELO Feasibility Study sampling plan	71
Selecting and engaging institutions in the study	71
Sampling students	72
Sampling faculty	72
Sample monitoring and quality control	73
Sample size and participation requirements	74
Documentation.....	74
Survey procedures and operations.....	75
Test administration preparations.....	75
Coding operations.....	75
Data capture and within-country verification	76
Training of data collection staff and coders.....	77
Data management seminar	77
Coding seminar	77
Support manuals	78
Monitoring quality and survey procedures	79
Data verification and processing	80
Overview	80
Standardisation of the national file structure	80
Cleaning rules and procedures	80
Computation of weights	81
Implementation of the Generic Skills assessment.....	82
An advanced relationship	82
Development of an analysis plan.....	82

Fieldwork preparation and management.....	82
Online platform for the generic assessment.....	84
Overview.....	84
Online test delivery.....	84
Human coding.....	84
Content delivery networks – Akamai.....	84
Testing centre (institution) verification.....	84
Student identification and tracking.....	85
Administration.....	85
Analysis and reporting.....	85
Analysis of results.....	86
The scaling model for test data.....	86
Analysis of test items.....	86
Analysis of context items.....	87
Data products and written reports.....	88
AHELO Feasibility Study database and codebooks.....	88
AHELO Feasibility Study compendia.....	88
AHELO Feasibility Study Technical Report.....	89
AHELO Feasibility Study Institution Reports.....	89
AHELO Feasibility Study Report.....	90
Module management.....	91
References.....	91
ANNEXURE.....	93
MODULE A WORK PLAN.....	93
MODULE B WORK PLAN.....	93
Module Objective.....	93
Module Tasks.....	93
1. Framework development:.....	93
2. Instrument development:.....	94
3. Item validation:.....	95
4. Contribute to project management:.....	95
5. Deliverables:.....	95
6. Milestone dates:.....	95
Roles and Responsibilities.....	96
MODULE C WORK PLAN.....	97
Module Objective.....	97
Module Tasks.....	97
1. Framework development:.....	97
2. Instrument development:.....	98
3. Item validation:.....	98
4. Contribute to project management:.....	99
5. Deliverables:.....	99
6. Milestone dates:.....	99
Roles and Responsibilities.....	99
MODULE D WORK PLAN.....	101
MODULE E WORK PLAN.....	101
Module Objective.....	101

Module Tasks	101
1. Project Management:	101
2. Assessment design and analysis plan:	102
3. Translation, adaptation and verification:	103
4. Data products and written reports:	103
5. Deliverables:	103
6. Milestone dates:	104
Roles and Responsibilities	104

ABBREVIATIONS

ACER	Australian Council for Educational Research
AHELO	Assessment of Higher Education Learning Outcomes
CAE	Council for Aid to Education
CHEPS	Centre for Higher Education Policy Studies
CLA	Collegiate Learning Assessment
EDPC	Education Policy Committee
ETS	Educational Testing Service
GNE	Group of National Experts
IC	Institutional Coordinator
IEA	International Association for the Evaluation of Educational Achievement
IMHE	Institutional Management in Higher Education
NIER	National Institute for Educational Policy Research
NPM	National Project Manager
NSSE	National Survey of Student Engagement
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PT	Performance Task
TAG	Technical Advisory Group

PREFACE: CONSORTIUM OVERVIEW

Introduction to this work plan

1. This Assessment Design documents the approach that is being taken by the ACER Consortium under the overall directorship of Associate Professor Hamish Coates to design and implement the OECD's Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study. The Consortium's partner organisations include:

- Australian Council for Educational Research (ACER) – team led by Hamish Coates;
- cApStAn Linguistic Quality Control Agency – team led by Steve Dept;
- Center for Postsecondary Research (CPR) – team led by Alexander McCormick;
- Centre for Higher Education Policy Studies (CHEPS) – team led by Jon File;
- Council for Aid to Education (CAE) – team led by Richard Shavelson;
- Educational Testing Services (ETS) – team led by Thomas Van Essen;
- International Association for the Evaluation of Educational Achievement (IEA) Data Processing and Research Center (DPC) – team led by Dirk Hastedt;
- National Institute for Educational Policy Research (NIER) – team led by Shuichi Tsukahara;
- SoNET Systems – team led by Mike Janic;
- Statistics Canada – team led by Jean Dumais;
- University of Florence School of Engineering – team led by Claudio Borri; and
- Westat – team led by Frank Jenkins.

2. The Consortium affirms that the aim of the AHELO Feasibility Study is to test the science of the assessment and the practicality of implementation. Please refer to module work plans provided in annexures for detailed information about specific objectives.

3. The Consortium has formed Expert Groups for Modules B, C and E. Modules A and D are overseen by the Module E Expert Group, which serves as the AHELO Technical Advisory Group. Members of the AHELO Feasibility Study Expert Groups are listed in Table 1.

Table 1: Members of AHELO Expert Groups

<p><i>Technical Advisory Group</i></p> <p>Dr Peter Ewell, United States (Chair) Professor Vaneeta D’Andrea, United Kingdom Professor Paul Holland, United States Professor Motohisa Kaneko, Japan Professor Lynn Meek, Australia Dr Keith Rust, United States Professor Frans Van Vught, Netherlands Professor Robert Wagenaar, Netherlands</p>
<p><i>Economics Expert Group</i></p> <p>Professor Cecilia Conrad, United States (Chair) Professor William Becker, United States Professor Fiorella Kostoris, Italy Professor Maria de Lourdes Dieck-Assad, Mexico Professor Henriëtte Maassen van den Brink, The Netherlands Professor Tatsuya Sakamoto, Japan Professor Vladimir Zuev, Russia</p>
<p><i>Engineering Expert Group</i></p> <p>Professor Robin King, Australia (Chair) Professor Giuliano Augusti, Italy Professor Michael Hoffman, Germany Professor Kikuo Kishimoto, Japan Professor Johan Malmqvist, Sweden Professor Nobutoshi Masuda (Japan) Professor Jim Melsa, United States Professor Lueny Morell, United States Professor Isao Satoh, Japan</p>

4. This AHELO Assessment Design is presented in a modular format. Following this preface, separate detailed work plans are provided for Modules A, B, C, D and E.

5. The Consortium builds on the experience of previous ACER-led consortia that have had central roles in implementing several major international assessments, including all PISA cycles as well as several IEA studies. The AHELO Assessment Design adopts the best and most successful features of this related work, and combines these with a number of new and innovative solutions.

6. Read as a whole, the AHELO Assessment Design advances the ACER Consortium’s integrated approach for designing and implementing the AHELO Feasibility Study. This approach has been designed to maximise the synergies across the different strands of the AHELO Feasibility Study, streamline communications, reduce resource demands on participating countries, and generate economies of scale.

7. It is important to note that AHELO is a feasibility study and an assessment on this scale has never before been undertaken in higher education. This AHELO Assessment Design therefore provides an overview from the perspective of the start of the project. It is likely that discoveries will be made during the project that are not documented in this plan. It is also possible, although far from desirable, that the work will be impeded by unanticipated delays. Hence this plan should be considered accurate at the date of its release, which is listed on the front cover.

General features

An ethic of collaboration

8. A distinctive strength of this approach is the scope and quality of collaborative threads: collaboration among clusters of expertise, both individuals and groups; collaboration between these clusters of expertise and the Consortium; collaboration between participating countries and the Consortium; collaboration between members of the Consortium; and collaboration with institutional leaders, faculty, professional staff and students.

9. A complex international study such as AHELO will be most effective if a strong working relationship is established among the Consortium partner organisations and staff, the OECD Secretariat, and the participating countries. It will be the responsibility of the Project Director to establish a level of trust and understanding between and among the key project managers, the OECD, and key representatives from participating countries. Several things will be done to build this level of trust and understanding including formal reports, face-to-face meetings, and internet-based communications.

10. Collaboration within the four AHELO Expert Groups is of key importance in two project areas. First, the Technical Advisory Group (TAG) will provide a strong, diverse range of expert input regarding technical and higher education matters. Second, the discipline-focused Expert Groups include highly knowledgeable people in the respective areas. In addition to the within-group collaboration, these groups will also collaborate with other individual experts to give the Consortium guidance and advice on the technology-related matters that will arise during the implementation.

11. To ensure collaboration between participating countries and this effort, the Consortium will optimise opportunities for dialogue on important policy, technical and implementation matters, with both national centres and relevant national experts.

12. As advanced in the AHELO Feasibility Study Terms of Reference, the Consortium itself was formed to maximise synergies across the different strands of the AHELO Feasibility Study, streamline communications and generate economies of scale. The collaborations among members of the Consortium will ensure that all elements of the project will be handled by highly experienced individuals and organisations:

- ACER is Australia's foremost educational research organisation and has increasing international reach through its extensive work especially in the Asia-Pacific region and Europe, and through its offices in India, the Middle East, and the UK.
- cApStAn is widely recognised as the leader in the area of linguistic quality control and equivalence checks, and both its staff and verifiers bring a wealth of experience from PISA, PIAAC and other international studies to the Consortium.
- CHEPS is a leading higher education policy centre that combines basic and applied research with education, training and consultancy activities.

- The Indiana University Center for Postsecondary Research (CPR) has led several major studies on the student experience and houses the National Survey of Student Engagement (NSSE) which is one of the most advanced and widely adopted evidence-based institutional assessment activities. CPR is home to the newly created National Institute for Learning Outcomes Assessment (NILOA).
- The Council for Aid to Education (CAE) has designed and led one of the largest assessments of higher education outcomes, the Collegiate Learning Assessment (CLA).
- ETS is recognised as one of the world's foremost educational assessment and research organisations, with a strong international focus, and is well positioned to provide significant input in the areas of questionnaire development, and in linking the proposed problem solving work to existing projects such as PIAAC, PISA and TALIS.
- The IEA DPC has more than 15 years of experience concerning data processing for large-scale surveys.
- NIER is Japan's premier educational research and development agency, and has participated in many OECD, IEA, UNESCO and APEC projects of direct relevance to AHELO.
- SoNET systems has substantial experience developing and delivering large and complex software development and IT infrastructure projects, including the development of several online testing systems.
- Statistics Canada is the national statistical office of Canada, and in recent years has contributed to numerous international research studies.
- The University of Florence School of Engineering has conducted significant work on engineering education, most recently via its leadership of the European and Global Engineering Education academic network (EUGENE), which represents 76 international partners.
- Westat has had an extensive history of involvement in international surveys such as PISA, and in large national studies such as NAEP, and offers considerable analytical capability.

The Consortium's management plan

13. This section details the management arrangements between the ACER Consortium and the OECD Secretariat to ensure the Secretariat is properly informed on all matters required for successful management of AHELO. In addition, a summary is provided of initiatives that are aimed at making project management and implementation more efficient.

14. As the Project Director, Hamish Coates is the primary spokesperson for the Consortium. He will maintain clear lines of communication with the Secretariat and ensure consistent decision-making processes and clear lines of accountability. Whenever required, other senior Consortium staff are available to provide direct advice on matters of detail. Arrangements for co-ordination among the different Consortium groups and individuals are described in subsequent sections of this AHELO Assessment Design.

15. The Consortium expects to communicate with the Secretariat on a day-to-day basis primarily through e-mail, but also via telephone and other means. A key mechanism for maintaining close communication between the Project Director (on behalf of the Consortium) and the Secretariat will be a proposed series of regular internet-based, video-based or telephone-based meetings.

16. The Consortium will make all meeting documents and records available to the OECD. In general, the highest levels of openness and accountability to both the OECD Secretariat and to AHELO participant

countries will be achieved on an ongoing basis through the use of a dedicated AHELO website as a repository for all key documents, meeting papers and records. The OECD Secretariat will have direct access to all documents via this website.

17. The Consortium will provide all required formal written reports of progress in AHELO implementation, as outlined in the AHELO Feasibility Study Terms of Reference. In addition the Consortium will regularly ask if any further steps should be taken to improve aspects of its work, to provide better information, or in any other way to give the Secretariat improved levels of support it needs to monitor and oversee project implementation.

18. Hamish Coates will also be responsible for communicating with the AHELO Technical Advisory Group and for co-ordinating the implementation of advice provided by the expert groups. Hence the accountability of expert and advisory groups will be directly through the Project Director.

The Consortium's guiding principles

19. Significant innovation will be required at this formative stage of the AHELO Feasibility Study and the following 'guiding principles' detail key aspects of the philosophy underpinning the work of the Consortium. These principles are forward looking and aspirational, and are intended to provide insight on what a feasible assessment of higher education learning outcomes would look like.

20. These guiding principles affirm and extend those presented in other AHELO documents concerning quality and comparability, innovation in measurement and process, efficiency and cost effectiveness, partnerships and collaborations among all actors, balancing country-level and international needs, maintaining the highest levels of transparency, and providing actionable information to institutions and systems.

New understanding

A picture of success

21. The Consortium's work plan is influenced by a vision of success. A successful feasibility study will prove that it is possible to undertake an international assessment of final-year students' capacity to use, apply and act on their knowledge and reasoning. Moreover, this study will prove that it is possible to assess these outcomes in an internationally comparable, efficient and scalable way. New methodologies and technical standards will be established for higher education policy research. Policymakers, institutional leaders, faculty and students will be engaged, and see the assessment of higher education learning outcomes as an essential checkpoint in the educational process. Institutions will begin taking steps to convert learning-outcome results into improvement-oriented changes. Industry and government leaders will see new possibilities for assessing graduate capability. International education will have a new educational foundation. Importantly, the outcomes will be seen to offer a significant, effective and additional means of understanding the outcomes of higher education.

22. This vision sets high standards for aspirations, engagement and outcomes. This Assessment Design details the Consortium's approach for bringing this change about.

Internationalisation

23. An abiding appreciation of internationalisation pervades all aspects of the Consortium's approach, from the structure of the team and expert groups, to the testing of assessment frameworks, to developing rubrics which are interculturally responsive, and to the reporting of results. This is vital, for higher education is a global interest and enterprise, even in the most local activities.

24. The Consortium has devoted extensive consideration to what internationalisation means for the study's focus, methods, implementation and outcomes. The Consortium believes that AHELO should be a powerful force for stimulating even greater international thinking and development. Simply asking students to participate in an international assessment such as this sends a message that they are part of an interdependent global knowledge economy. Providing multidimensional reports to institutions will emphasise international learning networks and the transnational flow of knowledge.

Assessing global graduate capability

25. Measuring later-year students' learning and capacity to perform is likely to become a routine facet of higher education practice. Such measures offer information that complement assessments of achievement, and which can facilitate progressions into further study and professional practice. At the same time, the information can provide institutions and teachers with a reference point against which they can estimate the efficacy of their own goals and practices.

26. To add value, these assessments need to go beyond testing knowledge to testing students' capacity to reason in complex and applied ways and effectively use these skills and competencies in different settings. The assessments need to be sophisticated and align with the forms of thinking and professional work in which most graduates will likely engage. They need to employ a wide-range of methods, provide for a more balanced view of higher education quality, and tap capabilities that both educators and the professionals recognise as important for educational success – capabilities such as collaboration and teamwork, oral and written communication, creative and analytic ability and leadership. It is important to take account of the discipline(s) within which students learn, and the likely trans-disciplinarity of professional life.

27. Developing such assessments requires innovation. It requires conceptualising new constructs, developing items and systems for capturing high-level reasoning, and reporting in informative and informed ways. The Consortium is aware of this mandate, indeed many of the members have led the new thinking in this area. Our methodology has been designed to build on such developments.

New perspectives on higher education

28. The Consortium understands the significance of AHELO to the global higher education community and OECD countries. Even at the feasibility stage, the study's process and outcomes have the potential to induce fresh thinking about higher education. This calls for a conceptually rich understanding of higher education, of the value of new methods, and of the implications which may flow from our successful design, implementation and evaluation of the study.

29. It is vital, for instance, that the feasibility study develops a means of informing diversification within higher education. We define diversity as a concept indicating the level of variety of entities within a system, and differentiation as a process in which new entities emerge in a system. While differentiation devotes a dynamic process, diversity refers to a static situation (Van Vught, 2008). The existing global rankings focus on vertical overall institutional diversity and are not able to address programmatic diversity (Marginson & Van der Wende, 2007). By developing multidimensional approaches that move beyond standardised institutional rankings, AHELO has the capacity to inform and enhance each institution's distinct mission and autonomy, and their subsequent efforts to improve performance.

*New methods**High-quality, efficient and meaningful methodologies*

30. The Consortium has developed a robust and effective methodology for AHELO. Many of these methods have been developed and tested in school-level research since the middle of the twentieth century. This AHELO Assessment Design outlines how the Consortium will adapt and extend these sound approaches within the unique operating context of higher education. It details, for instance, how instrumentation and fieldwork will involve sampling students and items to enhance the targeting and efficiency of assessment, and hence the precision and validity of outcomes. It details the Consortium's multifaceted quality assurance strategy, and our approach to building awareness of and confidence in the processes with higher education policymakers, practitioners and researchers.

31. The AHELO Feasibility Study makes assumptions about the nature and future of higher education that are considered by some to be radical and even controversial. Given this, our approach incorporates a number of features that have been designed to address challenges related to the validity of outcomes across institutional contexts, cultures and languages. As this Assessment Design outlines, this involves formal quality assurance procedures that will be implemented by national project managers, partner organisations and expert groups. It also includes more tacit approaches to quality monitoring in which, for instance, staff of the Consortium are continuously reviewing and questioning their methodologies and testing alternatives. A well structured and reflective methodology will advance core principles in ways that reflect the innovation and growth that is required for AHELO.

Advanced approaches to quality assurance

32. The assumptions on which AHELO is based challenge the traditional definitions of 'quality' and 'academic standards' in higher education. Simply focusing conversations on 'student learning outcomes' and 'graduate capability' brings sharper focus to these important educational fundamentals. Moving beyond this – producing valid, reliable and meaningful data that can be benchmarked internationally – will provoke powerful and innovative thinking about enhancing and assuring the quality of educational provision and the student experience. AHELO has the capacity to develop standards-referenced levels of performance, for instance, against which institutional performance levels may be observed to vary.

33. The Consortium will design and manage the AHELO Feasibility Study with a deep awareness of these contexts and possibilities. At each turn, we will consider not just how to manage the assessment, but how to position the development in ways that provide the most robust foundations for future analysis and development of higher education. It is vital, for instance, that AHELO provides information on capability that is multidimensional in scope and promotes diversification not standardisation. It's also vital that the process and outcomes offer a legitimate complement to quality considerations that focus on institutional reputation, research performance, inputs, and teaching processes. The study should help understand the development of professional capability, and development of the professional workforce more broadly. The institutional reports need to stimulate creative reflection on feasible means of converting evidence-based insights into change.

Enhancing the rigour of higher education research

34. Developing a scientifically defensible as well as practical international approach to measuring student learning outcomes will be a major advance in higher education research and development. Higher education is an enormously diverse activity that involves people and ideas crossing disciplinary, institutional, community and national boundaries. So far, however, this mobility and multidimensionality has not been matched by the development of robust large-scale data collections. Much research on higher

education is country-, institution-, discipline- or even course-specific, and much of it is commercial in intent and without transparent methodological foundations. The ACER Consortium will implement a design for AHELO that sets new standards for the scope and quality of empirical studies of higher education.

A plan for engaging systems, institutions, faculty and students

35. The design, implementation, evaluation and future of AHELO will not succeed without a sound plan for engaging institutions, faculty and students. Institutional- or program-level level assessments that build macro structures around practice can be difficult to link with that practice and energise the people that make change happen. The Consortium will use an overarching engagement strategy that links individuals and institutions, disciplines and teachers, leaders with systems, and makes participation an informative experience in itself. We will develop scalable strategies for communication, promoting participation, and data use by participating institutions.

36. The scope of the feasibility study allows for developing and testing key facets of this strategy. Once established, this foundation laid will make it possible to efficiently increase the number of systems and institutions in future replications, to assist policymakers and researchers in making better use of findings, to help institutional researchers and educational developers enhance practice, and to inform learner aspirations and choice. Along with these dividends, the strategy itself offers a policy instrument for exploring the engagement and investment of stakeholders in AHELO.

Practical contexts and constraints

37. While this AHELO Assessment Design is methodological in substance and intent, the design and development of the first international assessment of higher education learning outcomes is a non-trivial proposition that is invariably shaped by a range of practical constraints. An important constraint is imposed by the breadth of scope of the work – that is, by what is not able to be included in the study. First, it is obviously not possible to evaluate the feasibility of any processes and materials which are not included in the study. A second important kind of constraint is imposed by contexts and pressures that unfold once the study has commenced. Background on each of these constraints is provided here to contextualise the document that follows.

38. As discussed throughout this Assessment Design, the basic focus and scope of the AHELO feasibility study was formed through a series of meetings convened between 2006 and 2009. A suite of background work was conducted and, along with meeting deliberations, this work was used to draft the Terms of Reference for the AHELO Feasibility Study. These Terms of Reference were quite expansive in their scope and touched on most areas that might be included in a full-scale study. This Assessment Design was drafted in response to these Terms of Reference and in certain respects maintains the expansive aspirations for the project that these advanced.

39. Between November 2009 and July 2010 a series of scope reductions needed to be made in response to funding and other constraints. As a consequence, the scope of the work was significantly reduced. Some of the decisions made included:

- removing multiple choice items from the generic skills component and reducing the number of performance tasks;
- developing only a provisional assessment framework for the Economics strand based on limited research and consultation, and producing only a ‘mini instrument’ that does not include any innovative constructed-response item types;

- narrowing the scope of the Engineering assessment from three subfields to Civil Engineering alone, developing only a provisional assessment framework, and producing an instrument that measures ‘above content’ but does not include any innovative constructed response item types;
- eliminating certain context instruments altogether, and delaying design and development of the reduced set of context assessments;
- significantly reducing project staffing, Expert Group and National Project Manager meetings, coordination among contractors, and face-to-face meetings;
- reducing national resource requirements, and the operational and technical support available to countries;
- revising the adaptation, translation and quality control procedures to achieve cost reductions;
- making greater use of small-scale qualitative testing procedures and deferring quantitative/psychometric validation of the assessment instruments;
- deferring operationalisation of the instruments, either using paper or online systems;
- deferring design and development of all facets of implementation, including: sampling, fieldwork training, coder training, facets of quality management, data preparation and validation, and analysis and reporting; and
- deferring the design and development of certain reports, including databases and codebooks, compendia, technical reports, and institution reports.

40. While not overly detailed or exhaustive, this list outlines the breadth of the scope reductions implemented between November 2009 and July 2010, and the consequent reduction in the feasibility evaluation. Detailed work plans for Modules A, B, C and E provide in the Annexure of this AHELO Assessment Design provide detailed operational information about the focus and scope of the study. As these work plans show, current plans only cover the instrumentation phase of AHELO with plans for the implementation phase still taking shape.

41. While this Assessment Design is forward-looking it is imperative, given these scope limitations, to limit expectations and enthusiasm for what the AHELO Feasibility Study is able to achieve. Further, as the study is taking an evolutionary rather than planned approach to its development, its full scope is not yet clear. Given these uncertainties, it is important to stress at the outset of this Assessment Design that there are limitations on how much can be concluded on the basis of what is being undertaken in the feasibility study in its current form. Certainly, the work is non-trivial and will test key facets of a full-scale study, potentially providing proof that a larger study is possible, but the constraints that shape the feasibility study’s approach and hence outcomes mean significant caveats surround the reporting and interpretation of results.

MODULE A: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE GENERIC SKILLS STRAND

Introduction

42. This section presents work to be undertaken by the Council for Aid to Education (CAE) to carry out the adaptation of the Collegiate Learning Assessment (CLA) towards an international version of the instrument. This adapted instrument will be used in the generic skills strand of the AHELO Feasibility Study.

43. As part of the AHELO Feasibility Study, CAE will adapt two CLA tasks and their scoring rubrics now being used in the USA for delivery in the implementation phase of the AHELO Feasibility Study with extensive summaries of their content provided to OECD to provide understanding of what the CLA is measuring. A detailed discussion of the structure of the performance tasks and critical thinking measures is provided in chapter five of *The Architecture of the CLA: Returning To Learning In An Age Of Assessment* (Benjamin et al., 2009). Summaries of the content of CLA adapted tests will be based on the logic and test design principles presented here. Participating countries will deliver the agreed upon version of the CLA in their home language via the internet in a proctored environment on the Internet Testing Systems platform designed for the CLA.

Development of the CLA Performance Tasks

44. Module A involves translation and adaptation of two selected CLA Performance Tasks (PTs). This involves establishing in-country teams, and selecting, adapting and translating tasks. Following this, tasks will be prepared for operationalisation.

45. Initially, preparatory work will be undertaken with countries to establish National Project Managers (NPM) structures and resources. Participating countries (Egypt, Finland, Korea, Kuwait, Mexico, Norway, selected USA states, and any others to be selected), in collaboration with Group of National Expert (GNE) delegates and following established criteria, will select NPMs for the project. The countries must agree to carry out agreed-upon-tasks in a timely fashion to meet the strict time line. Non-compliance to the schedule by countries could significantly delay the study. All NPMs stipulate that they will have access to software and computer/telecom equipment for conducting work using agreed-upon software. CAE has discussed by phone with designated participating NPMs any special needs or concerns that need to be given consideration during the first meeting of the country representatives. CAE has discussed a conceptual framework and procedures for training NPMs to adapt each CLA PT's content to reflect their contexts. CAE has carried out an NPM meeting in New York and assembled materials for circulation to the NPMs.

46. The aims of the initial Module A NPM meeting in New York were to:

- introduce NPMs (ideally two representatives from each country – a GNE member and an assessment expert);
- familiarise teams with the CLA;

- review a subset of nine CLA PTs and select four PTs considered suitable and valid in an international context, two of which were selected for modification and use in the AHELO Feasibility Study based on recommendation of CAE (the subset was selected using the following criteria: (a) universality of performance-task theme, (b) ease of translating based on complexity of language in performance task, (c) ease of scoring based on USA experience with the CLA); and
- form the Assessment Adaptation Group (AAG) – consisting of the assessment experts from the NPM team, which CAE trained in task adaptation, the translation and review process, procedures, and the recruiting of test translation teams.

47. After the meeting, NPMs on return to their respective countries are to begin the process of reviewing the contents of the two selected PTs and provide recommended modifications to ensure fit within their country’s context following agreed upon procedures. Subsequent teleconferences will be held to:

- present each country’s recommended modifications of selected CLA PTs for discussion/review (with the AAG member taking lead); and
- gain consensus on the two PTs based on CAE and NPMs’ evaluations and recommendations for final development.

48. CAE will fully modify two PTs (tasks, scoring rubrics and IT administration procedures) following modifications agreed upon at telecom meetings and send draft PTs for review, comment, and revision as they are completed.

49. Translation will follow the adaptation work. For this, participating countries in collaboration with CAE will be responsible for recruiting two teams, each of which will ideally include a:

- translation team, comprised of two (or more if country desires) members who will translate the CLA following agreed upon translation protocol; and
- translation review team, comprised of a translator, a university professor knowledgeable in content area, and an assessment expert (or more if country desires) who will review and resolve differences with the translation team. Participating countries will be responsible for compensation (if any) of the translation team and translation review team members.

50. Figure 1 further stipulates the desired membership and capabilities and expertise required for the translation team and translation review team.

Translation team	Translation review team
<ul style="list-style-type: none"> • Two certified professional translators English-to-Target language • Assessment adaptation expert will serve as the leader • Desired additional team members: <ul style="list-style-type: none"> ○ Content specialist ○ Reconciler ○ Linguist 	<ul style="list-style-type: none"> • One translation review leader (assessment adaptation expert) • One specialist in the content area (university professor) • One certified professional translators English-to-Target language • Desired additional team members: <ul style="list-style-type: none"> ○ Linguist

Figure 1: Desired membership of Module A in-country translation teams

51. CAE representative(s), expert in translation and translation review, will visit and assist, train, and guide translation team and translation review team members in each of the participating countries. CAE will conduct, as needed, telephone meetings with the translation teams to support the translation process and the sharing of translations for review and finalisation.

52. The country translation teams will translate the two modified PTs and other testing information as they are finalised so that small pilots (of around 10 students) can be conducted in home countries following CAE protocol – including cognitive workshops – conducted by NPMs with guidance and assistance of CAE. Cognitive workshops provide information about the kinds of thinking students do when they engage CLA performance tasks. A small sample of students is asked to talk ('think') aloud as they carry out the task, only interrupted if they fall silent. Extensive research shows that these 'think alouds' provide access into the students' mental thinking processes and thereby enable us to insure that the thinking elicited by the performance task is the thinking sought. In this way, we can see if the performance task measures the same thinking across countries.

53. In summary, the following translation cycle is anticipated: Translation team creates first version in home language, circulates it to the country translation review team, gets feedback and modifies translation. The tasks are then pilot tested. The translation cycle is then repeated.

54. Only one language per participating country will be chosen for the field test (with the exception that Kuwait will use both English and Arabic versions). Translated PTs, scoring rubrics, and IT administration procedures will have the same online presentation formats as the USA counterpart PTs. Translating teams will translate the online testing instructions early in the process (because these take longer to adapt onto the test delivery website). CAE will produce the final versions of the translated performance tasks for incorporation into the internet testing application (testing instructions will already be translated by participating countries in the prior phase).

55. As a separate optional activity, CAE will conduct 'CLA Faculty Academies' to engage faculty in understanding the CLA approach to measurement and how it can be utilised not only for university-wide assessment but also for testing individual students.

Module management

56. Please refer to the Module A Work Plan in the Annexure for detailed information about specific objectives, focus and scope, schedules, milestones and deliverables, and personnel, responsibilities. This Work Plan pertains to only the first phase of the AHELO feasibility study.

References

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., Nemeth, A. & Shavelson, R. (2009). *The Architecture of the CLA: Returning To Learning In An Age Of Assessment*. New York: CAE.

MODULE B: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE ECONOMICS STRAND

Introduction

57. Economics education is often subdivided into a handful of broad categories: macroeconomics, microeconomics, statistics and econometrics. The study of macroeconomics focuses on the economy as a whole, whereas, microeconomics focuses more on how individuals and institutions interact and make decisions. Econometrics and statistics underpin many facets of Economics, particularly those focused on modelling data or which are more applied in nature. However, there is no common curriculum within a country, much less across countries, for studying Economics in general or any sub-strands more specifically.

58. The challenge for the team developing the AHELO Feasibility Study Economic framework and the instrumentation is to determine, given the constraints of scarce resources, which models and tools, which content within macroeconomics and microeconomics, and which other areas of Economics are most important and accepted worldwide as essential in an ever-changing world.

59. Module B will be led by staff at Educational Testing Service (ETS). ETS is a non-profit organisation with substantial resources to support the proposed research activities. Within ETS, the Assessment Development group in the Research and Development Division will primarily be responsible for this effort.

60. ETS will draw on its extensive experience and resource base to develop the Economics assessment for the AHELO Feasibility Study. The sections that follow provide more detailed description of the proposed item development process and the proposed scoring process.

61. ETS staff have extensive experience in developing both multiple choice and free response questions for Economics assessments for different groups of students. The purposes of the assessments and the populations tested vary considerably. ETS staff members played major roles in establishing the specifications and the formats of the tests based on the purposes of the assessment. These include the Graduate Records Examination® (GRE®), the Major Field Test for Economics, the Advanced Placement® (AP®) macroeconomics and microeconomics tests, the College Board's College Level Examination Program® (CLEP®) assessments in microeconomics and macroeconomics, Praxis™ Economics,

Approach to framework development

The purposes of assessment frameworks

62. An assessment framework serves a number of critical purposes. It provides an agreed upon definition of the domain to be tested. This becomes the basis for instrument development. Frameworks give a theoretical and technical basis for the investigation, reporting and discussion of educational outcomes. The framework can provide the link between test outcomes and educational practice. It sets out the landscape of the content domain. Ultimately the framework is the practical basis for allowing the ETS

team, the OECD, countries, institutions and stakeholders to consider implications of the AHELO Economics strand results for the feasibility of ongoing and broader assessments in Economics.

63. As flagged in the Module E chapter of the Consortium's Assessment Design, there are four questions that the OECD and ACER Consortium's Module B team will need to answer when the feasibility study is complete:

- Was the framework that was developed reflective of an international consensus about the important learning outcomes in Economics?
- Was the instrumentation that was developed on the basis of the framework faithful to the spirit and intent of the framework?
- Was the instrument delivered and scored successfully?
- Do the test results stand up to psychometric scrutiny?

64. If the answer to all of these questions is 'yes' (and if the instrument is successfully delivered) we would be able to say that the Economics strand of the AHELO Feasibility Study is a success.

65. It must be noted that for the purpose of the feasibility study a full assessment framework for Economics will not be developed. Given the short timeframe available for development, the constraints on consultation, and given the complexity of the field of Economics, the framework that results at the end of the feasibility study will be considered a 'provisional framework.' The 'provisional framework' will not be complete, but will contain most of those elements which the majority of experts in the field agree are essential.

Framework development approach

66. Developing a framework which the participating countries and the global educational community can feel committed to is the most important part of the work of Module B and the foundation for the project's success. The key to success lies in involving as many stakeholders as early and as thoroughly as possible. This includes the OECD, the world community of experts in Economics education and national experts in Economics from participating countries.

67. Developing a conceptually rigorous, internationally valid and operationalisable framework for the field of Economics is a major task and contribution in itself. To reach international agreement a considerable amount of conceptual development will be required, along with extensive consultation and review. This includes provision for a number of face-to-face meetings of the Economics Expert Group, along with several other videoconferences and online and telephone communications in order to obtain necessary engagement with and endorsement of the framework.

68. Much of the groundwork for framework development has already been laid through by the OECD through its sponsorship of the 'Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics' (AHELO Feasibility Study Tuning Framework) (Tuning Association, 2009). ETS will use this document as the basis of the development of the AHELO Feasibility Study Economics framework. Given the short time frame available for the development of the framework and given the complexity of the field of Economics, the framework that results at the end of the feasibility study will be considered a 'provisional framework'. Every effort will be made to make the framework as complete and final as possible, but ETS believes that by signalling to the Economics community that this framework is a 'work in progress', that we will be able ultimately to engage that community in a productive and ongoing dialogue. This dialogue will result in a generally agreed upon framework at the time the AHELO Feasibility Study will be ready for full scale implementation. ETS has invited many of

the experts who worked on the AHELO Tuning conceptual group to join the AHELO Economics Expert Group and to work on the development of the Economics framework.

69. The ETS test development team is working with Professor Rae Jean Braummuller Goodman, Professor of Economics at the US Naval Academy, to produce draft framework documents for the Expert Group to consider. This document will then be revised to reflect the thinking of the Expert Group. The ETS team is working between the Expert Group meetings and is in touch with the experts between meetings in order to solicit their views on the work in progress.

70. By the end of the second Expert Group meeting the Economics strand assessment framework will:

- broadly reflect the current thinking of experts in higher Economics education;
- be informed by the needs and views of the leading international Economics associations;
- provide scope for collecting data to address specific research questions and other relevant aspects of interest;
- take into account the characteristics of the target population — in this case, final year ‘first-cycle’ (bachelor degree) Economics students;
- define the subject domain in terms of features such as content areas, skills and processes;
- recommend the relative weights to be given to each part of the domain that is measured, reflecting its importance to the domain;
- be specific enough to be useful in the instrument development process, but not so tightly specified that the opportunity for the assessment of integrated and above-content skills and conceptual understandings is removed;
- take into account cultural and language differences of participating countries;
- be reviewed and discussed by all participants in the Economics strand of the study so that consensus can be reached; and
- recommend a suitable mix of item response formats.

Economics assessment instrument development

General approach

71. The overarching goal of the AHELO Economics strand is to find ways to measure those learning outcomes in Economics that are ‘above content’. ETS acknowledges that there are constraints of timing and budget which might make it difficult to build a test from the ‘ground up’ that is perfectly aligned with the to-be-developed framework. ETS will therefore implement a hybrid task development model which will be both efficient and sensitive to the spirit and intent of the framework.

72. Given project constraints it is necessary to develop a ‘mini instrument’. The items in this ‘mini instrument’ will assess the main areas of focus identified in the provisional framework. The ‘mini instrument’ will be able to serve as the foundation for the final instrument. It will also be useful to demonstrate to stakeholders what the final instrument will look like. It is anticipated that almost of all of the tasks included in the ‘mini instrument’ will be able to be used in the final instrument, the only differences between the two will be breadth of coverage. It is also anticipated that what is learnt by exposing the mini test to students will lead to general improvements which will be reflected in the final instrument.

73. ETS will begin with Economics test items from two of the Economics tests that it has already developed – the GRE Subject Test in Economics and the Major Field Test (MFT) in Economics. Both of these assessments are directed at approximately the same population as the AHELO Economics Assessment.

74. After the framework has been developed, ETS staff will audit the test items against the framework. Staff will perform a triage to determine:

- which items have appropriate content and sufficient cultural generality to be candidates for inclusion in the AHELO Feasibility Study;
- which items can possibly be revised so as to be candidates for inclusion in the AHELO Feasibility Study; and
- which items will not be useful.

75. With the exception of the determination of items that will not work, the final determination in relation to items to be used will be made by the Economics Expert Group. The team anticipates that the questions used for the AHELO Economics assessment will, to a large extent, be ‘situational’ such that students will be given a real or hypothetical scenario and students will have to determine the correct outcome of any action/correct policy to use, etc. These questions are designed to assess whether students are able to use the economic reasoning learned in school to solve problems and resolve issues. Preliminary analysis by ETS staff of the existing item pool suggests that approximately two-thirds of the multiple choice items needed for the AHELO Feasibility Study can be drawn from the current ETS item pool. As the framework is developed, it may be that either fewer or more items are available. Some items may be used as is, while others may require small revisions. The constructed-response questions will be developed based on the requirements of the framework.

Measuring ‘above content’

Overview

76. Since the goal of the AHELO Feasibility Study is to measure those outcomes that are ‘above content’ ETS will supplement the items from the GRE Subject Test and the MFT with two kinds of additional items:

- innovative, machine scorable multiple choice items that measure higher order skills that will work equally well in paper or electronic delivery modes; and
- constructed response items that measure higher order integrative skills that will work equally well in paper or electronic delivery modes.

77. ETS will use Evidence-Centered Design principles (Mislevy, Steinberg & Almond, 1999; Mislevy, 2000; Mislevy, Steinberg, Breyer, Almond & Johnson, 2001) as part of developing any new task that needs to be created. In this approach, initial specifications for tasks, scoring, and interpretation are developed as part of assessment planning. These specifications take the form of respondent, task, and evidence models. The respondent model constitutes the experts’ view of how the components of proficiency are organized in each domain. The task model specifies the kinds of behaviours that will provide evidence of these proficiencies. The evidence model describes how to connect student responses to the specific proficiencies we wish to know about.

Innovative multiple choice items

78. Multiple choice items are often criticised for failing to measure skills that go beyond mere recognition and recall. This criticism is, unfortunately, often justified. But it is possible, especially given a framework that values integrative skills, to develop machine scorable multiple choice tasks that measure higher order skills.

Situational judgment tasks in Economics

79. ETS's Research and Development Division has done a considerable amount of work developing what are called situational judgment tasks (Kyllonen, 2005; McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001; McDaniel & Nguyen, 2001). These types of tasks are typically used to measure so called 'soft skills'. In the case of a test designed to measure the interpersonal skills of medical professionals, for example, test takers are given a video clip depicting the interaction between a doctor and an elderly patient's family. Candidates are then asked to evaluate the doctor's behaviour using a Likert scale. The 'correct' answer is determined by the consensus of view of experts and highly qualified practitioners.

80. ETS will suggest to the Expert Group that we adapt this concept to the Economics assessment. The students would then be expected to apply what they have learned in their Economics courses in order to respond to the question. The virtue of this approach will be that it should be possible to measure the learning outcomes that the AHELO Feasibility Study is interested in while stepping over, as it were, the skills of recognition and recall that are typically tested in multiple choice tests. It should be possible to vary the difficulty of tasks of this sort by choosing scenarios that are more or less routine by asking students to choose between courses of action that are more or less similar to each other. We will look to the Expert Group to suggest economic scenarios that could become the basis of these kinds of tasks to ensure that the tasks do not reflect parochial or regional biases.

81. ETS Assessment Development staff members have extensive experience working with question writers and members of committees, such as the Expert Group, to develop, review, revise and finalise both multiple choice and constructed response questions. For our work on the AHELO Feasibility Study, ETS will consult with the Expert Group to make sure that the various terms (such as 'demand function' and 'utility value') can be translated precisely enough so as to make questions which depend on them fair and valid for students in different countries. If there are certain terms that are preferred translations of specific terms of art the ETS team will make note of these terms and pass them on in the instrument documentation so that the translators will certainly use those terms that are preferred by the members of the Expert Group.

Constructed response tasks

82. In the Tuning AHELO document one of the learning outcomes that is specified is 'Effective Communication', which is defined as "the ability to communicate and explain effectively economic arguments both to those with disciplinary knowledge and non experts". The document states that such communication skills should be both written and oral and could involve the use presentation technology as well. In the context of the AHELO Feasibility Study we will probably need to limit ourselves to 'written' communications, but it is certainly clear that this key skill – the ability to communicate effectively about issues in Economics – cannot be measured directly with multiple choice items.

83. We will, therefore, work with the Expert Group to develop an instrument that is a mixture of multiple choice and constructed response tasks (see the discussion of instrument design for more detail below). We estimate that we will have between 30 and 40 minutes for two or three constructed response tasks, the number depending on the complexity and timing of the tasks selected. These tasks, which will be

developed through an Evidence-Centered Design process, will assess higher order integrative skills as well as communicative competencies. It is impossible to specify at this time the precise content of these tasks, because that will grow out of the framework and the design process described earlier.

84. One possibility, however, might be to assess whether students are able to understand, use, and interpret economic models. In order to do so students would be given a set of equations describing a macroeconomic model of the aggregate economy, for example. Depending on the purpose of the question, students might then be asked to explain (in the form of an essay or short answer question) how an exogenous change in one variable in the macro model, perhaps government expenditures or taxes, will impact the other variables in the model and what information in the model led the student to that conclusion. The scoring guide for such a question might focus both on the ‘correctness’ of the explanation and on the student’s discussion of the ‘meta cognitive’ aspects of the task, the student’s discussion of why certain kinds of information seemed significant.

85. Alternatively, another possibility to suggest the Expert Group is to provide students the production function of a firm and ask the students to derive the elasticity of output with respect to a given input. Students would then be asked to explain why knowing the elasticity is important to the firm and how this knowledge impacts decision making at the firm. In this case both the mathematical manipulations and the individual student’s response would be evaluated on the person’s ability to make the leap from a mathematical expression to a real world situation. For a different question we might give students information about different countries and their current trading patterns and tariffs. We would then ask what the expected change in imports and exports between the countries might be if some countries form a customs union, but others do not and why that change is expected to occur. We will work closely with the Expert Group to devise constructed-response questions of the sort sketched out above in order to get at those “above content” learning outcomes that span national and linguistic boundaries.

86. As was mentioned in the discussion of Evidence-Centered Design above, ETS realises that developing a scoring rubric at the same time as a constructed-response question is developed is critical to the process. We will work with the Expert Group to make sure that all of the experts have a shared understanding of what would constitute acceptable as well as exemplary performance on the task. This level of consensus building will help to ensure that constructed-response questions can be scored as consistently and reliably as possible.

87. ETS has a great deal of experience in developing constructed-response Economics tasks that measure higher order skills. The Economics content lead for the AHELO Feasibility Study was for many years the lead test developer of the AP Macroeconomics and Microeconomics test, tests which include a 60 minute constructed response section, 10 minutes of which are devoted to the student planning and outlining answers and 50 minutes for actual writing.

88. We will recommend that all constructed-response questions be developed following ETS procedures outlined in its 2005 publication “Guidelines for Constructed-Response and Other Performance Assessments” (see: www.ets.org/Media/About_ETS/pdf/8561_ConstructedResponse_guidelines.pdf). This document is based on the experience of ETS Assessment Development staff with many decades of experience working in different content areas and for different testing programs. It will be shared with the Economics Expert Group before they meet to discuss the constructed response sections of the assessment.

Scoring of constructed response items

89. In order to earn the confidence of stake holders in AHELO, the team working on Module B will demonstrate that constructed-response items that measure higher order skill can be scored reliably across cultural and linguistic groups. Great care, therefore, will be taken with the scoring. Scorers will be

recruited from the academic communities of the participating countries and institutions in order to ensure that well-conducted scoring will advance the goals of the AHELO Feasibility Study in two ways. In the first place, the resulting technical documentation will convince individuals who might be sceptical and, in the second, the scorers themselves, when they go back to their institutions will see the scoring sessions as learning opportunities. They will emerge from the sessions as ambassadors of the AHELO Feasibility Study effort.

90. An important part of the Evidence-Centered Design process is to begin with a notion of what successful performance on a task might look like. Working with the team leading Module E, scoring guides (documents which specify expected performance at each score level) will be developed in English when the tasks are developed. Scoring and the training for scorers will include the following steps:

- After the instrument has been administered during implementation we will convene a scoring training meeting. ETS staff, the national project managers and three content experts from each participating country will attend. The three content experts from each country will represent the intellectual leadership of the national scoring.
- Attendees at the meeting will discuss (in English) the scoring guide and develop a shared understanding of the distinctions between score levels.
- Attendees will go to break out rooms and pick ‘training papers’ (i.e. representative papers at each score point) in their own language.
- Attendees will gather together and discuss the training papers (in English) in order to reach a shared understanding of how the scoring guide should be translated into action; if necessary the scoring guide will be revised.
- Each language group will leave the meeting with a set of training papers in their own language.
- The final version of the scoring guide will be translated into the languages of instruction of each participating country. The translation will be verified by cApStAn.
- Attendees will return to their home country to train scorers using training papers and the translated scoring guide.

Item validation

Pretesting

91. Those test items from the GRE or MFT item banks that have a long history of use in the field will not be pretested, unless in the view of ETS staff and the Expert Group, the tasks have been modified to such an extent (either for the purpose of updating or to achieve a greater degree of linguistic or cultural generality) that they will perform in different ways.

92. Newly written tasks and substantially revised tasks will be pretested on small populations of third year college students. ETS, through the NPMs will endeavour to enlist the aid of faculty members at participating countries in order to make the pretesting population as representative as possible. If logistical and scheduling constraints make that problematic, ETS will rely on existing networks of US Economics faculty to pretest the tasks at colleges and universities within the USA, possibly involving a group of international students to help check cross-cultural relevance.

Instrument design and reliability

93. Rick Morgan, the psychometrician on the ETS team, will advise the content developers and the Expert Group on how best to weigh the mix of item types given the constraints of testing time and the desired technical characteristics. ETS will also work with the TAG to establish targets for the technical characteristics of the instrument. Dr Morgan will work with both the Expert Group and the TAG to reach a consensus on the optimal mix of question types given the framework and other considerations.

94. Based on other ETS assessments with similar questions of similar difficulty, ETS expects that, on average, students will take about 1.3 minutes to answer a question. Thus for a 60-minute multiple choice section, approximately 45 multiples choice questions can be asked.

95. Reliability is an indicator of the extent to which test scores will be consistent across different conditions of administration and/or administration of alternate forms of the test. Reliability is reported as an index ranging from .00 (complete absence of reliability) to 1.00 (perfect reliability). The closer this index is to 1.00, the better the test scores can be said to be free from errors of measurement. Based on a preliminary analysis of the data from various ETS Economics tests, Table 2 gives the estimated reliabilities for various 90 minute test configurations.

96. An assessment of this 90 minute duration is proposed to ensure face validity – which is imperative to gain acceptance of the economics community, to ensure appropriate levels of reliability, to ensure sufficient content coverage (international consultations so far have suggested that shorter instruments are seriously compromised), and to allow time for the context assessment.

Table 2: Estimated reliabilities

Instrument configuration	Estimated reliability
90 minutes multiple choice (67 questions)	0.90
90 minutes long constructed response (3 tasks)	0.74
90 minutes short constructed response (6 tasks)	0.79
60 minute multiple choice (45 questions)	0.87
30 minutes long constructed response (1 task)	
60 minutes multiple choice (45 questions)	0.87
30 minutes short constructed response (2 tasks)	

97. Although the highest estimated reliabilities would come from an all multiple choice test, such an approach would not allow for the kind of content coverage that AHELO requires. ETS will work with the TAG and Expert Group to decide which of the suggested configurations (or some variant) goes the furthest toward meeting the over arching goals of AHELO.

Documentation

98. ETS staff will prepare reports as needed, including a report that links or maps all assessment items to the respective aspects of the framework. Since the item development process will be based on the principles of Evidence-Centered Design (Mislevy, Almond & Lukas, 2003), all the items and tasks will be linked to the Framework throughout the development process. ETS is committed to a transparent design process in which the OECD Secretariat, national program managers, and members of the Expert Group are always aware of the links between the assessment items, assessment design, and the framework.

99. ETS staff will produce formal updates as required, including a full report on the framework and detailed technical reports which provide estimates of technical quality throughout the process. Formal reports will include, but not be limited to:

- framework, with rational and test specifications;
- operational updates (as required);
- technical report on instrument construction; and
- technical report on constructed response coder training.

Module management

100. Please refer to the Module B Work Plan in the Annexure for detailed information about specific objectives, focus and scope, schedules, milestones and deliverables, and personnel, responsibilities. This Work Plan pertains to only the first phase of the AHELO feasibility study.

References

- Bennett, R. E., Persky, H., Weiss, A. R. & Jenkins, F. (2007, August). *Problem solving in technology-rich environments: A report from the NAEP technology-based assessment project (NCES 2007-466)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- College Board, Advanced Placement, AP, Advanced Placement Program, College Level Examination Program, and CLEP are registered trademarks of the College Board. The SAT Reasoning Test is a trademark of the College Board.
- Educational Testing Service, ETS, *the ETS logo*, Graduate Records Examination, and GRE are registered trademarks of Educational Testing Service. Praxis and The Praxis Series are trademarks of Educational Testing Service.
- Katz, I. R. (2007). Testing information literacy in digital environments: ETS's iSkills Assessment. *Information Technology and Libraries*, 26(3), 3-12.
- Kyllonen, P. C. (2005). The case for noncognitive assessments. *R&D Connections*, 3. Princeton, NJ: Educational Testing Service.
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103-113.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.
- Mislevy, R.J., Almond, R.G. & Lukas, J.F. (2003). *A Brief Introduction to Evidence-Centered Design*. Princeton: ETS.
- Organisation for Economic Co-Operation and Development (2004). *Problem solving for tomorrow's world – First measures of cross-curricular competencies from PISA 2003*. Paris: OECD Publications.

MODULE C: DEVELOPMENT OF THE ASSESSMENT INSTRUMENTS FOR THE ENGINEERING STRAND

Introduction

101. Module C will be led by ACER in collaboration with NIER and the University of Florence School of Engineering. ACER has been setting foundations for developing an assessment of Engineering capability for nearly two years. An update on this work was presented at the December meeting of the AHELO GNE (Coates & Radloff, 2008). Since this time, background preparations have continued, along with meetings with high level networks in Australia and Japan. This work has provided the ACER-led team with a deep contextual understanding of how to assess Engineering capability in ways that are ‘above content’ and relevant to contemporary graduate and professional needs.

102. The profession of Engineering is intrinsically concerned with creating systems and products that are ‘fit for purpose’ and whose performance can be assessed objectively and quantitatively. The instrument created for Module C will embrace these principles by focusing on assessment of the core learning outcomes that define graduates’ prospective effectiveness. The set of assessed outcomes will be determined and refined with reference to international specifications. Their embodiment in the assessment instrument itself will be guided by the expertise of an international Expert Group.

103. Over the past few decades, the profession of Engineering and the roles of engineers have changed rapidly. The problems faced by engineers in today’s world are increasingly complex and require engineers to have both strong technical knowledge and skills, and understanding of relevant environmental, social, economic and cultural contexts. In addition, as for other professions, engineers are expected to be good communicators, be able to work effectively in interdisciplinary teams, to conduct themselves ethically and professionally, and to be able to constantly update and improve their technical and personal skills. The required Engineering flavour of these generic skills areas are well covered in the Engineering education and professional literature (e.g. Bons & McLay, 2003, Walther, Mann & Radcliffe, 2005; Gill, Mills, Sharp & Franzway, 2005).

104. Such changing requirements are continuous, but they are also identified formally in reviews that are undertaken periodically by national professional peak bodies. The past decade or so has seen such reviews in the United States (National Academy of Engineering, 2005), the United Kingdom (Royal Academy of Engineering 2007), and Australia (Institution of Engineers Australia, 1996; King, 2008). The recommendations in such reviews are usually focussed on changing university level Engineering curriculum and pedagogy, revising professional accreditation requirements, and intensifying connections to both professional practice and to school education.

105. The common trend in modernising Engineering education is to increase the focus on graduates’ competencies in project work, communication, and collaborative skills, and increase their understandings of ethical practice in the contexts in which Engineering problems and projects exist (Boles, Murray, Campbell & Iyer, 2006; Walkington, 2001; West & Raper, 2003). Underpinning much of the curriculum redesign and revision are the agreed graduate outcomes as required by national Engineering accreditation processes. Over the past decade, these have increasingly been framed in terms of graduates’ learning outcomes and competencies, rather than focusing on input measures. Thus, Engineering curricula are

specified in terms of expected outcomes, rather than subject content. There is also substantial commonality in the statements of these terms as used internationally by bodies concerned with both professional and education accreditation (Washington Accord, 2009), European Network for Accreditation of Engineering Education (ENAE), 2008, USA Accreditation Board for Engineering and Technology, ABET 2008, Engineers Australia (EA) (2006), UK Quality Assurance Agency (QAA) (QAA, 2006) and EU Tuning Process (Tuning Project, 2004).

106. While educational processes and outcomes in Engineering are relatively well defined, a need remains to produce robust data on learning outcomes and graduates' potential for subsequent success in work and further study. An assessment of Engineering capability undertaken as part of the AHELO Feasibility Study provides an opportunity to contribute to a more evidence-based approach to ascertaining quality in higher education. In collaboration with teams leading other modules the work undertaken in Module C will explore the feasibility of directly measuring learning outcomes in Engineering and across different cultural, linguistic and institutional contexts.

Developing the Engineering assessment framework

The purposes of assessment frameworks

107. The assessment framework for Module C will serve a number of essential purposes. It will contain an agreed definition of the domain to be tested on which instrument development can be based. In addition, it will serve as a theoretical and technical basis for the investigation, reporting and discussion of educational outcomes. The framework will also specify links between test outcomes and educational practice. Taken together, these elements allow the team, the OECD, countries, institutions and stakeholders to consider implications of the AHELO Feasibility Study Engineering strand results for the feasibility of ongoing and broader assessments of outcomes from the Engineering education in higher education institutions.

108. The key to achieving an agreed framework, to which countries can feel committed, lies in conducting careful preparatory negotiations that involve all stakeholders. This includes the OECD, the world community of experts in Engineering education, and national experts in Engineering from participating countries.

109. Much of this groundwork has already been done through the efforts of the OECD and its agents in the development of the AHELO Feasibility Study Tuning Conceptual Framework of Expected/Desired Learning Outcomes in Engineering (AHELO Feasibility Study Tuning Framework) (Tuning Association, 2009). The Engineering strand framework development will build on but not be restricted by this work. ACER has recently undertaken considerable background work in this area (Coates & Radloff, 2008), and we will build on this prior work as well. There is a deep alignment between these frameworks and the various professional competency statements that have been developed by institutions, regulators and the profession itself.

110. The Engineering strand framework is being developed to include one of the three branches of Engineering articulated in the AHELO Feasibility Study Tuning Framework, namely civil engineering. Working with this one branch will provide insight into the feasibility of the process and reduce costs. The significance of civil engineering has emerged from ACER's consultations with the international Engineering community as the most important to consider. Further, the team is aware that based on earlier documentation participating countries have already started engaging with their civil engineering communities.

111. As flagged in Module E of this AHELO Assessment Design, there are four questions that the OECD and ACER Consortium's Module C team will need to answer when the feasibility study is complete:

- Was the framework that was developed reflective of an international consensus about the important learning outcomes in Engineering?
- Was the instrumentation that was developed on the basis of the framework faithful to the spirit and intent of the framework in all languages?
- Was the instrument delivered and scored successfully in all languages?
- Do the test results stand up to psychometric scrutiny in all languages?

112. If the answer to all of these questions is 'yes' (and if the instrument is successfully delivered) we would be able to say that Engineering strand of the AHELO Feasibility Study is a success.

Guidelines for developing and reviewing the framework

113. Overall, the Engineering strand assessment framework will:

- reflect the current thinking of experts in tertiary Engineering education;
- be informed by the needs and expectations of Engineering professional bodies and practitioners;
- consider findings from previous research in this area;
- provide scope for collecting data to address specific research questions and other relevant aspects of interest;
- take into account the characteristics of the target population – in this case, final year 'first-cycle' (bachelor degree) Engineering students;
- define the subject domain in terms of features such as content areas, skills and processes;
- recommend the relative weights to be given to each part of the domain that is measured, reflecting its importance to the domain;
- be specific enough to be useful in the instrument development process, but not so tightly specified that the opportunity for the assessment of integrated skills and conceptual understandings is removed;
- take into account cultural and language differences of participating countries;
- be reviewed and discussed by all participants of the study so that consensus can be reached;
- recommend an approach for coding student performance; and
- recommend the mix of item response formats.

114. The ACER-led Engineering strand team will continue development of the Engineering strand framework using the AHELO Feasibility Study Tuning Framework as the starting point. The Engineering strand team will be working in close collaboration with the Engineering strand Expert Group and will provide the OECD Secretariat with ongoing advice about relevant developments.

115. The ongoing Engineering strand framework development work is largely unprecedented and consequently is likely to raise some unanticipated conceptual and practical issues. This being said, particular issues of special relevance to the review and final development of the Engineering framework that currently have been identified include:

- the differences among participating countries and institutions in their use of concepts relevant to the AHELO Feasibility Study definition of Engineering program assessment, and the implications of these differences for the operationalising of the framework through item development;
- the relationship between basic Engineering competencies and the branch-specific competencies, including:
 - the extent to which a common set of basic competencies applies across all branches and, what role that plays in the assessment of students in their final year of study in civil engineering; and
 - how basic competencies can be assumed: as achieved by candidates much earlier in their study and not assessed; as likely to have been met and assessed by a very small proportion of items; or as assessed through a relatively large proportion of achievement items;
- the structure of representing competencies across the different branches, including:
 - whether the structure of knowledge and understanding of outcomes suggested in the AHELO Tuning framework is the best way to represent the outcomes; and
 - whether different scores (or groups of scores) should be weighted equally within each branch, and whether they are expected to exist as different scales within branches or contribute to a single scale;
- the role of generic Engineering skills, including:
 - how the generic Engineering skills described in the AHELO Feasibility Study Tuning framework relate to other (non-Engineering) generic skills measured in the AHELO Feasibility Study; and
 - what proportion of the assessment should be allocated to the assessment of generic Engineering skills.

116. It must be noted that for the purpose of the feasibility study a full assessment framework for Engineering will not be developed. Given the short timeframe available for development, the constraints on consultation, and given the complexity of the field of Engineering, the framework that results at the end of the feasibility study will be considered a ‘provisional framework.’ The ‘provisional framework’ will not be complete, but will contain most of those elements which the majority of experts in the field agree are essential.

Framework development process

Audit of existing resources

117. Work on the Engineering strand framework acknowledges, integrates and builds on the work completed in developing the AHELO Feasibility Study Tuning Framework. Initially a supplementary audit will be conducted of Engineering standards, frameworks and outcomes to add weight to the outcomes of the Tuning Process. This audit will rely heavily on input from the Engineering Expert Group in particular to provide suggestions for additional fertile sources and connections within participating countries. It is not anticipated that this process will necessarily uncover significant ‘gaps’ in the description of the field in the AHELO Feasibility Study Tuning Framework, but rather to enhance this work and to provide additional evidence from which to consider the framework development issues detailed in the previous section.

Framework development

118. The Engineering strand team will carry out the development of the framework under the guidance of the Engineering Expert Group, and with the input of the Contextual Dimension and Generic Skills Expert Groups, using open and transparent consultation processes that facilitate the involvement of National Project Managers (NPMs) and of relevant experts in all participating countries, and that engage to the greatest extent possible the broader Engineering education community.

119. Given the limited time available for the framework development, the initial framework drafting work has commenced in parallel with the audit of existing resources. This work is being shared by the Engineering strand team and selected members of the Engineering Expert Group.

120. At the initial meeting of the Engineering Expert Group the framework development methodology will be finalised and the outcomes of the audit process, together with draft framework specifications and content, will be considered and revised in preparation for wider consultation. Following this meeting, a series of wider and well-targeted consultations will be undertaken to further enhance and validate the framework. The ACER-led team will work between the Expert Group meetings and we will be in touch with the experts between meetings in order to solicit their views on the work in progress.

121. A second meeting of the Engineering Expert Group will be held a few months later to evaluate the penultimate draft of the framework and to make any further changes. The team will make these changes, and the Engineering assessment framework will be finalised.

Test specification

A broad conceptual model for Engineering capability

122. Coates and Radloff (2008) suggest a broad conceptual model for Engineering that comprises three sub-domains: technical competence; Engineering process; and professional attributes. This model will be the starting point for considering different possible assessment designs for the AHELO Feasibility Study Engineering strand instrument. In particular the sub-domains of technical competence and Engineering process require an assessment design that can measure both the technical knowledge skills and understandings that underpin Engineering practice but also the 'authentic', relatively more open-ended research, and problem solving process that typify Engineering practice. Coates and Radloff (2008), for example, list six processes as comprising Engineering process: contextual awareness; research; problem identification; Engineering design; implementation; and verification. These six processes reflect and extend on the Conceive – Design – Implement – Operate (CDIO) Initiative (CDIO Initiative, 2008).

123. The significant feature of these conceptualisations of the role of the practising engineer is that they suggest 'authentic' Engineering practice to involve a sequence of interrelated activities. Ideally the Engineering assessment design will allow for the assessment of the process of Engineering as well as its discrete knowledge skills and understandings.

Alternative assessment designs

124. This AHELO Assessment Design acknowledges that the stated overall intention of the AHELO Feasibility Study is, as described in the AHELO Terms of Reference, not necessarily to develop comprehensive instruments to assess student performance. Instead, the focus is on providing proof of concept, and it should be possible to take advantage of this feasibility study to explore different approaches, methodologies and instruments that might eventually be envisaged as parts of a fully-fledged assessment.

125. The initial instrument development phase will begin with the development of a table of specifications. Part of the framework and instrument development processes will be to evaluate the ongoing feasibility of the design in the context of their integrity to the Engineering strand framework contents and practical considerations such as the available development time. A critical balance will be struck in the assessment design between the authenticity of the Engineering information gathered by the instrument and the practical feasibility of operationalising each assessment design.

126. The relative weighting of items by Engineering outcomes, appropriate number of items, common or branch-specific content and relative proportions of different item types will all form part of the collaborative negotiation around the framework design, item development and test design. Prior to further review and consultation, this AHELO Assessment Design makes no assumptions about these elements of the test design.

Engineering assessment instrument development

Overview

127. The success of the AHELO Feasibility Study Engineering assessment relies on the development of valid and reliable assessment instruments. While frameworks for Engineering are reasonably well developed, no single instrument currently exists that is fit for deploying directly for AHELO and it is necessary to develop and validate new items. This situation contrasts to that of Economics and is reflected in the additional resources required for the construction of this instrument. In broad terms, Engineering has a framework but embryonic instrument, whereas Economics has an instrument but embryonic framework.

128. The ACER-led Engineering strand instrument development team is using a range of professional test development teams and rely, wherever feasible, on the modification of existing successful items and approaches rather than building the instruments from the ground up. Test developers will be based in well-respected institutions in Australia (ACER), Japan (NIER) and Italy (University of Florence) and will work in extensive consultation with the Engineering Expert Group and other experts.

129. The selection of test development partners and members of the Engineering strand Expert Group reflects a strategic decision to make connections with and to include as many Expert Groups in the instrument development process as possible but to ensure that the instrument development process can still proceed quickly and efficiently.

Item submission process

130. In order to maximise national input to the item development process and to expedite the assessment instrument development process, item submissions will begin as early as possible. ACER has initiated links with the National Council of Examiners for Engineers and Surveying who lead the USA Fundamentals Exams (NCEES, 2008). Given time constraints and the instrumentation requirements for the AHELO Feasibility Study, the item submission process will be relatively less formal than that of other well-established assessment programs. In the initial stages, members of the Engineering strand team and the Engineering strand Expert Group will invite submissions from members of their international networks. This process can be regarded as advantageous to those submitting items as they will potentially have an influence in shaping the final international instrument. In addition to this, the parallel and connected development of the Engineering framework may be influenced by the nature of submitted assessment materials. The Engineering strand team will also conduct an active search for existing item materials that may contribute conceptually to the instrument pool.

Item types

131. Two assessment configurations:

- A 90 minute assessment comprising five or six modules with contexts, each with approximate six to seven items, containing multiple choice, short-constructed (such as numerical or short text) and extended innovative response; or
- 90 minute assessment, with two components:
 1. Component 1: comprising 15-20 minutes of multiple choice items covering basic engineering skills and sciences; and
 2. Component 2: comprising three to four modules with contexts, each with approximately six to seven items. Containing multiple choice, short-constructed (such as numerical or short text) and extended innovative response.

132. An assessment of this length is proposed to ensure face validity – which is imperative to gain acceptance of the engineering community, to ensure appropriate levels of reliability, to ensure sufficient content coverage (international consultations so far have suggested that shorter instruments are seriously compromised), and to allow time for the context assessment.

133. These two proposed working assessment configurations suggest a range of item types for possible sourcing and development. During the initial development phase, items of both types will be sourced, refined and developed. These may include:

- multiple choice;
- short-constructed response;
- extended (innovative) response.

134. The selection of items is both influenced by and an influence on the characteristics of operationalisation. The item types can all be implemented via paper or online.

135. Multiple-choice and short-constructed response items can provide a fast and efficient way to collect data on students' Engineering knowledge, understanding and skills. The items may be stand-alone or grouped in 'units' comprising more than one item relating to information in a common piece of stimulus.

136. Extended (innovative) response items typically require students to do more than provide a simple numerical or short-text response to a question. In the Engineering context these items may, for example, require students to complete short Engineering designs (typically in their specialty branch), describe analytic processes or evaluate and make use of complex data to make recommendations or suggest solutions to Engineering problems.

137. ACER's prior research and consultation with the Engineering and higher education community suggests that these innovative types of extended response items will play an important role in enhancing the:

- authenticity and validity of the Engineering assessment;
- reception of the AHELO Feasibility Study by Engineering educators;

- capacity to evaluate the effectiveness of different item types;
- alignment with the assessment approach used in Modules A and B; and
- reliability of the assessment – via the use of multiple scoring check-points.

Maximising the validity of items

Conceptual framework

138. First, the items will be based firmly on the Engineering strand assessment framework. Given the limited time in which the Engineering strand framework and assessment are to be developed, it is envisaged that these two processes will occur in parallel and be mutually informative. One of the most important roles of the Engineering strand Expert Group will be to monitor the extent to which draft test items and tasks properly reflect the conceptual and other priorities arising from the Engineering strand framework. All items developed will be mapped onto the relevant framework concepts.

National diversity

139. The item development teams will make use, wherever possible, of test items and source material submitted by participating countries. Given that only a small number of countries are participating in the AHELO Feasibility Study Engineering strand, the Engineering strand team and the Engineering strand Expert Group will use their extensive international networks to maximise the range of countries from which possible test items are submitted. This will be essential to ensuring that from a cognitive and conceptual point of view, test items reflect diverse modes of thought, and diverse cultural and national perspectives, experiences and priorities, particularly with a view to considering the feasibility of developing an instrument that could be used in a broad range of countries in the future.

Cognitive rigour

140. Given the limited time available for test development, it will not be possible to undertake an extensive program of piloting items and, for example, conducting many cognitive interviews about the items and test-taking. This being said, the Engineering strand team will make use of its networks and those of the Expert Group and participating countries to conduct qualitative tests of the items and pilots of the materials with small numbers of students. This process will also be used to inform judgements about appropriate test length (i.e. number of items and tasks of different types) and timing. In parallel to this qualitative testing, a small-scale quantitative test of the item materials will be undertaken at two or three Australian universities. This test will yield sufficient data to enable review of the key psychometric properties of the items and of how the items work together as an overall instrument.

Finalisation of the items

141. A series of final reviews and quality checks will be conducted of the items. These will focus on design and layout, editorial proofing, and ensuring that the items are well mapped against the Engineering assessment framework. The items will be prepared for operationalisation, and the team will deliver source versions of the Engineering assessment in English.

Coding guides and coder training

Refinement of coding guides

142. Coding guides will be developed in conjunction with all items and tasks that require coding. These guides will go through the same process of collaborative review by the Engineering Expert Group, NPMs and other stakeholders and experts as the items and tasks themselves.

143. Following data collection, the coding guides will be reviewed in light of a large number of authentic student responses in the different languages. This review allows the Engineering strand team to:

- check and when necessary refine the descriptions of student achievement described in the coding guides in the light of actual student responses;
- refine the guides to accommodate any previously unanticipated valid responses; and
- supplement the guides with example student responses that are indicative of the different substantive categories described in the guides.

Coder training

144. Following the refinement of the coding guides there will be a coder training workshop for senior coders. It should be noted that the timing of the coder training workshop is unusual in the context of international comparative studies. The choice of this timing is intended to ensure that a complete range of student responses in all languages are available to coders at the coder training workshop. It is intended that the coder training workshop will be a highly collaborative exercise that will serve three main aims:

- to train senior coders so that the interpretation of student responses is consistent across coders from participating countries;
- to refine and adapt the coding guides where necessary to ensure that they support consistent coding across coders; and
- to evaluate the feasibility of coding higher education student Engineering work in an international context.

145. In order best to achieve all three aims it is therefore proposed that the coder training workshop take place following the test administration. Given that only a small number of countries are participating in the AHELO Feasibility Study Engineering strand study at this stage, this is regarded as reasonable.

146. Coders will need to be Engineering experts, preferably with some background in Engineering education. Ideal candidates will be academic staff at (participating) higher education institutions.

147. The coder training for the feasibility study student responses will be an extensive and collaborative process. As part of the coder training process there is still opportunity to refine the coding guides to refine the language used in them, remove ambiguities to ensure that the coders interpret them consistently.

Module management

148. Please refer to the Module C Work Plan in the Annexure for detailed information about specific objectives, focus and scope, schedules, milestones and deliverables, and personnel, responsibilities. This Work Plan pertains to only the first phase of the AHELO feasibility study.

References

- ABET (2008). Criteria for accrediting engineering programs: effective for evaluations during the 2008-2009 cycle. Engineering Accreditation Commission. Accessed 2 October 2008 from:
- Boles, W., Murray, M., Campbell, D. & Iyer M. (2006). Engineering the learning experience: Influences and options. Proceedings of the 17th Annual Conference of the Australasian Association for Engineering Education, Auckland, 10-13 December.
- Bons, W. & McLay, A. (2003). Re-Engineering Engineering curricula for tomorrow's engineers: Engineering education for a sustainable future. Proceedings of the 14th annual conference for Australasian Association for Engineering Education and 9th Annual Women in Engineering Forum, Melbourne, 29 September to 1 October.
- CDIO Initiative (2008) The CDIO Initiative. Accessed 1 November from: <http://www.cdio.org>.
- Coates, H. & Radloff, A. (2008). Tertiary Engineering Capability Assessment: Concept Design. Canberra: Group of Eight.
- Engineers Australia (EA). (2006). Engineers Australia National Generic Competency Standards – Stage 1 Competency Standard for Professional Engineers. Accessed 8 October 2008 from: http://www.engineersaustralia.org.au/shadomx/apps/fms/fmsdownload.cfm?file_uuid=E01497F2-F208-96A1-14C4-AA8CAE8CFD2A&siteName=ieaust
- European Network for Accreditation of Engineering Education (2008): EUR-ACE® Framework Standards for the Accreditation of Engineering Programmes, Accessed 16 July 2009 from www.enaee.eu (ENAAE Documents)
- Gill, J., Mills, J., Sharp, R. & Franzway, S. (2005). Education beyond technical competence: gender issues in the working lives of engineers. Proceedings of the 4th ASEE/AaeE Global Colloquium on Engineering Education, Sydney, 26-29 September.
- <http://www.abet.org/forms.shtml>
- Institution of Engineers Australia. (1996). Changing the culture: Engineering education into the future. Canberra: Institution of Engineers Australia.
- Institution of Engineers. (1996). Changing the culture: Engineering education into the future. Canberra: Institution of Engineers.
- King, R. & Bradley, A. (2008). Addressing the Supply and Quality of Engineering Graduates for the New Century. Sydney: Australian Learning and Teaching Council.
- King, R. (2008). Addressing the Supply and Quality of Engineering Graduates for the New Century. Sydney: Australian Learning and Teaching Council.
- National Academy of Engineering (2005). Educating the Engineer of 2020. National Academies Press, Washington, DC.
- National Council of Examiners for Engineers and Surveying (NCEES) (2008). Fundamentals Exams. Accessed 1 October 2008 from: <http://www.ncees.org/exams/fundamentals>.

- Organisation for Economic Cooperation and Development (OECD). (2008a). OECD Assessment of Higher Education Learning Outcomes (AHELO). Accessed 1 October 2008 from:
http://www.oecd.org/document/22/0,3343,en_2649_35961291_40624662_1_1_1_1,00.html
- Quality Assurance Agency (QAA). (2006). Subject benchmark statement: Engineering. Accessed 2 October 2008 from:
<http://www.qaa.ac.uk/academicinfrastructure/benchmark/statements/Engineering06.pdf>
- Royal Academy of Engineering (2007). Educating Engineers for the 21st Century. RAE, London.
- Tuning Association (2009). A Tuning-AHELO conceptual framework of expected/desired learning outcomes in Engineering. Paris: OECD.
- Tuning Project (2004). Tuning educational structures in Europe: Generic competencies. Accessed 30 April 2009 from:
<http://www.tuning.unideusto.org/tuningeu/index.php?option=content&task=view&id=183&Itemid=210>
- Walkington, J. (2001). Designing the Engineering curriculum to cater for generic skills and student diversity. *Australasian Journal of Engineering Education*, 9(2), 127-135.
- Walther, J., Mann, L. & Radcliffe, D. (2005). Global Engineering education: Australia and the Bologna Process. Proceedings of the 4th ASEE/AaeE Global Colloquium on Engineering Education, Sydney, 26-29 September.
- Washington Accord (2005). Graduate Attributes and Professional Competencies. Accessed 1 October 2008 from: [http://www.washingtonaccord.org/IEM_Grad_AttrProf_Compencies_v11\(2\).pdf](http://www.washingtonaccord.org/IEM_Grad_AttrProf_Compencies_v11(2).pdf)
- Washington Accord (2009). Graduate Attributes and Professional Competencies,
<http://www.washingtonaccord.org/IEA-Grad-Attr-Prof-Competencies-v2.pdf>
- West, M. & Raper, J. (2003). Cultivating generic capabilities to develop future engineers: An examination of 1st year interdisciplinary Engineering projects at the University of Sydney. Proceedings of the 14th annual conference for Australasian Association for Engineering Education and 9th Annual Women in Engineering Forum, Melbourne, 29 September to 1

MODULE D: DEVELOPMENT OF THE SURVEY INSTRUMENTS FOR THE CONTEXTUAL DIMENSION

Introduction

149. Module D of the AHELO Feasibility Study will be led by CHEPS, alongside ACER and CPR. This module involves developing a series of instruments to assess the factors and conditions known to affect learning. This contextual information includes system characteristics, institutional settings, teaching practices, and other characteristics of learning environments. Such insights will play a vital role in enhancing the capacity of the AHELO Feasibility Study to better understand how students, faculty and leaders engage and productively interact in an increasingly diverse and internationalised higher education environment (Van Vught, 2008; Marginson & van der Wende, 2007; Kuh, 2009; Codling, & Meek, 2006; Scott, Coates & Anderson, 2008).

150. The development of surveys as part of a contextual dimension of the AHELO Feasibility Study will:

- help assess the feasibility of capturing contextual variables across different national and institutional contexts in valid and reliable ways;
- help rehearse psychometric analyses to identify relevant contextual variables for longer-term development and demonstrate the analytical potential of the AHELO Feasibility Study for institutional improvement;
- help manage sampling, data collection and quality control; and
- inform bias analyses undertaken as part of the cognitive assessments.

151. Within the scope set for the AHELO Feasibility Study, the team will set foundations for implementing a technically and conceptually advanced assessment of higher education contexts. A framework will be produced which uses the prior work (see: Ewell et al, 2008, 2009) to chart an understanding of those contexts of most relevance to contemporary higher education. Items will be produced to operationalise this framework in efficient, technically informed and environmentally relevant ways. The instrumentation will yield data that are internationally relevant and link systems information with the insights of teachers and students.

152. Through the approach given below, the team leading Module D will prepare draft and final versions of:

- the revised organising framework for the contextual dimension instrument;
- the student and system survey questionnaires;
- the report mapping the contextual survey items to the organising framework; and
- coding guides for the surveys instruments.

Context surveys framework development

Principles guiding framework development

153. The team will develop a survey framework for the contextual dimension of the AHELO Feasibility Study, building on development work already completed by the contracted contextual dimension experts (Ewell et al, 2008, 2009) and the AHELO GNE (as per the AHELO Terms of Reference). Peter Ewell acted as rapporteur for the commissioned work and has agreed to act as Chair of the Technical Advisory Group which will oversee the work in Module D, ensuring continuity, efficient progression, and excellent oversight.

154. In broad terms, the framework for the context surveys will:

- set the conceptual focus and scope for this aspect of the work;
- provide technical foundations for managing collection, analysis and reporting; and
- enhance the efficiency of the assessment.

155. The context surveys cross all strands of the AHELO Feasibility Study and will have close technical, practical and substantive links with all other facets of the work. Development of the assessment framework for the context surveys will be undertaken in partnership with teams leading the Economics assessment, Engineering assessment, Generic Skills assessment and fieldwork. Forming close linkages in this regard will play a vital role in ensuring the validity and efficiency of the AHELO Feasibility Study.

156. As with the cognitive frameworks, a number of guidelines and processes will be observed in developing the framework for the context surveys for the AHELO Feasibility Study. The framework will be designed with the following principles in mind:

- It will reflect the views of the participating countries as well as the thinking of leading experts in the field. A key is that the questionnaire Expert Group be made up of knowledgeable individuals with strong theoretical understanding of the contextual constructs and how they should be measured. These individuals will work with the team and the participating countries to reach consensus on how best to measure the constructs, including how best to communicate findings to key stakeholders.
- The framework will take into account the characteristics of the target population. The AHELO Feasibility Study will be conducted in a wide range of countries. Countries vary culturally, in how higher education fits into society and in many other ways. These country and cultural differences will be taken into account by the team working with the members of the Expert Group.
- The framework will provide an overall definition of the construct. An initial task will be to further develop working definitions of the background domains and constructs. The goal will be to reduce the variety of interpretations that can be associated with a given construct.
- The framework will operationalise the construct. Once the definition is developed, it is important to think about the kinds of questions or other information sources that can be used to measure the given construct. Those tasks must then be categorised, or organised, to inform the item writing process. There must be consideration of the kinds of evidence that could be gathered to determine the reliability and validity of inferences that could be drawn based on that scale.
- The framework will help identify task characteristics. The next step in constructing a framework consists of identifying a set of key characteristics that will be used in constructing tasks for the

assessment. For contextual questions this could include features such as item level, the item type and potential validity threats.

- The framework will reflect the relative importance of constructs and scales so that question inclusion can be decided and communicated to the various stakeholders.
- The framework will provide a template for mapping items. As outlined below, the team accepts responsibility for mapping each item back to the framework. For this aspect, the framework needs enough specificity about the key constructs being used to guide item development and classification.

Background reviews

157. The team's work on the contextual dimension will begin with a series of background reviews. This integrated series of reviews will detail relevant substantive, pragmatic and technical considerations of relevance to the development. This work will reflect and build on rather than repeat the extensive background work already undertaken.

Review of research

158. Using the existing work as a guide, an extensive review of relevant research will be conducted. Both published and unpublished (but accessible) research will be reviewed, and the team will consider documents which are scholarly and also applied in nature. Methodological papers will be considered so as to highlight potentially important technical characteristics of the instruments. This research will help to further establish the relevance of the concepts to be addressed, to chart technical characteristics of instruments in use, and to clarify and position the development approach. The review will be international in scope. As many of the most important developments in this area are included in the background work, our analysis will focus on identifying recent contributions and on prioritising relevant studies. This research review will help position the AHELO Feasibility Study development within the higher education community, and will shape and help establish the currency of the selected focus and approach.

159. This review will be expedited and enhanced by drawing on the extensive research undertaken by team and Expert Group members in this area over the last decade. Research by CHEPS has focused on understanding and measuring institutional contexts and diversity in higher education, and how such insights can be reported and used (Van Vught, 2008, 2009; Schwartz & Westerheijden, 2004). Research at CPR has explored student engagement (Kuh, 2001, 2002, 2003, 2009; McCormick, 2009), links between the subjective and objective assessment of student outcomes (Carini, Kuh & Klein, 2006; Klein, Kuh, Chun, Shavelson & Benjamin, 2005), faculty views on student learning (Kuh, Chen & Nelson Laird, 2007; Nelson Laird, Chen & Kuh, 2008) and, importantly, how to help institutions convert evidence-based insights into change (Kuh, 2008). ACER research has covered the design of large-scale context surveys (Coates, Tilbrook, Guthrie & Bryant, 2006; Coates & Hillman, 2008), the development of large-scale quality indicators (Coates & Hillman, 2008; Coates, 2007a, 2007b, 2007c, 2009), student engagement (Coates, 2006, 2008, 2009), contexts linked with success graduate capabilities (Coates & Edwards, 2009), the student experience (Coates & Ainley, 2006, 2007; Coates & Edwards, 2008, 2009; Edwards & Coates, 2008; Griffin, Coates, McInnis & James, 2003) and faculty work and academic leadership (Coates, Goedegebuure, van der Lee & Meek, 2008; Goedegebuure & Coates, 2009; Goedegebuure, Coates, van der Lee & Meek, 2009; Scott, Coates & Anderson, 2008).

Audit existing resources and practices

160. Simultaneously, an audit will be conducted of existing context survey resources and practices. This will focus on initiatives which are international, national and institutional in nature. These three levels

of analysis are important, because many contemporary collections are trans-national rather than national in focus, and many of the largest collections are cross-institutional in nature.

161. The Module D team will compile a database of relevant and available questionnaire items. These will be categorised according to pertinent substantive, technical and practical criteria. The team will attempt to source psychometric information on selected items. The item inventory is intended to be a working document that will facilitate the item development process.

162. The review will consider what existing documentation is available in participating countries and institutions with a view to minimising data collection. This will involve reviewing the confidentiality, currency and reliability of existing information.

163. Part of this practice review will seek to identify particularly innovative approaches to using survey and related assessment data for continuous improvement. A potential risk to the AHELO Feasibility Study is that data is collected that is not fully utilised by institutions to drive continuous improvement, or is not used at all. Hence identifying effective strategies for using evidence to drive change will help refine and position the contextual work.

164. This audit will be undertaken via National Project Managers (NPMs), in parallel to the research review, and through direct liaison with selected institutions and research institutes. The audit will further enhance general insights developed via the research review, will identify possible sources of items, will provide initial insights into policy and educational emphasis placed on different forms of data, will chart common and more advanced analysis and reporting practices, and will help flesh out and validate assumptions about the most efficacious means of collecting data.

Synthesis of existing practice

165. The results of the reviews will be analysed and distilled into a synthesis of existing research and practice. This will build directly on the existing work undertaken by Ewell et al (2008, 2009), providing validation and extension where necessary. Documentation of the background reviews is an intrinsically important activity, but the main operational purpose of doing this will be to inform the formation of a parsimonious but sufficiently rich framework.

A general structure for the framework

166. The framework will play a vital organising role in maximising the efficiency of the collection while at the same time ensuring conceptual alignment and technical rigor. A single high-level framework will be developed to situate the constructs and variables to be measured, assist with item generation and management, and facilitate statistical analysis and reporting.

167. The structure proposed by Ewell et al (2009: 4) aligns well with the hierarchical input-process-output framework. Over the years, this basic but robust framework has been used across a large number of diverse education systems and contexts (Bottani & Tuijnman, 1994; Astin, 1978, 1985; Ewell & Jones, 1996; Jaeger, 1978; Nuttall, 1994). The most established contemporary international manifestation is the OECD's indicators of education systems (INES) framework (OECD, 2008: 19). To build links with other analytical and empirical work we will use the INES framework as our most general organising principle. The INES framework organises policy and educational issues that might be considered in the AHELO Feasibility Study using two dimensions:

- the level of the higher education system to which the resulting indicators relate (systemic, provider, instructor, learner); and

- whether they relate to outputs or outcomes, policy-amenable determinants of these outcomes, or antecedents or constraints.

168. Adopting a taxonomy such as this carries several benefits. It provides a general model in which antecedents and processes independently and jointly (interactively) determine outcomes. It provides a scheme from which specific models can be developed and tested using techniques such as structural equation modelling. It also serves as a framework into which items can be classified, and gaps can be discovered. The taxonomy is comprehensive, such that additional constructs that might be useful in the AHELO Feasibility Study can fairly readily be classified into one of the cells.

169. By way of example, Figure 2 situates some of the variables listed by Ewell et al (2009) and in the AHELO Terms of Reference. This mapping is partial and indicative only, but serves to illustrate how a wide range of variables will be positioned within the framework. As flagged by Coates (2007b) and Krause, Coates and James (2005), there is a need to develop and validate a set of indicators for higher education. Further development and validation will be undertaken in the early stages of the project to ensure the relevance and veracity of this conceptual structure for higher education. This framework is indicative, and it will not be possible to capture data on all or even most of these measures are part of the AHELO Feasibility Study.

	Outputs	Processes	Antecedents
Higher education students	Self-reported gains	Student engagement	Gender Age Disadvantage status Parent occupation background Parent educational background Socioeconomic status First-in-family status Individual ability
Faculty and support staff		Pedagogies Academic support services Student support services Location	Curriculum structure Curriculum depth/breadth Teaching staff characteristics
Higher education institutions	Completion/persistence rates Employment rates within field	Learning environments Physical plant Teaching culture Research culture Institutional spending patterns	Student numbers Student diversity/mix Diversity in student ability Reputation and rankings Industry links Institutional selectivity Financial resources Teaching/research emphasis
Higher education systems	Completion/persistence rates Employment rates Further study rates	National quality assurance system	Institution numbers and types Tertiary participation Institutional capacity Import/export of tertiary students Student characteristics

Figure 2: Indicative conceptual structure for the contextual dimension surveys

170. This structure will be populated with variables nominated by Ewell et al (2009), the GNE (as per the Terms of Reference), and the further background reviews. As illustrated by Ewell et al (2008), a wide variety of context variables could be addressed by the AHELO Feasibility Study. Given the need for parsimony, we propose the use of the following priorities:

- Is the construct of policy relevance and interest? Is the issue of timely interest? Is there a public interest in the particular issue? Will the information obtained be of value in understanding student performance and other outcomes and taking steps to improve them? Is the issue related to a construct in which trends over time are of importance?
- Is the factor appropriate for international assessment? Is the issue one that can be addressed cross-culturally? That is, will it have a similar meaning across countries? Will comparisons relate to important policy issues? Is addressing the issue in the AHELO Feasibility Study adding value beyond what might be accomplished with an institutional evaluation?
- Within the context of the AHELO Feasibility Study design, is it technically feasible to address the issue? Are the variables used to address the issues reliable and valid? Do we know that from prior variable usage, or from the research literature? Is the time and cost associated with the variables used to address the issue reasonable? How do we evaluate whether items are working properly? Are there less expensive ways to collect the information?

171. The INES structure provides a sufficiently broad foundation locating the concepts of interest to the AHELO Feasibility Study contextual dimension. Following the process detailed in the 2003 PISA Technical Report (OECD, 2005) and the European Commission classification work (Van Vught & Kaiser, 2008; Van Vught, 2008, 2009), it will be necessary to add additional dimensions to this framework to both reflect the multidimensional nature of contemporary conceptions of higher education quality and enable specification of substantive concerns at a relevant level of detail. As the framework is fleshed out, it can be used to flag specific links between various factors and then design analyses that explore the direct or indirect relationships between these.

Consultations and review

172. Developing a framework which is conceptually robust, technically referenced and capable of efficient operationalisation will be vital for the success of the contextual strand of the AHELO Feasibility Study. It is important that key stakeholders see their policy and educational concerns and interests expressed by the framework. Numerous cycles of consultation have been built into framework development, and further feedback will be sought of a more summative nature to finalise development of the framework. This feedback, which will be managed via an online feedback mechanism and through the research team, NPMs, the Expert Group, will play a vital role in identifying the strengths and limitations of the framework. These somewhat summative consultations will also validate plans regarding operationalisation of the framework.

Documentation of the framework

173. The design, development and validation of the context surveys framework will be documented. This will provide a record of the development process, justify the steps undertaken for development, and link the process and outcomes with other aspects of the AHELO Feasibility Study. The team will deliver this revised organising framework for the contextual dimension surveys to the OECD.

Development of the context surveys

Overview

174. With a robust framework in place, the team will develop and prepare the following conceptually and psychometrically linked instruments:

- Student Survey Instrument (SSI), to be administered alongside the Generic Skills, Economics or Engineering assessments; and
- Faculty Survey Instrument (FSI), to be administered to teaching staff.

175. There will be a need to collect information that is system-wide in scope or which is available at the system level. To reduce the burden on NPMs, the team will collect this information centrally, making use of existing public sources where possible. Where necessary, the team will work with and support NPMs to capture and validate this information.

176. It is also likely that information at the institution level will be required to assist with management, sampling, fieldwork, analysis and reporting. Where necessary, the team will work with NPMs and ICs to collect information that is absolutely necessary from a technical and practical perspective to undertake the project. Information will be collected on programs (eg: Dean, Head, Program Coordinator) and institutions (eg: President, Vice Chancellor).

Consultative and rigorous item development

177. The approach used to develop and validate the context items align with those used for the objective assessments. Its defining features are that it will be highly consultative, phased, conceptually based, and iterative. In recent years, the approach has been tested and refined in several national and international studies. It is detailed here in terms of defining design specifications, mapping and drafting items, qualitative testing, quantitative testing, and final review. The same approach will be used for all instruments regardless of variation in the target population.

Defining design specifications

178. Instrument development will be guided by a number of general design considerations to enhance the power of measurement and ease of administration. These will align with the standards set for international data collections (Schmidt & Cogan, 1996; Mullis, Martin & Stemler, 1999; Siniscalco & Auriat, 2005), characteristics of large-scale existing context assessments (for instance: NSSE, 2008; AUSSE, 2009) and link with other survey design specifications recorded during the background reviews.

179. In summary, the instruments will be designed to:

- measure the target constructs;
- have high levels of face, content and construct validity;
- provide reliable and precise measurement of target constructs;
- be efficient to administer, analyse and report;
- align with and enhance existing instruments and practices; and
- provide a basis for ongoing research and development.

180. Item specifications will be set to enhance the quality of measurement and minimise response interference effects. Item specifications will be defined from studies that have sought to determine the characteristics of effective survey items (Kuh, 2002; Laing, Sayer & Noble, 1988; Yamaguchi, 1997; Converse & Presser, 1989; Andrich, 1993; Bradburn & Sudman, 1988). Such specifications include, for instance, that:

- items should be as succinct as possible;
- grammar and spelling should be accurate;

- the cognitive demand of each item should be less than the cognitive demand of the phenomenon being measured by that item;
- only a single phenomenon should be tested in each item;
- each item should have distinct and independent content;
- items should be distributed across the range of the variable;
- use of negative phrasing should be minimised or avoided;
- the information should be known to the respondent;
- items should refer to recent activities;
- items should be reviewed for possible bias to environmental or individual contexts;
- items should be phrased clearly and unambiguously;
- item design should encourage serious and thoughtful response; and
- items should not threaten, embarrass or violate the privacy of the respondent or encourage them to respond in socially desirable ways.

181. The SSI and FSI instruments will be designed for completion within 15 minutes, either in online or paper format.

182. The survey instruments will not be static in nature but rather dynamic instantiations of a complex matrix-sampling process. Item sampling will be used to create different versions of each instrument and provide an efficient means of getting content coverage, efficiency of administration, avoid the influence of item non-response, help target different items to different contexts, and help to test a larger number of items. The item sampling design will be developed to inform each country's national selection and use of optional sections of items. It will be sufficient for conducting the kinds of multivariate analyses that will be undertaken following full data collection. To enact the modular approach described in the Ewell et al (2009), the contextual items will be stratified by:

- target population group:
 - system/institution/faculty (precise groups to be confirmed); and
 - student;
- desired generalisability:
 - core (for all countries, institutions and respondents);
 - discipline-specific; and
 - optional sections for countries.

183. The development of an efficient and robust item sampling approach will play a vital role in balancing efficiency with administrative, validity and substantive considerations. It will allow for planned variation of the survey instruments across countries depending on their preferences and the amount of existing documentation in their national context.

Mapping and drafting items

184. Work will be undertaken to operationalise the multidimensional framework for the purposes of measurement. This will involve identifying the desired level and unit of analysis for each construct in the

framework, and then mapping the constructs to the different instruments, potentially in a many-to-many fashion. This exercise will link the framework with the instruments and chart the data structures and algorithms required for item sampling (which will be dynamic in online implementations). Correct specification of the level and unit of analysis is vital, as this carries implications for the efficiency of measurement and for sampling and reporting.

185. With this foundation, the team will undertake initial item drafting and review. For this, we will draw on the item inventory compiled earlier which will contain the most advanced items in use in relevant existing instruments. Every effort will be made to link with widely used available resources, including through establishing licensing arrangements that may be required. This will provide an assurance of quality, enhance the efficiency of the development, and help with analysis, reporting and interpretation. Where new material is developed, this will be panelled within the team, with the Expert Group, and with NPMs. As part of this process, item developers will be instructed to consider whether questions will be treated individually or as part of a scale.

186. The items will be drafted with reference to the underpinning assessment framework, a process that will be managed continuously throughout the development. This mapping process will provide information about where each survey item fits in the framework and how the combination of survey items in the surveys and existing resources achieves the overall measurement goals described in the framework.

187. Items will be developed for deployment in online versions. The item sampling strategy (varied for countries and institutions as per paragraph 177) will be best operationalised online. However, the items themselves will be trans-modal in structure.

188. Survey items and materials will be developed by the contractor in English. As specified in the Terms of Reference, the team understands that translation of survey items from English into the languages used in national assessments will be the responsibility of participating countries, and that responsibility for ensuring the quality of translations will form part of Module E. The team will work with countries and other partners to ensure that cross-national comparability of instruments is achieved in the translation process. The Module D team and Expert Group will be cognisant of language and cultural nuance when developing the instruments. The consultative nature of framework and item development will ensure cultural relevance across participating countries.

Item validation and review

Qualitative testing

189. A multifaceted empirical testing and review process will be undertaken to establish the face and content validity of the draft items and scales. The process of item review and development will be an inclusive one, involving the countries participating in the AHELO Feasibility Study and engaging experts from relevant scientific communities. As with other aspects of research design, survey instruments invariably reflect a compromise between practical, methodological and substantive considerations. A highly iterative and consultative validation process is one means of finding a balance between these forces.

190. Feedback on the draft instrument will be sought from a range of stakeholders. This feedback will help refine and shape items to align them with the assessment framework, educational contexts, existing quality assurance activities and research foundations.

191. An important consideration underpinning the instruments will be to ensure that the language used in the instruments is appropriate. An appropriate reading level will be set, as will common semantic conventions.

192. A series of focus groups will be undertaken in participating countries to capture student and faculty insights into the range and characteristics of the items. Resources will be developed and distributed to NPMs who will be trained, where necessary, to undertake the work. This testing will help determine whether the items measure appropriate phenomena, whether they are pitched at the right level, and are seen by potential respondents as being appropriate and useful. Through the process of probing and exploring responses, the focus groups will help explore reactions to the items, while simultaneously generating rich qualitative feedback to enhance the face and content validity of the instruments.

193. In addition to the focus groups, cognitive interviews will be conducted with members of the target population. These interviews will be conducted using a verbal probing method. In this type of cognitive interview, after the interviewer asks the proposed survey question out loud, and the interviewee responds using the proposed response set, the interviewer then asks for other specific information relevant to the question or provided answer. In essence, the interviewer probes further into the basis for the response given by the interviewee. As with the focus groups, these will be undertaken by NPMs.

194. Together, the focus groups and cognitive interviews will be used to study the response burden imposed by the context instruments. Due to the use of item sampling, there will be various permutations of each instrument. Steps will be taken to ensure that each permutation of the SSI and FSI can be completed within a maximum of 15 minutes.

195. The Expert Groups from all strands will have considerable input into the process of item development. This includes both thinking about how key reporting issues can best be addressed, and what particular wording in the questions is most likely to produce sound items with good psychometric (reliability, validity) qualities. The context surveys items will be considered at Expert Group meetings so there will be opportunities to interact face-to-face with the Expert Groups. In addition to the formal meetings, we anticipate informal communications with Expert Group members regarding discussion about specific items, and ways to get at important contextual information.

Quantitative testing

196. A small-scale trial test will be conducted to collect data to undertake an initial psychometric examination of the context survey items, and to provide further information that would help refine items and scales. Around 50 staff and 200 students will be asked to take part in this field test of the SSI and FSI. This will be conducted at two institutions in English in Australia, which is the first country to be involved in testing. Data will be entered, verified and compiled into files for analysis.

197. The instruments will be distributed to a heterogeneous sample of current learners. The faculty-focused instruments will be distributed to a sample of current faculty with teaching activities. For efficiency, these tests will be conducted online. Data will be entered, verified and compiled into files for analysis.

198. A range of psychometric analyses will be conducted to explore the characteristics of learners' interactions with the SSI items, the behaviour of the items, and relationships between items and target constructs. A suite of analytical approaches will be deployed to undertake the psychometric analyses. These include congeneric measurement modelling, item response modelling, and classical test analyses. The precise nature of the analyses will be tailored to the nature of items and instruments, and will include review of:

- links with population marker variables;
- item descriptive statistics;
- links between items;

- coding (and scaling for any composite variables);
- construct (internal, convergent and divergent) validity and concurrent (where possible);
- reliability (for any composite variables)
- response category performance;
- reliability generalisability;
- test processes for standard error calculation;
- differential item functioning;
- planned and unplanned item non-response; and
- response interference effects.

199. As possible given data characteristics, similar analyses will be done on the data collected through the FSI.

Final review and documentation

200. A range of technical reviews and developments will be undertaken to bring together the various validation activities, cross check the measurement properties of the instruments, and develop a range of resources for managing and analysing the items and instrument.

201. The final set of items will be proofed and cross-checked against project objectives and instrument specifications. The items will be reviewed in terms of the generic measurement criteria specified at the start of the development. The item mapping initiated at the start and managed through the development process will be verified.

202. A design scheme and series of templates and operational checkpoints will be developed to maximise potential respondents' engagement and response. Research on 'survey engagement' (Coates, Tilbrook, Guthrie and Bryant, 2006) has shown that people are more likely to respond to a questionnaire and provide a high-quality response if they see the resources and processes as a valid means of having their voice heard.

203. The Module D team will document the development and characteristics of the items. It will pay particular attention to outlining the content and construct validity of the items. This report will describe how the combination of survey items in the surveys and existing documentation achieves the overall measurement goals described in the framework. Instructions will be prepared for managing item sampling, implementation, analysis and reporting.

204. A codebook will be developed to manage the operationalisation of the items, assist with item sampling, underpin data file production, and guide analysis and reporting. This codebook will be documented and delivered to the OECD.

205. The team will gather the validated item library, support documents, and the SSI and FSI instruments and deliver these to the OECD.

Module management

206. Please refer to the Module D Work Plan in the Annexure for detailed information about specific objectives, focus and scope, schedules, milestones and deliverables, and personnel, responsibilities. This Work Plan pertains to only the first phase of the AHELO feasibility study.

References

- Andrich, D. (1997). A hyperbolic cosine IRT model for unfolding direct responses of persons to items. In W. J. Linden & R. D. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 399-414). New York: Springer.
- Astin, A.W. (1978). *Four Critical Years*. San Francisco: Jossey-Bass.
- Astin, A.W. (1985). *Achieving Educational Excellence: A critical analysis of priorities and practices in higher education*. San Francisco: Jossey Bass.
- Australasian Survey of Student Engagement (AUSSE). (2009). *Australasian Survey of Student Engagement*. Accessed 1 May 2009 from www.acer.edu.au/ausse
- Bottani, N., & Tuijnman, A. (1994). International Education Indicators: Framework, development and interpretation. In A. Tuijnman & N. Bottani (Eds.), *Making Education Count: Developing and using international indicators*. Paris: OECD.
- Bradburn, N. M., & Sudman, S. (1988). *Polls and Surveys: Understanding what they tell us*. San Francisco: W. H. Freeman.
- Carini, R.M., Kuh, G.D., & Klein, S.P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, 47, 1-32.
- Coates, H. & Ainley, J. (2006). *Graduate Course Experience, 2005: The report of the Course Experience Questionnaire (CEQ)*. Parkville: GCA.
- Coates, H. & Ainley, J. (2007). *Graduate Course Experience, 2006: The report of the Course Experience Questionnaire (CEQ)*. Parkville: GCA.
- Coates, H. & Edwards, D. (2009). *Engaging College Communities: The impact of residential colleges in Australian higher education*. AUSSE Research Briefing 4. Camberwell: ACER.
- Coates, H. & Edwards, D. (2009). *The 2008 Graduate Pathways Survey: Graduates' education and employment outcomes five years after completion of a bachelor degree at an Australian university*. Canberra: Department of Education, Employment and Workplace Relations.
- Coates, H. & Hillman, K. (2008). *Development of Instruments and Collections for the AQTF 2007 Quality Indicators*. Canberra: Department of Education, Employment and Workplace Relations.
- Coates, H. (2006). *Student Engagement in Campus-based and Online Education: University connections*. London: Taylor and Francis.
- Coates, H. (2007a). *Developing generalisable measures of knowledge and skill outcomes in higher education*. *Proceedings of AUQF2007 Evolution and Renewal in Quality Assurance*. Melbourne: Australian Universities Quality Agency.

- Coates, H. (2007b). Excellent Measures Precede Measures Of Excellence. *Higher Education Policy and Management*, 29(1), 87-94.
- Coates, H. (2007c). Universities on the Catwalk: Models for performance ranking in Australia. *Higher Education Management and Policy*, 19(2), 1-17.
- Coates, H. (2008). *Australasian Student Engagement Report*. Camberwell: ACER.
- Coates, H. (2009a). *Attracting, Engaging and Retaining: New conversations about learning*. Australasian Student Engagement Report. Camberwell: Australian Council for Educational Research.
- Coates, H. (2009b). What's the difference? A model for measuring the value added by higher education in Australia. *Higher Education Management and Policy*, 21(1), 69-88.
- Coates, H., Goedegebuure, L., van der Lee, J., & Meek, L. (2008). The Australian academic profession: A first overview. *The Changing Academic Profession in International Comparative and Quantitative Perspectives*. Report of the International Conference on the Changing Academic Profession Project, 2008. Hiroshima University: Research Institute for Higher Education.
- Coates, H., Tilbrook, C., Guthrie, B. & Bryant, G. (2006). *Enhancing the GCA National Surveys: An examination of critical factors leading to enhancements in the instrument, methodology and process*. Canberra: Department of Education, Science and Training.
- Codling, A. and Meek, V.L. (2006), Twelve Propositions on Diversity in Higher Education, *Higher Education Management and Policy*, 18(3), 23-47.
- Converse, J. M., & Presser, S. (1989). *Survey Questions: Handcrafting the standardised*
- Edwards, D. & Coates, H. (2008). *Graduate Course Experience, 2007: The report of the Course Experience Questionnaire (CEQ)*. Parkville: GCA.
- Ewell, P. T., & Jones, D. P. (1996). *Indicators of "Good Practice" in Undergraduate Education: A handbook for development and implementation*. Colorado: National Centre for Higher Education Management Systems.
- Ewell, P., Aamodt, P.O, Allen, J., Bettinger, E., Brennan, J., Federkeil, G., Kaiser, F., Llavori, R., McCormick, A., Ogata, N., Prades, A. & Torres Sola, J.M. (2009). *Analytical Framework for the Contextual Dimension of the AHELO Feasibility Study*. Paris: OECD.
- Ewell, P., Aamodt, P.O, Allen, J., Bettinger, E., Federkeil, G., McCormick, A., Ogata, N. & Prades, A. (2008). *Report on the 1st meeting of experts on the AHELO contextual dimension*. Paris: OECD. Francisco: Jossey Bass.
- Goedegebuure, L. & Coates, H. (2009) *Diversity in Australian higher education: an empirical analysis*. Australian Universities Review.
- Goedegebuure, L. Coates, H., van der Lee, J. & Meek, V.L. (2009). *International dimensions of the Australian academic profession*. *The Changing Academic Profession in International Comparative and Quantitative Perspectives*. Report of the International Conference on the Changing Academic Profession Project, 2009. Hiroshima University: Research Institute for Higher Education.

- Griffin, P., Coates, H., McInnis, C. & James, R. (2003). The development of an extended Course Experience Questionnaire. *Quality in Higher Education*, 9(3), 259-266.
- Jaeger, R.M. (1978). About educational indicators: Statistics on the conditions and trends in education. *Review of Research in Education*, 6, 276-315.
- Krause, K-L., Coates, H. & James, R. (2005). Monitoring the internationalisation of higher education: Are there useful quantitative performance indicators? In M. Tight (Ed.), *International Perspectives in Higher Education Research*. London: Elsevier.
- Klein, S.P., Kuh, G.D., Chun, M., Shavelson, R., & Benjamin, R. (2005). An approach to measuring cognitive outcomes across higher education institutions. *Research in Higher Education*, 46, 251-276.
- Kuh, G.D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change*, 33(3), 10-17, 66.
- Kuh, G.D. (2002). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. Bloomington, IN: Indiana University, Center for Postsecondary Research. Retrieved July 11, 2009 from <http://nsse.iub.edu/html/pubs.cfm?action=&viewwhat=Journal%20Article,Book%20Chapter,Report,Research%20Paper>
- Kuh, G.D. (2003). What we're learning about student engagement from NSSE. *Change*, 35(2), 24-32.
- Kuh, G.D. (2008). High-impact educational practices: What they are, who has access to them, and why they matter. Washington, DC: Association of American Colleges and Universities.
- Kuh, G.D. (2009). The National Survey of Student Engagement: Conceptual and empirical foundations. In R. Gonyea and G. Kuh (Eds.), *Using student engagement data in institutional research*. New Directions for Institutional Research, No. 141 (pp. 5-20). San Francisco: Jossey-Bass.
- Kuh, G.D., Chen, D.P., & Nelson Laird, T.F. (2007). Why teacher-scholars matter: Some insights from FSSE and NSSE. *Liberal Education*, 93(4), 40-45.
- Laing, J., Sawyer, R., & Noble, J. (1988). Accuracy of self reported activities and accomplishments of college bound seniors. *Journal of College Student Development*, 29, 362-368.
- Marginson, S., & van der Wende, M. (2007). *Globalisation and Higher Education* (No. EDU/WKP(2007)3). Paris: OECD)
- McCormick, A.C. (2009). Toward reflective accountability: Using NSSE for accountability and transparency. In R. Gonyea and G. Kuh (Eds.), *Using student engagement data in institutional research*. New Directions for Institutional Research, No. 141 (pp. 97-106). San Francisco: Jossey-Bass.
- Mullis, I.V.S., Martin, M.O., & Stemler, S.E. (1999). TIMSS questionnaire development. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), *TIMSS 1999 technical report*, Chestnut Hill, MA: Boston College.
- National Survey of Student Engagement (2008). *Promoting engagement for all students: The imperative to look within*. Bloomington, IN: Indiana University Center for Postsecondary Research.

- Nelson Laird, T.F., Chen, P.D., & Kuh, G.D. (2008). Classroom practices at institutions with higher than expected persistence rates: What student engagement data tell us. In J. M. Braxton (Ed.), *The role of the classroom in college student persistence*. San Francisco: Jossey-Bass.
- Nuttall, D.L. (1994). Choosing indicators. In; K.A. Riley & D.L. Nuttall (Eds.) *Measuring Quality: Education Indicators: United Kingdom and international perspectives*. London: The Falmer Press.
- OECD (2005). *Technical Report for the OECD Programme for International Student Assessment 2003*. Paris: OECD.
- Organisation for Economic Cooperation and Development (OECD). (2008). *Education at a Glance 2008: OECD indicators*. Paris: OECD.
- questionnaire. California: Sage Publications.
- Salazar-Clemeña, Rose Marie and Meek, V.L., eds., *Competition, Collaboration and Change in the Academic Profession: Shaping higher education's contribution o knowledge and research*, Manila, UNESCO and De La Salle University-Manila, 2008, 1-209 + vi.
- Schmidt, WH, & Cogan, LS. (1996). Development of the TIMSS Context Questionnaires. In Martin, MO, & Kelly, DL (Eds.), *Third international mathematics and science study technical report: Vol. 1. Design and development* (p. 5-1—5-22). Chestnut Hill, MA: Boston College
- Schwartz, S., & Westerheijden, D. F. (Eds.). (2004). *Accreditation and Evaluation in the European Higher Education Area*. Dordrecht: Kluwer Academic Publishers.
- Scott, G., Coates, H. & Anderson, M. (2008). *Academic Leadership Capabilities for Australian Higher Education*. Sydney: Australian Learning and Teaching Council.
- Siniscalco, M. T., & Auriat, N. (2005). *Questionnaire Design: Module 8*, Paris: International Institute for Educational Planning (UNESCO).
- Van Vught, F. & Kaiser, F., et al. (2008). *Mapping Diversity: Developing a European Classification of Higher Education*. Eschede: CHEPS.
- Van Vught, F. (2008). Mission diversity and reputation in higher education. *Higher Education Policy* 21(2), 151-174.
- Van Vught, F. (Ed.). (2009). *Mapping the Higher Education Landscape: Towards a European Classification of Higher Education*. Dordrecht etc.: Springer.
- Yamaguchi, J. (1997). Positive versus negative wording. *Rasch Measurement Transactions*, 11(2), 567.

MODULE E: PROJECT MANAGEMENT, SURVEY OPERATIONS AND SUPPORT FOR THE ANALYSIS OF THE AHELO FEASIBILITY STUDY RESULTS

Introduction

207. This section of the AHELO Assessment Design provides information on project management, survey operations, and on support provided for the analysis of the AHELO Feasibility Study results. It includes a broad overview of the Consortium's management structure and approach.

208. As specified in the Terms of Reference, the sections that follow detail how the team leading Module E will:

- implement an effective leadership and management process for the study that builds links, synergies and economies of scale between the various AHELO Feasibility Study strands of work – including the Generic Skills strand – and assists teams leading these strands to contribute to project management, survey operations and analytical support;
- develop an overall assessment design and related analysis plan, in cooperation with the OECD Secretariat and the AHELO GNE, that outlines the processes to establish quantifiable criteria for gauging the success of various facets of the AHELO Feasibility Study;
- verify national adaptations and the translations of assessment instruments and context questionnaires;
- engage institutions and sample students within participating institutions, giving focus to the type and level of support to be provided to participating countries;
- deliver the cognitive assessments and context questionnaires;
- implement survey operations – including training and supporting NPMs and ICs – that assure the quality of the cross-national, cross-cultural and cross-institutional comparability of the results from the AHELO Feasibility Study;
- implement the Generic Skills assessment in partnership with the Council for Aid to Education;
- prepare data, conduct scaling, and analyse results; and
- produce data products and written reports.

209. Quality control is an integral part of all assessment activities and is critical to help ensure that the AHELO Feasibility Study provides data that are comparable across institutions and subgroups. As the AHELO Assessment Design outlines, quality control procedures are involved in each part of the assessment and evaluation process, including test design and development, translation, sampling, data collection, coding, data cleaning and weighting, scaling, data analysis and the determination of feasibility. Hence this Assessment Design outlines an approach that will ensure high standards of process and outcome. As the following sections suggest, key facets of this include the implementation of well-tested methodologies, a phased management approach, transparent communication with key stakeholders, ongoing monitoring, training and support activities, the development of an explicit quality assurance plan,

strategies for engaging and managing the participation of key stakeholders, consultation with Expert Groups, and reporting to the Secretariat and AHELO GNE.

Project management

Project leadership

210. Associate Professor Hamish Coates is the AHELO Feasibility Study Project Director. In this role, Dr Coates devotes the vast majority of his time to the AHELO Feasibility Study, and is available for enquiries from and communication with the Secretariat on an ongoing basis. He is supported by individuals from the Consortium who will fill various roles specified in the following sections.

211. The different teams and individuals involved in delivery of the AHELO Feasibility Study work closely and are in communication on a daily basis or as needed to ensure:

- effective and efficient completion of implementation tasks;
- the consistent use of the highest quality standards in the work completed;
- openness and transparency in all dealings with each other, with AHELO Feasibility Study participants and with the Secretariat; and
- the application of strict financial controls.

212. ACER understands the unique governance arrangements that have been put in place to manage the AHELO Feasibility Study. ACER will report on a routine basis to the OECD Secretariat, report bi-annually to the GNE, receive advice and input from several Expert Groups, and work with NPMs to manage the assessment. We understand that the Secretariat will provide regular reports to the AHELO GNE, the OECD EDPC and the IMHE GB.

213. ACER will lead the team in a way that optimises technical and managerial synergies across strands. The team has been designed and developed in a collaborative fashion over the last year. Partners are aware of the strengths and complementarities across the group, with many having worked together on previous assignments. ACER will draw upon the extensive resources within the team to develop the most valid and effective technical solutions for the AHELO Feasibility Study. ACER will enter into subcontracting agreements, and provide regular formal and informal means for discussion and analysis.

214. The AHELO Terms of Reference call for advice on the impact of additional countries joining the assessment. This is a possibility, as several countries have expressed an interest in participating in the feasibility study. Three additional countries have already joined since the release of the AHELO Terms of Reference. Given the very short timeframe, ACER will work with the OECD Secretariat to identify a date beyond which new participants may not join the study.

Lead staff and areas of responsibility

215. The Consortium formed for the AHELO Feasibility Study comprises a diverse range of institutions and consultants from several different OECD countries. Below is an outline of the individuals and organisations involved in the Consortium and their broad roles in the project as a whole (incorporating all five modules).

The **Australian Council for Educational Research (ACER)** is taking primary responsibility for overall project management, including liaison with the Secretariat (Hamish Coates, John Cresswell). ACER will also coordinate Module C (Julian Fraillon), provide technical support for Module D (Wolfram Schulz), and assist with coordination and analysis in Module E (Alisdair Daws, Tim Friedman, Ali Radloff, Sarah Richardson, Wolfram Schulz). Dr Coates will ensure that, in accordance with the expectations outlined in the Terms of Reference, the team is represented at GNE meetings by appropriate senior staff when required. This may include virtual attendance where appropriate and where suitable facilities are provided at meeting venues.



cApStAn taking responsibility for translation verification operations (Andrea Ferrari, Steve Dept, Laura Wäyrynen, Raphaël Choppinet, Shinoh Lee).



The **Center for Postsecondary Research (CPR)** is providing input into the staff and student context survey instruments (Alexander McCormick, Bob Gonyea, Tom Nelson Laird).



The **Centre for Higher Education Policy Studies** is providing input via their coordination of Module D (Jon File, Frans Kaiser and Don Westerheijden).



The **Council for Aid to Education** providing input via their coordination of Module A (Richard Shavelson, Jim Hundley, Guillermo Solano-Flores, Amy Kurpius, Stephen Klein, Marc Chun, Scott Elliot and Jeffrey Steedle).



The **Educational Testing Service (ETS)** providing input via their coordination of Module B (Thomas Van Essen, Claire Melican, and Rick Morgan).



The **IEA Data Processing and Research Center** is managing fieldwork and data operations, and provide expert input into quality management procedures (Dirk Hastedt, Oliver Neuschmidt, Falk Brese, Tim Daniel).



Japan's **National Institute for Educational Policy Research (NIER)** providing a key Asian partner in the framework and test development process (Shuichi Tsukahara, Satoko Fukahori, Fumiko Yasuno).



SoNET Systems is managing the online systems required for item development and deployment (Mike Janic, Stephen Birchall).



Statistics Canada is taking primary responsibility for overall sampling arrangements, for variations to standard procedures required to support national and international options, and for weighting of data (Jean Dumais, Sylvie LaRoche).



The **University of Florence School of Engineering** is providing input on facets of the Engineering assessment developed in Module C (Claudio Borri, Alberto Tesi).



Westat Inc is undertaking structural equation modelling of the context survey data with a particular emphasis on exploring differential functioning across contexts (Frank Jenkins).



216. ACER recognises that for a development of this complexity and scale, it is necessary to draw on additional consultants, as required, during the course of the work.

217. ACER will establish contractual relationships with each of its partners and consultants. Development and administration of these contracts will be overseen by Hamish Coates in close liaison with ACER's legal services personnel, and this will be done in such a way to ensure full financial accountability of partner organisations and consultants through the lead contractor (ACER), with a view to satisfying the requirements of the lead contract that will be between ACER and the OECD, and under the leadership of the Project Director.

Management process

218. As lead organisation in the team, it is important that the arrangements be made within ACER to ensure a strong, continuing and well-supported focus on the AHELO Feasibility Study's management and implementation throughout the life of the project are clearly stated and understood. Within its research division, ACER has a number of research programs. Associate Professor Hamish Coates directs ACER's higher education research program. Core AHELO Feasibility Study staff will be located in this area. In addition, members of staff from other business units within ACER contribute significant time to other specialised AHELO Feasibility Study activities such as test and item development, IT support, database development, psychometric analysis, materials production services, financial management and support services, personnel services, library and information services, and so on. The AHELO Feasibility Study will be well staffed, centrally located within ACER's broader business arrangements and it is controlled according to the stringent financial management and accountability processes that are applied across the organisation.

219. ACER's AHELO Feasibility Study staff will work closely with those in partner organisations on every aspect of project implementation. In implementing the AHELO Feasibility Study, the lead staff from each partner organisation will meet regularly using email and telephone, as well as computer and video-based remote meeting technology. Collaborative technologies and project management software will also be used where possible to speed up communication and sharing of information.

220. Further, with such a diverse group of partners many of whom are in close proximity to one another, close co-operation, including physical meetings where required, will be arranged between key individuals and teams working together on particular aspects of the project.

221. Developing open communications with AHELO Feasibility Study participants will be given the highest priority among the team members. To this end, centralised communication via ahelo@acer.edu.au is being used as a key element in the communication strategy. A dedicated AHELO information management system, the AHELO Exchange, has been established (see: <https://ahelo.acer.edu.au>). The team will also develop and implement open, task-based communication protocols between participants and the key personnel within each organisation.

222. The ACER team has established regular teleconference meetings with the Secretariat. By this means or otherwise, the team will apprise the Secretariat of any consequences of the failure of countries to meet deadlines or otherwise conform with AHELO Feasibility Study technical standards at the earliest possible time, so that remedial action can be taken wherever possible.

223. The exchange of timely and accurate information is critical for the success of any complex project. As stipulated in the AHELO Terms of Reference, the team accepts responsibility for producing quarterly reports signed by the Project Director detailing the progress of the project (such as planned activities, major accomplishments, problems and their solutions, significant findings, noteworthy events, decisions that may be needed from the OECD Secretariat or the expert panels, and plans for the next reporting period). The team will also provide an annual report detailing project expenditures by task at the end of January of each calendar year. Using internal financial systems, projected versus actual expenses are easily monitored and early detection of potential problems is feasible. Any significant deviation will be brought to the attention of the OECD Secretariat for discussion and resolution. The team working together on Module E also agrees to produce the set of reports mentioned in the Terms of Reference. These include sampling plans, reports on survey procedures and quality, and a technical report. The teams leading Modules A, B, C and D will contribute to these reports.

Expert Groups

224. In accordance with the expectations defined in the AHELO Terms of Reference, the team has formed Expert Groups in relation to the Economics assessment (under the leadership of John Beath), the Engineering assessment (under the leadership of Emeritus Professor Robin King) and the broader Technical Advisory Group (TAG, under the leadership of Dr Peter Ewell). Given Dr Peter Ewell's expertise in this area, the contextual assessment will be overseen by the TAG.

225. The team has engaged many leading disciplinary, technical and higher education experts (see Table 1). Arrangements have been made to adapt group composition to accommodate new members from among potential nominees recommended from participating countries according to a process and timeline determined by the Secretariat and AHELO GNE.

226. Nominated experts will be respected and eminent in their field, and will have strong links with the relevant disciplinary or policy community in their countries and beyond, in order to maximise the technical input from a range of participating countries.

227. Senior staff of the Module E team will attend all Expert Group meetings, and meeting papers and records will be posted on the AHELO Feasibility Study website either before or as soon as practicable following each meeting. This will facilitate communication of decisions among the different groups, and awareness of issues and discussions taking place among members of different groups.

228. When it is appropriate, Expert Groups will be represented at meetings of National Project Managers (NPMs) to facilitate direct communication between these groups. In particular it is envisaged that the TAG chair will attend NPM meetings from time to time. NPMs will be at liberty to attend Expert Group meetings.

National Centres

229. It is important that participating countries are able to provide appropriate infrastructure for managing key facets of the AHELO Feasibility Study. The in-country infrastructure plays a vital role in supporting the development and implementation of the study. The Consortium's broad involvement in cross-national higher education assessment work suggests that not all countries have such facilities in place, and it is helpful to sketch what would likely be required.

230. First it is necessary for countries to establish a National Centre. These may be headquartered in government offices, or contracted to specialist service agencies. Given the autonomy of institutions, and the risk that AHELO may be viewed in some systems as a compliance mechanism or even funding indicator, the Consortium's preliminary advice is that the National Centre be established at 'arms length'

from funding or regulatory agencies. Of course, discussions about AHELO have been underway since 2006, and many national centres have already been established. While the import of these considerations will vary across contexts, the broader point is that National Centre should be able to engage institutions on a scholarly and quality improvement basis.

231. Whichever model is chosen, it is vital that the National Centre has the means of communicating efficiently with the OECD, government agencies, the Consortium, and institutions. This requires nuanced knowledge of the system, effective leadership capacity, an (optimally) established rapport with opinion leaders, and sound technical footings.

232. The National Centre requires standard office infrastructure, funds to support in-country and international travel, funds to undertake consultations in-country, access to document preparation and review facilities, staff or networks to manage translation and adaptation processes, the facilities to manage relationships with Institutional Coordinators (defined below). While the staffing of the National Centre depends on its location and history, core staff would include:

- National Project Manager (NPM) (see following section);
- Research Assistant;
- Administrative Assistant;
- Translation/Adaptation Advisor;
- Technical Advisor; and
- Editorial Support.

233. Indicative costs for in-country participation were provided in the commercial section of the Consortium's tender. While it is difficult to be overly prescriptive about costs without detailed analysis of national contexts, these estimates are provided to assist with national planning. These estimates assume the approach given in this AHELO Assessment Design, and are based on review of costs of leading in-country activities of a similar nature and scale to AHELO in Australasia, Asia, North America, the United Kingdom, and Europe.

National Project Managers

234. The National Project Managers (NPMs) nominated by each country will be the primary contact point for the team in the day to day dealings with that country. It is expected that NPMs will be assessment/survey specialists. NPMs will liaise with the team on all issues related to the implementation of the AHELO Feasibility Study in their country. NPMs play a vital role in ensuring that the AHELO Feasibility Study is administered in accordance with prescribed technical standards and survey operations guidelines, and in documenting processes implemented at national level toward the development of the AHELO Feasibility Study's final reports. Wherever appropriate, the same individuals may serve as representatives on the AHELO GNE and AHELO NPM.

235. The team sees the AHELO Feasibility Study's NPMs as filling at least two critical roles. The first is that they lead survey implementation in their respective countries. This is complex job, and one of the prime responsibilities of the team is to provide accurate, detailed and timely information to NPMs that guide their work on key implementation tasks. The team will use a variety of methods to communicate with NPMs, and to deliver the support and information they require to perform their functions. These methods will include:

- maintaining regular and direct email communication with NPMs as a group and with individual NPMs;

- using telephone and video-based meetings where this is appropriate;
- maintaining and developing an AHELO Feasibility Study website as a user-friendly resource and central tool for project implementation; and
- conducting briefing and training meetings at which key information is provided for discussion and review, and at which input from NPMs is obtained to assist the team in the conduct of its business.

236. The second major role for NPMs is to provide a channel through which national interests are represented in the implementation of the AHELO Feasibility Study. The team recognises the importance of providing proper opportunities for participating countries to shape project implementation, and will seek to maximise such opportunities. The use of surveys and review documents, providing an online discussion forum, and including meetings sessions that are specifically designed to collect national input and the views of NPMs are examples of the ways in which these objectives will be achieved. The team will place great emphasis on obtaining and using input and feedback from NPMs on such matters as the draft frameworks as they are developed, the proposed test and questionnaire items, operational issues with survey implementation, the conduct of meetings, workshops and training sessions, and of course NPMs are in a key position to provide the most informed advice on the survey data collected through the AHELO Feasibility Study assessment, and its interpretation.

237. Elaborated guidelines on proposed timelines and procedures for NPMs and other international and national experts to contribute to project development in a variety of ways, such as in submitting draft assessment items, are in communications and manuals distributed during the study.

238. Senior staff of the team will attend all NPM meetings, and meeting papers and records of all meetings will be posted on the AHELO Feasibility Study website either before or as soon as practicable following each meeting.

239. An important aspect of the team's work on the AHELO Feasibility Study's implementation with NPMs will relate to the role of the AHELO Feasibility Study technical standards. The team will develop and review these standards and will present proposed revisions of the standards firstly to the Technical Advisory Group then to the Secretariat for consideration by the AHELO GNE.

Institutional Coordinators

240. Each participating institution within a country will nominate an Institutional Coordinator (IC) whose role will be to liaise with the NPM, assist the NPM draw samples of students and faculty, and organise the administration of the AHELO Feasibility Study assessment and context surveys at institutional level.

241. The IC will provide the primary point of contact at each participating institution. As such, it is expected that between 100 and 200 ICs will be involved in the study (between 10 and 15 in each country), depending on the number of participating countries and institutions.

242. For the most part, the Module E team will not interact directly with ICs, but instead support NPMs to manage this function. Pending consultation with the Secretariat, GNE and NPMs, there may be value in taking steps to develop a community of practice among ICs such that in partnership with NPMs they can enhance their understanding of key contexts and processes. Given the relatively small size of the IC community and the short timelines for the study, the team will put procedures in place to manage this networking and ensure that it enhanced the efficiency and validity of communication.

Meeting arrangements

243. It is recognised that international meetings are expensive and should only occur if they are seen as essential, or the most economical means of conducting essential business of the project. Wherever possible, business will be conducted remotely, particularly when the tasks at hand involve only a small number of people. However, it is also recognised that project participants and the project more generally can benefit enormously from the interactions that can only occur during a face-to-face meeting.

244. An indicative schedule of meetings is included in the Work Plans reproduced in the Annexure to this AHELO Assessment Design. This schedule has been designed to maximise synergies across strands and enable efficient use of delegates and venues. The team will negotiate with the Secretariat on the conduct, timing, location and agenda of all meetings of Expert Groups and NPMs and any other significant project implementation meetings that might be planned.

245. Factors to be considered when determining the nature, timing, location and business of each meeting proposed will include the cost to participants, the cost to the project, convenience of the venue to participants, accessibility of transport to and within the host country, the suitability of available facilities, and the desirability of giving all participating countries the opportunity to showcase their country or institutions by hosting a meeting.

246. Meetings will generally be of two days duration, depending on the nature of the business to be considered. While keeping in mind the benefits of networking, wherever possible the agenda of these meetings will be arranged to minimise the time that individuals need to attend. Agendas and meeting papers will be distributed in advance of the meetings, and also made available on the AHELO Exchange. Delegates will be encouraged to be very selective about their attendance on particular days and at individual sessions in order to maximise the benefit they gain from attending and to minimise costs and disruptions to their busy work schedules.

Assessment design and analysis plan

247. The AHELO Assessment Design and the AHELO Analysis Plan will be produced early in the study. They will be delivered for approval by the OECD Secretariat and the AHELO GNE. When ratified, these documents will provide a significant guide for shaping the development and analyses that follow.

AHELO Assessment Design

248. This AHELO Assessment Design draws on a wide range of methodologies associated with international education assessment, large-scale educational evaluation, and higher education policy analysis. The AHELO Feasibility Study is landmark initiative, and the teams, experts national representatives from all modules will work together to deliver a design to the AHELO GNE that is cogent, sound and feasible.

249. The AHELO Assessment Design identifies the best methods and analyses for assessing the cross-national and cross-cultural validity of various assessments. Drawing on the material presented below, the assessment design specifies the AHELO Feasibility Study's:

- project management approach;
- population definitions and sampling strategy;
- translation, adaptation and verification approach;
- data collection plan including proctoring;

- delivery mechanism;
- coding procedures;
- data analysis plan;
- means of preparing data products and written reports; and
- overall quality assurance and risk management process.

250. This study's design advances methods for assessing the cross-national and cross-cultural validity of the AHELO Feasibility Study, and the kinds of quantifiable criteria that could be used to measure success. The overall assessment design optimises the synergies between the various strands of work through adequate coordination between the different teams and their work.

AHELO Feasibility Study analysis plan

251. The derivative analysis plan will be more specific and operational in nature. It will include specification of the research questions posed by the various strands, outline the best methods and analyses to assess the cross-national and cross-cultural validity of the instruments being used in the various assessments and context surveys, and chart the quantifiable criteria to assess the various dimensions of the feasibility study. The design and plan will be structured to deliver data and insights that assist countries decide on next steps.

252. A framework with criteria for establishing feasibility will be developed as part of the overall design. In line with the study's desired outcomes, this framework will separate feasibility considerations into those which are scientific in nature, and those which are practical in nature. Questions will be set, which will be underpinned by indicators that can be used to respond to questions about feasibility.

253. By way of example, the analysis of scientific feasibility will look at technical design and operationalisation matters such as:

- Was the assessment design itself valid and feasible?
- Was the study implemented in a technically appropriate fashion?
- Were Generic Skills, Economics and Engineering appropriate initial areas and fields to assess?
- Were the frameworks developed reflective of an international consensus about important learning outcomes?
- Were the instrumentations developed on the basis of the framework faithful to the spirit and intent of the framework?
- Were the item types used to assess the prescribed material effective from a measurement perspective?
- Do the test results stand up to defined psychometric standards of validity and reliability?
- Has the AHELO Feasibility Study been successfully generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally?
- Has it been possible to train people in different countries to score higher-order performance tasks?
- Have different educational and institutional contexts been effectively measured by the contextual assessments in ways that help understand differences in learning outcomes and demonstrate the analytical potential of the AHELO Feasibility Study?

254. An indicative list of the kinds of questions that will be posed to analyse practical feasibility – matters associated with contexts and implementation – includes:

- Was it possible to engage systems and institutions in the AHELO Feasibility Study?
- Was it possible to engage staff and students in the AHELO Feasibility Study?
- Has the AHELO Feasibility Study been received well by countries from a political perspective?
- Have the processes and near-term outcomes of the AHELO Feasibility Study been seen to add value within countries?
- Were the instruments delivered and scored successfully?
- Were promotion and institutional engagement activities successful?
- Has the process been seen to add value to institutional continuous quality improvement initiatives?
- Has the higher education community within each country been convinced that the AHELO Feasibility Study is a useful initiative?
- Is the AHELO Feasibility Study feasible from an economic perspective?

255. A range of evidence will be collected to help test scientific and practical feasibility. Formative evidence will be sourced and quantified throughout the study's reflective and collaborative design, development and implementation processes. More structured feedback will be sourced from participants – through context items administered to respondents during fieldwork. The GNE, TAG, Expert Groups, NPMs and ICs will provide rich feedback throughout the project, and structured mechanisms will be put in place to collect this and quantify it where possible. The team's analytical approach will deliver evidence of a psychometric and statistical nature.

256. An evaluation approach will be designed to guide the process of interpreting evidence in light of the specified criteria and reach a determination about feasibility. It is very likely in an initiative of this scale that the outcomes from this feasibility assessment will, like higher education itself, be multidimensional in nature. Hence the evaluation approach will provide a means of synthesising a vast and diverse amount of evidence into a productive means of determining whether the AHELO Feasibility Study has provided a proof of concept.

Translation, adaptation and verification

Approach to translation

257. This section describes translation, adaptation and verification processes designed to maintain cross-national comparability of assessment materials (Modules B and C) and context survey instruments (Module D). cApStAn is the Module E team in charge of linguistic quality control for the AHELO Feasibility Study. As detailed below, the Module E team will work closely with the Module A team, the Council for Aid to Education, to ensure consistency of process and outcome.

258. While the procedures envisaged for the different modules feature specific characteristics, which are described in the relevant sections of this AHELO Assessment Design, an integrated quality assurance and quality control approach to linguistic aspects of the AHELO Feasibility Study is desirable.

259. The methodology planned for AHELO is underpinned by the translation, adaptation and verification design that has been successfully deployed in PISA and also draws on experience acquired in other surveys. However, to verify the content of discipline-specific strands in Economics and Engineering,

additional input will be required: the translation verification should be carried out in dyads consisting of an expert verifier (linguist) and a domain specialist.

260. While the approach presented here is applicable to both paper-and-pencil and computer-delivered data collection instruments, it fosters computer-based materials through intense upstream preparation work by the team, with a view to providing a stable translation environment and adequate support to participating countries. The Module E team will endeavour to adapt an existing, user-friendly, fully integrated translation management system. This environment will be used to produce national versions of computer-delivered AHELO Feasibility Study materials but will accommodate paper-and-pencil tests. The Module E team would undertake the authoring of the national versions of new test and questionnaire items destined for computer delivery.

261. The set of procedures outlined here is structured in two distinct parts:

- quality assurance, which consists in defining processes that will ensure the production of national versions of instruments that meet stringent equivalence standards; and
- quality control, which checks whether the standards are met and proposes corrective action when they are not.

Quality assurance

262. Quality assurance includes the definition of the translation and adaptation design; the preparation of translation and adaptation guidelines; the early resolution of potential equivalence issues; training modules for and assistance to national teams of translators; and the provision of consolidated monitoring instruments in which both team recommendations and translation/adaptation issues addressed by countries are documented, followed up and archived.

263. The AHELO Feasibility Study test adaptation and translation procedures will include the following key components:

- A preliminary scrutiny of test materials by a panel (test developers, verifiers, domain specialists) to anticipate potential adaptation issues, ambiguities, cultural issues or item translatability problems. The information gathered at the earliest stages will be integrated in a centralised record of the translation and adaptation history of each item.
- AHELO Feasibility Study Translation and Adaptation Guidelines will be developed to assist national translation teams. This document will include general guidelines as well as item-by-item guidelines for Modules B, C and D. The latter will list adaptations that are mandatory, desirable or ruled out and be echoed in the centralised monitoring tool for translation, adaptation and verification. The item-by-item guidelines will draw the translators' attention to possible terminology problems, translation traps, literal matches (e.g. between stimuli and items), patterns in response options etc. As far as possible, these guidelines should be produced jointly by test developers, domain experts and linguists.
- As regards Module A, communication channels will be opened between CAE and the team with a view to developing synergies and so that CAE can benefit from team work done for Modules B, C and D, and vice-versa. Integration will be limited by contingencies associated with Module A starting translation ahead of other strands. Comparison of the two different approaches will be made and reported on.

- An appropriate file management system will be used to coordinate file transfer between participating countries. It is anticipated that this will be the Open Language Tool which has been used in the ERA option of PISA 2009 and is currently used for the PIAAC Field Test instruments. Guidance in the form of clear, concise and comprehensive instructions on how to use the SoNET environment or Open Language Tool for the purpose of developing national versions of the AHELO Feasibility Study instruments will be provided to (trainers of) translators. Based on lessons learnt from ERA and PIAAC, the team will prepare separate user guides for the different roles (translators, editors, content experts, reconcilers, verifiers).
- The double translation and reconciliation design will be a requirement. Participating countries will be asked to prepare two independent translations into their target language(s) of test instruments and questionnaire items. A reconciler will merge the two independent translations and prepare a final national version. Compared to more traditional procedures such as single forward translation or back translation, the double-translation design has definite advantages in terms of standards provided to translators, and early detection of translation difficulties: the two translators and the reconciler are three different persons working with both the source and the target version. It is unlikely that two translators would make the same mistake, and it is highly likely that the reconciler, if confronted with a correct and an incorrect translation of an item, will be able to exercise discernment. For a full discussion see Hambleton (2002) and Grisay (2003).
- Should some countries lack financial and human resources to comply with the double translation and reconciliation requirements, it can be envisaged to allow these countries to produce a single translation of part of the materials. Unlike in PISA or PIAAC, where usually there is only one verifier per national version (sometimes with a Maths or Science teacher to assist the verifier on certain issues), in AHELO there will be dyads of verifiers: a linguist plus an economist for at least the test materials of Module B, a linguist plus an engineer for at least the test materials of Module C. Under those circumstances, and, within the framework of a feasibility test, one can reasonably expect that verification will minimise the risk of a weaker translation in cases where the double translation plus reconciliation requirements were not met by a given country.
- While closed-response items will be machine-coded, open-ended items require complex coding guides. Experience has shown that unclear translations of coding criteria may lead to differential item functioning due to coding rather than response patterns. However, the translation of coding rubrics is less sensitive than the translation of assessment items. It is suggested that countries produce a single translation of the coding guides and have this translation vetted by their local domain specialists. The translation team would then undertake a thorough equivalence check of the translated coding guides to ensure that coders will understand them the same way in each country. In principle, adaptations to coding guides are not foreseen, but if exceptions need to be made they are typically to be adjudicated by item developers rather than by translators or verifiers.

Quality control

264. Quality control includes thorough verification of target versions against source versions; effective reporting of residual errors and undocumented deviations; expert advice in case corrective action is needed; a final check procedure, and standardised quantitative and qualitative reports on the extent to which the different target versions followed the translation and adaptation guidelines.

265. The linguistic quality control of national versions and the verification of their equivalence versus the English source version will be carried out by dyads, consisting of experienced verifiers appointed by

the team on the one hand and of domain specialists in Engineering and Economics for each language version on the other.

266. The verifiers are selected from cApStAn's experienced team: they are native speakers of each of the target languages, highly proficient in English, trained to assess whether translation and adaptation guidelines are followed and to document possible deviations, insert corrections as needed and provide expert linguistic advice. The verifiers are all familiar with the Open Language Tool Translation Editor, with the use of 'verifier intervention categories', severity codes and verifier comments in a standardised form. This makes it possible to produce reports that are useful feedback for the national centres and also provide a qualitative and quantitative proxy indicator of the linguistic quality of national versions and of their equivalence with the source versions.

267. The domain specialists will be selected through networking and will not be linked to higher education institutions that are being assessed. They will need to demonstrate up-to-date expert knowledge in higher education level Economics and Engineering in their country and will be trained to work in dyads with the verifier.

268. Verifiers' and domain specialists' verification feedback will be entered in separate cells of the centralised monitoring tool, thus usefully supplementing the translation/adaptation history of each item.

269. Once countries will have processed edits suggested by their verification dyad, they will release pre-final versions of their instruments and submit these to an international team of final check reviewers. This team will verify the final layout of the items and questionnaires according to a detailed checklist. While this final check takes place, the international verifiers will check correct implementation of those verifier and domain specialist interventions that had a high severity code.

Adaptations

270. Adaptations fall on a continuum, from simple differences in spelling or typographic conventions, or lexical differences that a competent native translator may be able to implement, to much more subtle differences that require expert understanding of the item demands. The major role in implementing adaptations to specific local contexts should be played by national staff with extended expertise in test development and in the domain assessed, rather than by external translators or verifiers.

271. Therefore, while the common aim of both the national teams and the Module E team should be to ensure the highest quality of the final instruments, the specific strengths of national experts would best be used in ensuring the linguistic fluency, cultural adequacy and terminology appropriateness of the materials for undergraduate students. The specific strengths of the group of international verifiers appointed by the team would best be used to monitor linguistic correctness of the target versions and their equivalence with the source versions.

272. The contextual dimension survey will contain adaptations that require extended negotiations between the national research team and the Module E team. Consequently, national adaptations have to be submitted for a review by the Module E team prior to verification.

273. Accurate documentation of all adaptations implemented in national versions will be an important requirement. This adaptation approval process must be fully built into workflow.

Instrument delivery

274. The Consortium will develop all items for delivery online and, in exceptional circumstances, by paper. On balance, the Consortium recommends relying primarily on online development and delivery, while at the same time building items that could be converted seamlessly into paper format if required.

275. It is important that items can be deployed on paper. Paper test instrument delivery has the advantage of being relatively uncomplicated and is a known entity that can and has been successfully used across all countries in a broad range of international assessments (such as PISA, TIMSS and PIRLS).

276. Paper delivery has several limitations, however, which strengthen the case for using online delivery. Paper delivery is arguably less authentic and provides fewer opportunities to deliver innovative assessments of problem solving and reasoning for higher education students on the verge of further study or entering the workforce. In the field of engineering, for example, most project-based work is conducted on computer using a range of software types and using electronic information sources. This is only likely to increase in the future with paper methods being further outmoded. The use of paper assessments also includes several forms of risk which can be controlled or eliminated with the use of contemporary online systems.

277. In contrast, therefore, contemporary online systems provide a robust technological foundation that enhances item development, item validation, formative collaboration and review, quality control, translation, assessment delivery, coding, data verification, data dissemination, and reporting. Building an item in the system that will eventually be used for its deployment, for instance, reduces design and handling risks, and enhances the efficiency of readying the item for deployment. It is possible to deliver the materials over the internet and for candidate responses to be captured and, in some cases, scored or coded immediately. Computer-based delivery can therefore dispense with the need for printing, distribution and return, and any associated transposition of the paper-based information to electronic form (such as through data entry). Typically the software development costs are borne by the international study centre (and indirectly therefore the OECD) and so this is a cheaper option for participating countries.

278. ACER and its Consortium partners have much experience designing, developing and deploying online assessment systems. The team has worked in many countries, and is aware of key complexities and solutions. The team explored the opportunities available to work with different software partners to deliver the AHELO assessments online.

279. The Assessment Master system developed by SoNET systems in collaboration with ACER has been selected for AHELO. The SoNET software has been engineered to support internationalisation, and further development will be undertaken for the AHELO Feasibility Study. A significant advantage of the SoNET system is its capacity to support a wide range of item types, its cost efficiencies, and its capacity to be developed quickly for AHELO. The SoNET system can also be administered online (using the internet to connect institutional computers to remote servers), through the local area network (LAN) within institutions, or using USB drive plugged into institutional computers. The Module E team recommends the use of the web delivery with USB as an in-country backup solution as it is likely to be the most robust and easiest to administer across institutions with a broad range of network infrastructure, resources and internet connectivity.

Engaging institutions and sampling students

Population definitions

280. Operationally, the AHELO Feasibility Study consists of three populations: countries, institutions and respondents. There are also sub-populations of respondents. Students are the main respondents, and there are also NPMs and ICs.

281. The Secretariat is responsible for the specification of countries. The team will support the Secretariat in this role. The AHELO Assessment Design is based on the countries listed in the Terms of Reference plus three others that have joined subsequently.

282. The AHELO Assessment Design assumes that a convenience sample of 10 or more institutions per country will be invited to participate in the study. These institutions should reflect the diversity of the higher education sector at national level.

283. Discussions between the Expert Groups, NPMs and the sampling team will help establish working definitions of what the AHELO Feasibility Study Terms of Reference refer to as “students approaching the end of their first undergraduate bachelor-type degree”. The team will use the Bologna degree structure as the central reference point. Populations will be established for Generic Skills, Economics and Engineering. Issues to consider include:

- the length of the program, e.g. how a three year degree relates to a four year or five year program;
- the criteria for inclusion in the population definition, e.g. how much Economics a student must have studied in order to be considered as majoring in this area; and
- differing course structures, and differing levels of specialisation, e.g. between ‘general’, ‘specialised’ and ‘combined’ Economics or engineering strands.

284. Given the specifics of each participating country and each participating institution, the population definitions will likely need to be tailored to each field and institution. Developing a structured approach to managing such complexity will be vital to ensuring the validity of the study’s process and outcomes. At this stage the Module E team will systematically record the institution level working definitions, and map these onto the wider national and international frameworks.

285. So as to maintain comparability among institutions and in order to avoid bias in institution-level results, working categories for student level exclusions will be developed and agreed internationally. Exclusions will be kept to the strict minimum and be justifiable. For example, the Terms of Reference flag part-time students for possible consideration. Students studying via an external/distance mode may also be potential candidates for exclusion. The internationally agreed criteria for, and level of exclusions will be discussed and determined with the OECD Secretariat and AHELO GNE.

286. The team will work also with the Secretariat, AHELO GNE, Expert Groups and NPMs to finalise a valid and effective approach to sampling university staff. Specifications for the population definition of faculty will be developed in consultation with NPMs and ICs. The specification will include a description of staff in the excluded population. The team recognises that it is difficult for many institutions to identify teaching faculty, particularly in a generalisable way, and the specifications will be developed to account for this. The work will take account of complexities such as sourcing up-to-date lists, identifying members of the target population, and human resources considerations.

Preparation of the AHELO Feasibility Study sampling plan

Selecting and engaging institutions in the study

287. Following the AHELO Feasibility Study Roadmap (OECD, 2009a) and Terms of Reference, the first-stage sample will be a convenience sample of higher education institutions. Guidelines will be provided to the NPMs to help them select ten suitable institutions. As much as possible, geography, institution size (as measured by, say, staff, enrolment numbers or annual budget), linguistic or socio-cultural community or any other nationally relevant factor will help determine which institutions should be invited to participate in the study. Questions of statistical efficiency or precision of national-level estimates are not relevant at this stage for this study, but in order to test the feasibility of implementation across contexts the aim will be to achieve as much representation diversity in the set of the ten invited institutions as possible. The feasibility of implementation may depend on factors such as the agreement of institutions and of departments response burden and response rates, timing of events, communications and transmission of documents, delineation of target populations, ease of access to administrative sources and other key features that affect how easily NPMs and Institutional Coordinators (IC) can perform their duties.

288. The success of AHELO depends on engaging institutions in the study. Hence it is of extreme importance to encourage participation at the institution and individual respondent level. To this end, a sense of ownership for the study among the leaders and decision-makers involved in institutions will be created. The team will develop an institutional engagement strategy that is designed specifically for higher education and AHELO. This strategy will be designed to support the selection of institutions, and also to sustain institutional interest and participation in the study.

289. This strategy needs, among other characteristics, to:

- be multilevel and engage the interests of leaders, faculty and students;
- involve academic and disciplinary communities;
- respond to contemporary trends and pressures facing institutions, staff and students;
- demonstrate and build consensus around the value of the AHELO Feasibility Study in research-based ways;
- provide a means of engaging with a wide range of stakeholders, perhaps via the Stakeholders' Consultative Group;
- provide a means of managing communication with a wide range of stakeholders – via development of the AHELO Feasibility Study website; and
- support the Secretariat in communication and dissemination activities.

290. The Survey Operations Manual will include example letters for each of the levels of decision-makers with advice on how to contact the institutions and convince people to cooperate. As part of NPM meetings, time will be devoted to the presentation, exchange, and discussion of strategies that have proved successful in the past.

291. National Project Managers will be asked to secure senior executive sign-off from each institution participating in the study. Such endorsement helps avoid unexpected institutional withdrawal from the study, helps ensure confidentiality and publishing protocols, helps reduce administrative delays, ensures that a senior executive can take institution-level responsibility for participation, and provides an appropriate level of endorsement for the study within institutions.

292. Once an institution has agreed to participate in the survey, an Institutional Coordinator (IC) will be appointed. Detailed instructions on how to appoint and to train the Institutional Coordinator will be given in the Survey Operations Manual. The IC's role is vital because they serve as the operational link between institutions and NPMs and need to communicate in both directions. Although the method of selection may vary greatly between systems, it is very important that there is sufficient input both from the institution and the national centre in the selection of this individual to be sure the person is capable of and willing to take over the responsibility for this task crucial to a successful implementation of the study. The person appointed as IC should be of sufficient seniority and experience to ensure the work is properly conducted in a manner that meets the international specifications. Typically, the IC would be a senior administrator in the institution, or a representative of the national centre with attributes that will ensure good entry and cooperation in the institution. The IC will be responsible for collecting information about the students in the target strands, and the faculty members. In close collaboration with the national centre staff, the IC will take care of the data collection, making sure that the response rates are as high as possible, and that the instruments are fully and adequately completed and collected, and kept secure.

Sampling students

293. Students are at the core of the AHELO Feasibility Study. Unbiased, probabilistic sample-based assessments need to be designed in each participating institution so that sound inferences can be produced.

294. Though a general sample design can be drafted from scratch, the possible solutions should be as similar as possible to the approaches that have been adopted for other assessment studies and to the solutions that would likely be implemented in a full-size study. However, the particular conditions of the feasibility study relax some of the sampling constraints usually imposed in this type of assessment. Rather than appoint an official Sampling Referee, for instance, the team will draw on advice and input from the advice and input from the Technical Advisory Group.

295. For sampling within institutions, a number of different approaches will be undertaken, depending on local factors such as the size of the institution, frame availability and institutional preference. For smaller institutions it may be most appropriate to undertake a census that includes all students fitting within the target population definition in the study. Where a sample of students is required, this can be efficiently done using equal size samples from each list, yielding an equal probability (and self-weighted) sample of students. Availability of enrolment lists will be crucial to a timely, successful and easy completion of this step. Other models might also be decided upon, for example the selection of class groups. In these cases, the sampling procedures and sample size will be adjusted to account for clustering effects and to maintain the effective size of the sample of students, though the clustering effects are not expected to be as large as they are often seen in assessments done at lower education levels. Past experience with school-based and higher education surveys will be used to adjust the sample size for clustering effects where this is deemed appropriate or needed.

296. However, local conditions may not yield themselves to a direct implementation of a general sample design. The local availability of up-to-date frame information for the sampling of students may vary quite a lot from country to country and from institution to institution within a country. Hence, adaptations of the general sample design to local conditions may be required. The team will provide support to NPMs and ICs to ensure that such adaptations do not jeopardise the validity of the study.

Sampling faculty

297. In many, if not most instances, and noting that the population at the institutional level will be small, a census of faculty will be conducted at participating institutions. Where a larger number of

individuals exist are included in the frame, the probabilistic approach used for students will be adapted for use with staff.

Sample monitoring and quality control

298. At least two NPM meetings and discussions will be organised between each participating country and the sampling team prior to data collection. The first round will serve to go over the details of the criteria to be used to select institutions, the identification of the programs of interest and the delineation of the target population in participating institutions, and an overview of the activities required for the sampling of students. The second round will be required to approve the adjustments suggested by the participants. A third meeting, following collection and the early processing stages, may be required to discuss with each participant the final status of their implementation. All discussed and agreed upon institution sampling plans will be documented by the sampling team. Copies will be made for the Secretariat and the respective participants.

299. The sampling team will provide the NPM with a step-by-step guide that will instruct the IC in the procedures to be taken to sample students and faculty and the documentation to be assembled for adjudication. IEA DPC software was successfully applied in most IEA studies but also in OECD TALIS to sample groups and individuals within schools and institutions. Moreover, IEA DPC software was used in the IEA TEDS-M study to sample finishing ‘future teachers’ within teacher preparation programs. The software allows the import of electronic lists of selected institutions, and it supports the printing of tracking forms and labels for all survey instruments (or alternatively for letters to be sent to online data collection participants) in various formats.

300. Labels will include the name (if a country’s data protection laws allow this) and the identification codes (IDs) of students and faculty members and other relevant information like the validation code. This code is important for data entry after paper and pencil data collection. Only the ID plus the validation code allows entering the data in the data entry system in order to avoid miss-punches during the data capturing process. The validation code is an important factor of quality and confidentiality assurance.

301. After survey administration and when the tracking information is entered (or alternatively) imported into the within-institution sampling software, the sampling and tracking information database can be exported and sent to the IEA DPC for further processing. The exported database will not include direct identifiers such as names but only IDs to fulfil data protection laws.

302. The database will be used to validate the selection of respondents and the administration of instruments as well as to calculate participation rates, exclusion rates, weight adjustments and estimation weights.

303. In general, the sampling process will be a random procedure to assure comparability of the data from various institutions and countries – as much as is possible given convenience sampling of institutions. Countries that would choose to select the samples themselves without the use of software provided by the IEA DPC would have to submit to the sampling team all supporting documentation (copies of paper forms and computer listings, seeds of random procedures, etc.) so that the qualities of the samples can be assessed. All selection steps need to be documented thoroughly (listing, sorting, random numbers, sampling intervals), so that the sample could be reproduced at any time. In addition, individual response and non-response on the various instruments need to be tracked. In order to conduct these tasks in a systematic way, specific guidance will be given to NPMs or specific software may be used.

304. After collection, the sampling team will verify that there is no discrepancy between the agreed plan and the actual implementation and that the collected units correspond to the selected units. The

sampling team will ask the participating country to justify or correct any discrepancy it would have found before the sample can be verified.

Sample size and participation requirements

305. As the sample of institutions is one of convenience, institutional non-response is not an issue for this feasibility study.

306. Since the feasibility study is a proof of concept, high precision of institution-level estimates need not be as important as it might be for a full-size study. Nonetheless, to allow for the possible rotation of test booklets and enough completed test booklets to construct and confirm psychometric scales, it is expected that at least 200 final-year students will optimally be required for each participating institution. If fewer than 200 people are enrolled in the final year of a program at a participating institution, 200 responses cannot be expected from that institution, and a census of all enrolled students should be conducted. Such situations will be dealt with on a case-by-case basis and a recommendation will be made to the concerned NPM and to the Secretariat.

307. The suggested sample size can be used by country representatives as a means to estimate field costs. At the stage of implementation, the size of the sample would be adapted to the local conditions and population size.

308. An allowance for non-response must be made. The acceptable minimum participation rate has not yet been established, but at least 75 per cent student participation (OECD standard for TALIS) if not 80 per cent (OECD standard for PISA) or even 85 per cent (IEA standard for TEDS-M) might be considered. This would push the sample size to between 235 (with response at 85 per cent) and 270 students (response at 75 per cent) per program, degree or institution, as agreed between the Expert Groups, NPM and the sampling team.

309. Institutional response rates will be computed and recorded by NPMs for weighting and analytical purposes. As well, NPMs will be encouraged to document and report the amount of effort and the steps taken to successfully convince institution and department heads to participate in the AHELO Feasibility Study. This evidence will be reported to the OECD Secretariat and shared among the NPMs as 'lessons learned'. Replacement of non-responding students should not be allowed.

310. The minimum size of the sample will be confirmed after the instruments are developed and adopted, their possible rotation pattern established, and a decision concerning the minimum participation rate is made by the TAG and GNE.

Documentation

311. The sampling team will work with other members of the team to prepare a Sampling Manual. This will provide details for the institutional sampling procedures. This manual will clearly describe all different sampling steps to design national sampling plans, select institutional samples, and identify students to be sampled. The instructions will likely be tailored to each country's specificities, bearing in mind the population(s) invited to the AHELO Feasibility Study in that country.

312. The sampling team will document all the decisions made in the course of its monitoring of the national implementations of the general sample design. Copies will be made for the Secretariat as well as for the Expert Groups and NPMs.

313. A report on the sample yield, the response rate and other topics related to the efficiency of the data collection campaign will be prepared for each participating country.

Survey procedures and operations

Test administration preparations

314. The Institutional Coordinators are responsible for all procedures associated with assessment and questionnaire administration within institutions. They will make preparations to administer the test according to the uniform procedures and scripts described in the Institution Coordinator Manual.

315. Major tasks to be covered for the test administrations and described in detail in the manuals include:

- identifying eligible students and faculty members;
- preparing for the testing sessions (scheduling, timing, etc.);
- administering testing material and context questionnaires as well as calculating preliminary response rates;
- returning the material and instructions for keeping test material confidential;
- ensuring that technical equipment for the online administration is available, installed and pre-tested; and
- ensuring that support staff are available in case technical problems occur during test administration.

316. For countries where data protection laws prohibit personal information being transmitted to a national centre, procedures are in place that will assure that data protection and confidentiality laws will be adhered without a negative effect for conducting the study.

Coding operations

317. In addition to multiple-choice items, the assessments for the different AHELO Feasibility Study strands will likely include constructed response items. The latter categories will need to be scored by human coders using a procedure that ensures reliability of coding and minimises variability in coding across coders, countries and languages.

318. In developing and implementing these elements the Module E team will make use of best practices that have been developed in TIMSS, PISA and PIRLS. These will be adapted and developed for the AHELO Feasibility Study in collaboration with each of the instrument development teams.

319. A seminar for the coders responsible of each participating country will be held. During this seminar, participants will be trained using practice papers and example material in order to standardise the coding process and the interpretation of student responses. A detailed coding manual and detailed coding guides giving clear definitions for each category of constructed-response questions will ensure the reliability and validity of the scores assigned. Analogous electronic coding procedures will be developed for computer-based assessments.

320. The training and management of the human coding of constructed response items is the responsibility of the participating countries and will be coordinated by the Module E team in partnership with teams leading instrument development. In each country, a random sample of assessment booklets selected by the within-institution sampling software solution provided by the IEA DPC will be double scored to assess coder reliability. Coder reliability statistics will be computed including percent agreement and as well as using additional reliability measures such as Cohen's kappa. Internationally accepted

standards (such as defined in TIMSS, PIRLS, PISA) of reliability will be used to evaluate the coding reliability.

321. As in other parts of the study, quality control is embedded in the whole process through:

- input control through the development of guidelines, training materials and manuals;
- process control through training and documentation; and
- output control by analysing the multiple coding and by audits where the activities of NPMs will be audited about plans for conducting the constructed-response coding, including questions about the qualifications of the coders, the number of coders to remain on schedule, and the training program envisioned and implemented for the coders.

Data capture and within-country verification

322. During implementation data will be generated and collected from different sources and at different stages of the assessment. There will be databases with information specific to each country, yet these databases will need to eventually be put together and combined in a sensible way so that the information collected across the participating countries can later be compared minimising the data manipulation burden, and an international database for use by OECD can also be created.

323. If the option of paper and pencil administration is followed, NPMs will be responsible for transcribing the information from the questionnaires and assessment instruments into computer data files.

324. In the case that the online test and context questionnaire administration options are applied, NPMs will be required to verify that the data of respondents entered to the system matches the selection of participants from the within-institution sampling software and that all coding codes have been captured.

325. The following databases are expected to be developed and used during field operations:

- a sampling database with one record for each student selected for the survey;
- a student background database including the data from the student survey;
- a faculty background database containing the responses from the faculty survey;
- a institution background file;
- different strand-specific assessment databases containing the original responses and main codes from the students to the Generic Skills strand, the Economics strand, or the Engineering strand; and
- different strand-specific databases with the double scored open ended item scores.

326. All of these databases 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. Because of the complexity of the different sources of data, it is essential to provide tools to the participating countries so that errors in data capture are corrected as close as possible to the source of the data.

327. To assist with data capture and data verification in the case that paper and pencil instruments are used, the IEA DPC will supply a data entry software and manual and will hold a training session on the use of the software. The IEA DPC will also provide countries with codebooks describing the structure of the different response files. The codebook information includes (among others) descriptive labels for the variables, the values, format, missing codes, and validation criteria. The codebooks and data files will be

structured to match the international version of the tests and questionnaires. This means that for each assessment instrument there will be a corresponding codebook, which serves as a template for creating the corresponding instrument data file.

328. Before sending the data to the IEA DPC for further data processing, countries will be responsible for verifying the data files using the data entry software provided by the IEA DPC. For this purpose the software will provide built-in procedures to conduct valid range checks, checks for duplicate identification codes within files and across files, and linkage checks between files. For example, the software will verify that each record in the background databases has a matching record in the sampling database.

Training of data collection staff and coders

329. Modern communication infrastructure such as emails, phone, fax and video conferences will be used for the day-to-day business, and regular face-to-face meetings will maintain best possible information and transparency between Module E team members in order to establish an effective project management.

330. However special training seminars will be needed to familiarise the national teams with major survey operation procedures in order to achieve the highest possible data quality.

Data management seminar

331. While NPMs can be briefed about the survey operation procedures in the course of the regular NPM meetings, an explicit training seminar will be scheduled for the technical staff. This data management seminar, which would be attached to a regular NPM meeting, should be scheduled about three months before the test administration period. It is targeted to the person who would function as the technical coordinator within the national AHELO Feasibility Study project management team. The seminar will prepare participants to conduct all data management related procedures of the AHELO Feasibility Study in their countries.

332. The seminar will consist of a mixture of presentations and hands-on practical assignments. Major topics to be covered include:

- introduction of within-institution sampling procedures and the documentation of all sampling steps including the tracking of response/non-response (students and faculty members);
- national adaptation of questionnaires and related codebooks; and
- the use of the software application for data entry and data verification of any paper and pencil materials; and
- the administration and evaluation of the online assessment system.

333. Training for Institutional Coordinators will also be undertaken. This will be facilitated by the NPMs. Clear guidelines, description and tasks as stated in the Survey Operations Manual and the Institutional Coordinator Manual will give countries the necessary input.

Coding seminar

334. As noted, a coding seminar will be scheduled. This training will be organised centrally and follow the 'train the trainer' approach. Exchange between the participants is an important aspect of the training. Furthermore, practical advice will be provided and examples will be drawn from the wide experience that the Module E team has in international surveys. The master trainer material being used will

cover two aspects: how to score the constructed-response questions; and how to train the coders in the participating countries.

Support manuals

335. The Module E team will provide the participating institutions with all necessary manuals for conducting the AHELO Feasibility Study. The following manuals will be provided:

- A Sampling Manual will provide details for the institutional sampling procedures. This manual will clearly describe all of the different sampling steps required to design national sampling plans and to select institution (as still necessary at this stage) and student samples.
- A Survey Operations Manual will include general instructions and, separated into different parts, will give an overview of all AHELO Feasibility Study related survey operation procedures. This manual will outline the role and responsibilities of the NPM and the NPM's team. It will include information about:
 - contacting institutions;
 - within-institution sampling procedures;
 - translating and preparing the assessment materials;
 - administering the test;
 - coding activities; and
 - data capture and data verification process.

336. A number of additional manuals will be provided or be included as part of a multi-part Survey Operation Manual. The following will be covered:

- The AHELO Translation and Adaptation Guidelines will specify the rules and procedures for translating the survey instruments.
- An Institution Coordinator Manual will specify all tasks of the appointed Institution Coordinator within each selected institution.
- A Survey Administrator Manual will specify the procedures and scripts to be followed during test administration.
- A Quality Control Monitor Manual will provide detailed descriptions on the quality assurance program of the AHELO Feasibility Study.
- A Coding Manual will be developed that is dedicated to organising the coding of the constructed response questions and includes information on:
 - coding requirements;
 - recruiting of coders;
 - preparing for coding;
 - training of coders and table leaders;
 - single and multiple coding of assessment instruments;
 - item types and coding requirements;
 - schedule for coding;

- multiple coding of assessment instruments set aside for reliability study; and
- problem solving.
- Coding Guides and Master Trainer Materials will be developed for each language, which include information about the development of rubrics/coding schemes and how they relate to the codes. These materials manual will cover all the questions that need to be coded with the coding schemes and also guidelines, general as well as specifics such as how to deal with spelling and grammar, repeated errors, superfluous parts in the answers, answers that are not like the examples given, and partial responses. The main topics covered include:
 - a description of the correct answer or answer elements;
 - examples or descriptions as the item writer expects a candidate to formulate the response;
 - examples of very differently formulated responses that also receive ‘full credit’;
 - general guidelines and criteria to apply to decide whether a response deserves ‘full credit’; and
 - similar descriptions as the once above for coding ‘partial credit’ and ‘no credit’.
- If online assessment is used, a Manual on Computer-based Assessment will be produced that contains instructions on using the online testing system.

Monitoring quality and survey procedures

337. A number of steps will be undertaken to ensure that the survey is conducted under standardised conditions using agreed procedures, including the provision of detailed procedural manuals, practical training sessions, and expert advice and support. NPMs will be assisted in planning and implementing each step of the project. The Module E team will systematically monitor the sampling, the preparation of the instruments, the coding of responses, and the data collection and data management.

338. The ultimate responsibility for the appropriate implementation of the AHELO Feasibility Study, however, rests with participating countries. Therefore a program is proposed, which will combine both the internal and external quality of the survey implementation monitoring. The following materials will be developed:

- NPM Quality Control Checklist;
- a schedule for interviewing NPMs; and
- a schedule for interviewing ICs.

339. Using the Quality Control Manual, each NPM will register the compliance of operations on the national level with the international requirements at each step of survey implementation. Especially, the NPM will need to review the quality and all deviations from the described procedures in relation to:

- identifying and selecting samples;
- translation and national adaptation of the instruments;
- assembling and distributing survey materials;
- processes of data collection;
- the security of survey materials;
- implementation of coding procedures; and

- data management (data entry and verification).

340. The NPM will be asked to report systematically to the IEA DPC at each occurrence of non-complete implementation of procedures, in order to discuss implications for the data and possible corrections.

Data verification and processing

Overview

341. The main objective of the process is to ensure that the data adhere to international formats, 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. The program-based data cleaning will cover the following steps:

- file structure and valid range checks;
- identification-variable (ID) cleaning;
- between-file linkage checks;
- implementing flow and filter edits;
- cleaning of background inconsistencies;
- re-arranging the file structure towards data analysis and application of general cleaning rules; and
- quality control cleaning.

Standardisation of the national file structure

342. The first step in the data processing at the IEA DPC is to verify the comparability of the national datasets with the international file structure. The following deviations from the international file structure will be identified:

- international variables omitted;
- national variables added;
- different variable length or number of decimal positions; and
- different coding schemes or out-of-range values (valid range check).

343. Together with the inspection of the national data files, the data documentation (i.e. data management forms and survey tracking forms) submitted by each NPM will be reviewed. As a result of this initial review, the IEA DPC will outline and implement necessary changes in the national data to make the files compatible with the international format. Any structural national adaptations of the participating countries will be documented.

Cleaning rules and procedures

344. The following standard cleaning procedures are typically applied to large-scale international studies. All steps will be discussed with the OECD Secretariat and the expert instrument development groups to verify if they make sense in the context of the AHELO Feasibility Study given the expected analyses and reports.

345. When each data file matches the international standard as specified in the international codebooks, the IEA DPC cleaning program will apply several standard cleaning rules to the files received from the country. These rules will be applied using software developed at the IEA DPC in order to report and to correct inconsistencies in the data.

346. Each problem will be labelled with a unique problem number, a description of the problem and the action taken by either the program or by the staff of the IEA DPC.

347. If problems are identified that could not be automatically adjusted, they will be reported to the responsible NPM so that original data-collection instruments and tracking forms can be checked to trace the source of errors. Wherever possible, staff at the IEA DPC will suggest a solution and ask NPMs to either accept it or to propose an alternative. Afterwards data files will be updated in order to reflect the solutions agreed on. Where the NPM cannot solve problems by inspecting the instruments or forms, a general cleaning rule will be applied.

348. After application of all automatic updates; remaining corrections to the data files will be applied manually, using a specially developed editing program for the data files.

The following types of problems will be identified, and if possible, corrected automatically:

- ID-cleaning, will address problems with identification, tracking, or other indicator variables within files (for example duplicate identification numbers (IDs) in one and the same file or inconsistencies between student participation and data availability);
- linkage cleaning will address problems with identification variables between files (for example inconsistencies in the linkage between institutions and faculty members: faculty members without institution level information or institutions without faculty members, etc.);
- background cleaning will include:
 - split variable checks which refer to questions for which the answer is coded into several variables (for these groups of variables it will be checked if e.g. only ‘yes’ is coded and missing codes appear in the data, because according to experiences with previous studies it can be assumed that the missing responses actually mean ‘no’ in most cases and therefore will be recoded accordingly);
 - filter-dependent checks will consider cases where questions are introduced as filter questions and are followed by dependent questions, and the following rules will be applied: if the dependent questions are answered and the filter question is coded to ‘no’ or to missing then the filter is recoded to ‘yes’ or applicable; if the filter question is still coded to ‘no’ then all dependent questions are recoded to a newly introduced missing code ‘not applicable’; or if the filter question is ‘omitted’ and the dependent questions are coded to ‘not administered’, then the dependent questions are recoded to ‘omitted’; and
 - inconsistency checks, in which background data files are checked for logical consistency of the given responses.

Computation of weights

349. In parallel with the preparation of the data file, the sampling, data preparation and analysis teams will collaborate on the estimation of student-based weights for each strand and institution. These weights will be adjusted for student non-response within institution or program (as required by the sampling plan adopted in each institution).

350. The approach to variance estimation will depend on the sampling approach adopted within institutions. Where a simple random sample of students is a feasible approach, variance estimates will be obtainable from procedures available in standard software packages such as SPSS or SAS. For complex designs involving clustered samples of students and/or unequal selection probabilities and weights, replication methods as used in other studies such as PISA, TALIS and TEDS-M will be more appropriate. To ensure a consistent approach to analysis across institutions within a country, or across countries, it would likely be appropriate to develop replicate samples and apply a replication method to variance estimation for all institutions. For most simple statistics (such as means, totals and ratios of totals) the value of the sampling error estimated by replication will be identical to the value obtained using non-recursive formulae.

Implementation of the Generic Skills assessment

An advanced relationship

351. This section details how the ACER-led team leading Module E will collaborate with the CAE to implement the Module A generic skills assessment. The CAE team will work closely with other members of the ACER-led Consortium, and the Module E team in particular. To enhance the consistency, efficiency and validity of implementation, it will be vital to integrate and aligning various practical and technical facets of the development. By way of example, the Module E team will work closely with CAE to ensure that essay rating procedures align with coding procedures used with other assessments, that rubrics are adapted appropriately and internationally generalisable, and that translation and national adaptation processes align with other aspects of the project.

Development of an analysis plan

352. Along with teams leading other modules, CAE and ACER will collaborate to develop an analysis plan. In line with the overall design work for Module E, this plan will identify the best methods for assessing the cross-national and cross-cultural validity of the CLA instrument, and develop with quantifiable criteria to measure success.

Fieldwork preparation and management

353. CAE and the ACER-led team will begin planning for fieldwork at the start of the project. Preparations for the testing phase will be undertaken as part of the overall program of work and include:

- defining the student target population in the Generic Skills strand, selecting a convenience sample of institutions, and in drawing random samples of students within participating institutions;
- delivering two translated performance tasks (as per separate CAE contract with the OECD);
- making available the translated performance tasks; and
- proctor training by CAE with NPMs and one bilingual (English speaking) testing coordinator from each participating institution.

354. The Module E team will work closely with CAE to ensure that development processes and task characteristics align with the processes outlined above. Key links will be made: with the assessment design and analysis plan; with translation, adaptation and verification procedures; with the approach used for engaging institutions and sampling students; with overall survey operations and procedures (particularly the development of rubrics and coding); and with the analysis of results.

355. Preparation for coding will begin when countries finish testing. In close collaboration with the Module E team this will include:

- agreeing on coding criteria (rubrics and scoring guidelines) for performance tasks and translating into country languages;
- CAE, in consultation with NPMs, selecting a set of benchmark papers for coder training from each country (evidencing the full range in response quality for each performance task);
- translating country papers into English for common coder training;
- CAE recruiting a Chief Reader/Scoring Leader for each participating country/testing language. The Chief Reader/Scoring Leader is trained to:
 - use the rubrics and score student responses reliably; and
 - identify and resolve issues of coding rubric equivalence across languages and cultures.
 - More specifically, this involves:
 - CAE training coding trainers (Chief Readers/Scoring Leaders) and coders in English and the testing language using a common set of training papers and coding criteria at a series of two day training sessions in each individual country;
 - CAE working with coding trainers (Chief Readers/Scoring Leaders) to identify, discuss, and resolve issues on coding that result from the lack of entire equivalence across cultures and languages, such as the interpretation of rubric descriptions in English and in the translated languages and the equivalence of sample responses across cultures;
 - CAE continuing training until adequate inter-coder reliability with trainers is achieved;
 - with guidance and monitoring from CAE staff, coding trainers (Chief Readers/Scoring Leaders) training monolingual coders in their respective native languages. For example, the Korean coding trainer (Chief Reader/Scoring Leader) would train Korean-speaking coders to score Korean responses. For training, they would use the coding criteria (translated into Korean) and the set of Korean papers used for common training (which would also work if bilingual coders were used); and
 - training will continue until adequate inter-coder reliability on the common set of papers is achieved.

356. The testing phase involves CAE, working with its subcontractor Internet Testing Systems, overseeing testing operations at the ten institutions in each of the participating countries. This will include:

- internet delivery of tests in Korea, Kuwait, Finland, Norway, Mexico and USA;
- on-line grading of tests by trained coders working remotely in their home or work location;
- linking seamlessly with the context instruments developed for use with the Economics and Engineering assessments; and
- compiling a database of test results.

Online platform for the generic assessment

Overview

357. As noted, ITS will be a supplier to CAE for the OECD program. A key benefit of selecting ITS is that the Performance Tasks and the delivery interface currently reside on the ITS system. All that is needed for this project is to add the localisation for the five non-English languages. To do so ITS will create translated versions of the web sites, examinations and coding system. Only a single context instrument will be used for the study, so the ITS system will link to the SoNET system so that respondents can provide relevant contextual information. This mirrors an approach deployed by CAE and ACER in prior international assessment activities.

Online test delivery

358. The ITS test delivery retrieves each question delivered to the student from a central server over the internet. With this model each student response is recorded centrally each time they navigate from one item to the next or, at a minimum, every thirty seconds. The key benefit of this approach is that no exam data is lost should a disruptive event occur (such as testing workstation failure, loss of power or loss of connection to the internet). In that event the exam will be restarted from the point of interruption as soon as the disruption is resolved.

Human coding

359. The human coding interface allows coders to access student responses and then guides them through their evaluation. Results are then provided to CAE for reporting purposes. CAE will create a localised version of this coding interface for each language.

Content delivery networks – Akamai

360. An important consideration for international programs is the latency (delay) that can occur when the testing locations are at some distance from the host servers. One proven method of minimising latency is the use of a content delivery network.

361. A content delivery network or content distribution network (CDN) is a system of computers networked together across the Internet that cooperate transparently to deliver content to end users. When a request is made to a system using CDN's, the request is actually sent to a local edge server. This server processes the request, and determines whether any of the data requested is cached, and if not, whether a standard Internet route or a private route is faster.

362. The ITS online test delivery system supports using the Akamai Content Accelerator Network. Akamai provides ITS with several thousand edge servers around the world, and the use of caching can be configured down to the program level.

Testing centre (institution) verification

363. The ITS system includes a 'system check'. Run on individual workstations the system check automatically evaluates upload and download speeds, launches a sample test, runs a trace route test and provides feedback on the system readiness including:

- browser compatibility;
- operating system compatibility;

- monitor display settings;
- internet connection suitability which, in addition to speed reporting, includes an estimate of the number of concurrent tests that may be delivered using that internet connection;
- HTTP latency;
- HTTP compression;
- cookies readiness; and
- flash readiness.

Student identification and tracking

364. The ITS system supports the use of a key identifier for students. For the CLA program the student's email address has been used to uniquely identify and track each student and their testing history. This may be changed to another identifier if OECD desires. Every time a student tests they must login to the system. Logging in links their email address to each of their test result files.

Administration

365. The ITS system allows administrators to authorise users and control their level of system access. A user with administrator rights has the ability to create users at that institution. These users are typically test proctors. Each user is created with a unique ID, normally their email address, and a password. At initial login the user is required to change their password. The use of ID and password authenticates each user. Students are authenticated as described in the Student Identification and Tracking section.

366. Student confidentiality is maintained online by encrypting all questions and responses. For proctoring convenience the student's name is displayed at the top of the testing screen. This allows proctors to verify the correct student is taking the test.

367. Online tests can be suspended and resumed as desired. Proctors can pause the test through the use of an 'administrator's key sequence' at the student's workstation. The test driver is configured to check for an existing test record, against the student's ID. If one is found it is restarted from the point it was paused.

368. ITS has a secure browser for both PC and Macintosh. A test will not start unless the secure browser is detected. The secure browser locks down the workstation, disabling most function keys and key sequences, such as Ctrl-Alt-Del, Alt-Tab, Ctrl-C, Ctrl-V, etc. Students are unable access other applications, the local hard drive, the local network or the internet. The secure browser does not require installation or system administrator rights.

Analysis and reporting

369. CAE's measurement science staff will assist the team leading Module E with coding, the analysis of test data and overall determinations of feasibility. CAE and the Module E team will provide technical and analytical support to the OECD Secretariat during the preparation of the final report summarising the results of the feasibility study, including the preparation of analyses, tables and technical documentation.

Analysis of results

The scaling model for test data

370. The ACER team proposes to use scaling methodology based on item response theory (IRT) for the AHELO Feasibility Study. This methodology is widely used in international studies in the field of education and enables researchers to assess scaling characteristics. It is also proposed to place special emphasis on the review of the cross-cultural comparability of all instruments used in this study.

371. More specifically, the team proposes that the Rasch model, in its general form as implemented in the ACER ConQuest software, be used to analyse the AHELO Feasibility Study assessment data. The use of this analytic method is proposed because:

- of all available item response theory models, it provides the strictest assessment of psychometric validity;
- it supports the construction and validation of meaningful described proficiency scales, which are taken as a requirement for the useful reporting of the AHELO Feasibility Study performance data;
- it has been widely generalised to deal with the range of analytic requirements of complex cross-national studies similar to the AHELO Feasibility Study (such as exploring and controlling for coder effects and item position effects, and supporting multidimensional scaling); and
- it also supports equating tests for the purposes of maintaining and monitoring item sampling and trends.

372. As ACER ConQuest has in part been enhanced for use in the OECD PISA studies, the team will make the software available to the AHELO Feasibility Study participants and is willing to provide support to participants in their own use of the software.

Analysis of test items

373. At the end of the fieldwork, and after data cleaning and scaling, preliminary reports will be presented at national project managers and Expert Group meetings. These reports would include, among others, the following results:

- preliminary item statistics (IRT and classical item statistics);
- statistically undesirable characteristics in the items;
- test targeting and dimensionality;
- item-by-institution and item-by-language interactions; and
- analysis of non-systematic missing data.

374. Item fit will be assessed using a range of item statistics. The weighted mean-square statistic (infit), which is a residual based fit statistic, will be used as a global indicator of item fit. Weighted infit statistics will be reviewed both for item and step parameters. The analysis of item fit and the estimation of item parameters will be carried out with the ACER Conquest. In addition to this, item characteristic curves (ICC) will be generated for every item, which provides a graphical representation of item fit across the range of student abilities for each item (including dichotomous and partial credit items). The functioning of the partial-credit coding guides will further be reviewed through investigation of the proportion of responses allocated to each response category and the differences in mean abilities of students by response category.

375. In international studies the issue of cross-national validity of instruments is of crucial importance in order to ensure that data are comparable across institutions and languages. In the AHELO Feasibility Study, such analyses will play a central role, for they will help identify the extent to which the assessments have been successfully generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally. The cross-contextual validity of the test items will also be explored by assessing differential item functioning (DIF) (for groups that have sufficient sample). Specifically, IRT will be used to detect variance of item parameters across contexts. Such variance indicates that groups of students with the same ability have different probabilities of responding correctly to an item. This is commonly referred to as 'item bias' as it indicates that the probability of successful performance is a function of group membership as well as individual ability.

376. The extent to which missing student responses are due to problems with test length ('not reached items') will be reviewed to assess the appropriateness of test length. Unreached responses will be defined as all consecutive missing values starting from the end of the test except the first missing value of the missing series, which will be coded as 'missing'.

Analysis of context items

377. The contextual instruments will use different item types from those in the test instruments, and it is envisaged that a large proportion of them would be rating-scale (Likert) items. Different constructs will be measured with sets of Likert-type items (5-10 items). Another category of items will measure background information (age, gender, family background).

378. Review of percentages across categories will be computed and will provide information about the possible skew of items and the amount of missing responses. In principle, preference should be given to items that have sufficient percentages in each category. Distributions of core background variables will also be compared with those from previous surveys both the international and national level.

379. For items that are designed to measure latent constructs, it is recommended to use exploratory and confirmatory factor analysis to review the dimensionality of items. Once the dimensionality has been explored, classical item statistics (reliabilities, item-score correlations) as well as IRT-based statistics will inform further the scaling characteristics of these items. Scale scores for these constructs would be derived using the IRT partial credit model. This is similar to what is done in international large-scale assessment like PISA.

380. Validity, reliability, DIF and scale precision will be provided in psychometric reports describing the technical properties of the items and scales developed for the contextual analysis. DIF and validity assessment will also be conducted on assessment forms in the various languages to determine whether and to what extent language and cultural differences impact the instruments' measurement properties.

381. For the context survey data, a general descriptive analysis will be performed in which cognitive outcomes are disaggregated by major student groupings and institute characteristics within country and subject area. The outcomes will encompass cognitive measures (Economics and Engineering), institutional characteristics (size of department/faculty/institution, selectivity, research orientation, ethnic profile) and student characteristics (gender, ethnicity, SES, prior academic level, etc).

382. In order to glean a more nuanced understanding of the data and test for differential performance across contexts, we propose employing multilevel structural equation model (SEM) approach (Muthen, 2002, Raudenbush & Bryk, 2002, Brown & Rasbash, 2004). This approach will enable the Module E team to estimate how performance is related to key factors at various levels of nesting, students, departments, institutions and countries. Such a model would most directly address the question of the nature of

contextual effects and the value added by individual institutes. Contextual and confounding variables are accounted for at each level of nesting while the design effect of clustering is taken into account for the estimation of standard errors. These models not only describe the relationships between context and outcomes, but they also show how some effects vary from institution to institution, and ultimately enable the team to identify sources of variation.

383. Two types of multilevel SEM models will be employed. First, multilevel regression with random effects modelling will be conducted (Raudenbush & Bryk, 2002). In this model, individual, department and institution factors predict student outcomes, while controlling for confounding factors occurring at each level of aggregation. If some prediction effects are determined to vary across institutions, these effects are themselves predicted by contextual variables to determine which precursor, department and institute variables systematically account for the variability.

384. Second, multilevel SEM models with a network of paths (Muthen, 2002) will be run. Multilevel SEM models can go beyond regression models in that a network of relationships among factors can be described including mediating effects as well as direct effects of predictors. A network of paths can be used to test more complex hypotheses of how a pedagogical system functions. This modelling is multilevel in that a within-institutes and a between institutes path model can be posited. Also the variation of within-institute path coefficients can be modelled. Robust standard errors will account for clustering in the outcomes.

385. Given that the data come from a feasibility study, the results would not allow any inferences to the underlying populations of institutions and students. Multilevel analysis would be carried out using pooled international samples to ensure sufficient numbers of clusters at the level of institutions.

Data products and written reports

386. To support data analysis and assist an effective international dissemination of the results from the AHELO Feasibility Study, the team will provide the products and services described in the following paragraphs.

AHELO Feasibility Study database and codebooks

387. The team will provide a fully documented database that will be delivered to the OECD Secretariat by the end of the study. This database will include all student proficiency scores; all final weights and replicate weights for sampling variance computation any context composite indices derived from the questionnaires, together with students' responses to the questionnaire and the test questions. This database will allow the OECD Secretariat, the AHELO GNE, NPMs and institutions to conduct their own further analyses. Data will be provided in ASCII format and in SAS and SPSS format. In addition, codebooks will be provided for all databases that list the variable names, variable labels, format, column position in the ASCII data files, categories and labels.

AHELO Feasibility Study compendia

388. Compendia will be prepared and made available in Adobe PDF, MS Word and MS Excel formats. Compendia will be developed that include a set of tables showing statistics for every item in the questionnaires, and the relationship of background variables with performance. The tables will show the percentage of students per category of response and the average performance by assessment and domain for the groups of students in each category.

AHELO Feasibility Study Technical Report

389. The team leading Module E will be responsible for preparing the AHELO Feasibility Study Technical Report. The team will design a comprehensive technical report and draft materials pertaining to Module E. As well, the team will collaborate with the teams leading Modules A, B, C and D to incorporate material from the technical reports prepared as part of this work. A draft of this report will be delivered to the OECD Secretariat for review, followed by a final form by the end of the study.

390. The AHELO Feasibility Study Technical Report will summarise all aspects of the AHELO Feasibility Study. It will clearly describe all data and statistical conventions and approaches applied in the study, information on matters of test and questionnaire design, field operations, sampling, data adjudication and quality control mechanisms, methodologies used to analyse the data and other technical features of the project will be described at a level of detail that would allow other researchers to understand and replicate its analyses.

AHELO Feasibility Study Institution Reports

391. Informative and well-targeted reports will be prepared for each participating higher education institution. Such reports would provide a major means of developing the value of the AHELO Feasibility Study to institutions, and considerable emphasis would be placed on their design and production. These reports will provide institutions with additional information on the capability of their students. They will be designed to supplement and enhance existing data on student achievement, graduation rates, industry feedback, teaching quality.

392. The production of AHELO Feasibility Student Institution Reports can be managed in various ways. Reports for each institution could be produced centrally by the Consortium and distributed to NPMs. Alternatively, a report template and manual could be produced and delivered to NPMs who could use this as a basis for reporting to institutions in their country. A third innovative option, which could usefully underpin the previous two, would be to use the online assessment system to generate programmed institution reports. These reports could be accessed in a secure fashion by NPMs and used as a basis for developing their own customised versions.

393. It is very important to be clear with institutions about the nature and use of the AHELO Feasibility Study data and results. To this end, a code of practice will be developed to provide guidelines on how AHELO Feasibility Study data should be used, as well as provide ideas on ways to use the data and results most effectively. As noted in the Terms of Reference, the team understands that participating institutions will sign an agreement to ensure that AHELO Feasibility Study results are not to be released in the public domain.

394. The team will draw on its expertise developing through related national and international assessments to produce documentation that includes, for instance:

- an executive summary of results and overview of the AHELO Feasibility Study;
- protocols for appropriate use of results;
- institution-level results and profiles;
- disciplinary benchmarks (that maintain the confidentiality of other institutions);
- breakdowns for any contextual groups (as available given data availability and privacy considerations);
- recommendations for continuous improvement; and

- summary feedback sheets for leaders, faculty and students.

395. Institutions and programs are the level of analysis in the AHELO Feasibility Study, and no comparative data at the national level will be provided.

396. Each AHELO Feasibility Study Institution Report will include a copy of the full dataset for a given institution. As one of the deliverables specified in the AHELO Feasibility Study Terms of Reference, a well documented institutional database containing all cleaned data files of the participating countries including accompanying data products will be created. The Module E team has expertise with many databases and database systems (for example SAS, SPSS). A database architecture and technology will be chosen, in consultation with the OECD Secretariat and the Expert Groups that is cost effective and can be quickly queried and extracts harvested for analysis. It will be ensured that:

- information coded in each variable is internationally comparable;
- national adaptations are reflected appropriately in all variables;
- questions that are not internationally comparable have been removed from the database;
- all entries in the database can be linked to the appropriate respondent;
- sampling weights and any scores as well as derived indicators are available for international comparisons; and
- accompanying documentation listing national deviations from the concerning international version of the background instruments will help analysts interpreting their results correctly.

AHELO Feasibility Study Report

397. The team will start planning the AHELO Feasibility Study report early in the project during development of the assessment design and analysis plan. An initial structure will be developed through this collaborative planning process. This will provide a basis for subsequent consultation and development. The team will meet with Secretariat staff at routine intervals to update them as the draft takes shape. The team will adopt all formats and guidelines on style and content that are provided by the OECD Secretariat.

398. In particular, the draft final report will consider:

- methodological and technical questions raised by an international AHELO – including issues of domain definition, conceptual assessment frameworks, validity of instruments, translation and cultural adaptation, field implementation, coding, scaling and reliability of results, data analysis;
- issues that arose during implementation and in the analysis of results; and
- conclusions on the scientific and practical outcomes of the feasibility study as well as guidance for the longer-term development of an AHELO should the initiative be taken forward.

399. The OECD Secretariat will be responsible for the preparation of the report that will summarise the results of the AHELO Feasibility Study, based on the analysis plan and analyses prepared by the contractor. In general, the team will provide technical, statistical and analytical support to the OECD Secretariat during the development of the AHELO Feasibility Study report according to a program and schedule that is to be negotiated between the Secretariat and team. As detailed in the Terms of Reference, the team will support the OECD Secretariat in the preparation of the AHELO Feasibility Study report in the following ways:

- provide statistical and technical support to the OECD Secretariat during the development of this report, concerning aspects and results related to all components of the AHELO Feasibility Study;

- design tables and describe analysis, in co-ordination with the OECD Secretariat;
- review the final AHELO Feasibility Study report prepared by the OECD Secretariat for technical consistency and coherence; and
- review draft reports that other contractors complete.

Module management

400. Please refer to the Module E Work Plan in the Annexure for detailed information about specific objectives, focus and scope, schedules, milestones and deliverables, and personnel, responsibilities. This Work Plan pertains to only the first phase of the AHELO feasibility study.

References

- Braun, H. I. (2005). Value-Added Modeling: What Does Due Diligence Require? in R. Lissitz (Ed.), *Value Added Models in Education: Theory and Applications*, pp. 19-39. Maple Grove, MN: JAM Press.
- Browne, W. J. and Rasbash, J. (2004). *Multilevel Modelling*. Bryman, A. and Hardy, M. (Eds.), *Handbook of Data Analysis*
- Dumais (2009), Higher education institutions, sampling issues preparing the AHELO Feasibility Study, OECD, available from www.oecd.org/edu/ahelo/callfortenders
- Ewell, P., Aamodt, P., Allen, J., Bettinger, E., Brennan, J., Federkeil, G., Kaiser, F., Llavori, R., McCormick, A., Ogata, N., Prades, A., Sola, J. (2009). Analytical framework for the contextual dimension of the AHELO Feasibility Study, OECD, available from www.oecd.org/edu/ahelo/callfortenders
- Grisay, A. (2003), Translation procedures in OECD/PISA 2000 international assessment, *Language Testing*, 20(2), 225-240.
- Hambleton, R.K, Merenda, P.F. and Spielberger, C.D., Eds (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, NJ: Laurence Erlbaum Associates.
- Hambleton, R.K. (2002), *Adapting Achievement Tests into Multiple Languages for International Assessments*, in A. C. Porter and A. Gamoran, Eds, *Methodological Advances in Cross-National Surveys of Educational Achievement*, Board on Testing and Assessment (BOTA). Washington, DC: National Academy Press.
- Lockwood, J.R., McCaffrey, D.F., Mariano, L.T. & Setodji, C. (2007). Bayesian Methods for Scalable Multivariate Value-Added Assessment, *Journal of Educational and Behavioral Statistics*, Vol. 32, No. 2, 125-150
- Muthen, B. (2002). Beyond SEM: General latent variable modelling. *Behaviormetrika*, 29, 81-117.
- OECD (2009a), *Roadmap for the OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study*, 2nd version, Paris: OECD, June 2009.
- OECD (2009b) *Call for Tenders EXD/BPF/CPG/EDU/EMI/09/57 Feasibility Study for an Assessment of Higher Education Learning Outcomes (AHELO)*, Paris: OECD, July 2009.
- Raudenbush, S. & Bryk, A. (2002). *Hierarchical Linear Models*, 2nd Ed, Sage, Thousand Oaks, CA.

Raudenbush, S.W. (2004). What are value-added models estimating and what does this imply for statistical practice *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.

Rubin, D.B, Stuart, E. A., and Zanutto, E.L. (2004). "A Potential Outcomes View of Value-Added Assessment in Education." *Journal of Educational and Behavioral Statistics*, 29:103-116.

Scott, G., Coates, H. & Anderson, M. (2008). *Learning Leaders in Times of Change: academic leadership capabilities for Australian higher education*. Sydney: ALTC.

ANNEXURE

MODULE A WORK PLAN

Forthcoming

MODULE B WORK PLAN

Module Objective

401. This Annex presents a Statement of Work for the Australian Council for Educational Research (ACER) in collaboration with the Educational Testing Service (ETS) to develop the assessment framework and instrument for the Economics Strand (Module B) of the Assessment of Higher Education Learning Outcomes (AHELO) feasibility study.

402. ETS will develop a provisional assessment framework and assessment instrument. ETS will provide support resources on coding and analysis that will be required to measure those learning outcomes in Economics that are ‘above content’. The instrument will be tested qualitatively across those countries involved in the Economics Strand of the AHELO feasibility study.

Module Tasks

1. Framework development:

403. ETS will work with an international committee of Economics experts in developing a framework for the field of Economics. Given the short timeframe available for development, the constraints on consultation, and given the complexity of the field of Economics, the framework that results at the end of the feasibility study will be considered a ‘provisional framework.’ The ‘provisional framework’ will not be complete, but will contain most of those elements which the majority of experts in the field agree are essential. Completions of the work on the framework will take place when the AHELO project gets to the next stage by taking advantage of the responses of the larger community of international Economics experts to the ‘provisional framework.’ This approach to the development of the ‘provisional framework’ will involve:

- Utilization of the ‘Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics’ and the ‘QAA Subject Benchmark Statement for Economics 2007’ as the basis of the framework development.
- Establishing and managing an Economics Expert Group who will meet once face-to-face and at other times via teleconference and email to consult and revise a draft framework which will be created by ETS staff.
- Creating a provisional framework which:
 - broadly reflects the current thinking of experts in tertiary Economics education who are listed as members of the Economics Expert Group, and represents their consensus view;
 - takes into account the characteristics of the target population — in this case, final year ‘first-cycle’ (bachelor degree) Economics students;
 - defines the subject domain in terms of features such as content areas, skills and processes and different schools of thought;
 - is specific enough to be useful in the instrument development process, but not so tightly specified that the opportunity for the assessment of integrated skills, conceptual understandings and ‘above content’ learning outcomes is removed; and
 - takes into account cultural and language differences of participating countries.

2. Instrument development:

404. As the provisional framework is being developed, ETS will develop a ‘mini instrument’ for the Economics Expert Group to consider and revise. The items in the ‘mini instrument’ will assess the main areas of focus identified in the provisional framework. The ‘mini instrument’ will be able to serve as the foundation for the final instrument. It will also be useful to demonstrate to stakeholders what the final instrument will look like. It is anticipated that almost of all of the tasks included in the ‘mini instrument’ will be able to be used in the final instrument; the only differences between the two will be breadth of coverage. It is also anticipated that what is learnt by exposing the mini test to students will lead to general improvements which will be reflected in the final instrument. Instrument development will involve:

- Identifying and revising relevant items from existing ETS instruments used for similar populations to those targeted in the AHELO feasibility study.
- Mapping all potential items against the provisional framework to determine if the items have sufficient cultural generality for inclusion in the AHELO feasibility study.
- Working with the Economics Expert Group so that items that contain field-specific terminology (such as ‘demand function’ and ‘utility value’) can be translated precisely enough that items which depend on them will be fair and valid for students in different countries.
- Working with the Economics Expert Group to develop a ‘mini instrument’, consisting of approximately 20 multiple choice items and one free response task, which will not cover the entire domain but be constructed in such a way that all the tasks included will be considered as appropriate to a measure of tertiary learning outcomes in Economics.

- Linking all the items and tasks in the assessment to the provisional framework.

3. Item validation:

405. ETS will work with the Module B Economics Expert Group to assess the validity of items. This will involve:

- Checking that all items reflect the consensus view of a group of well-regarded international experts so that the items represent appropriate content and that they reflect diverse cultural and national perspectives, experiences and priorities.
- Using drafts of English versions of the items as the basis for a small number of cognitive labs and think aloud interviews with United States university seniors, which will be conducted and recorded.
- Pretesting the ‘mini test’ with small groups of students in a convenience sample of institutions identified by AHELO National Project Managers and/or by the member of the Economics Expert Group in each of the target languages. Staff in participating countries will replicate the earlier cognitive labs and think aloud interviews. Students will be asked to complete the items and then asked to complete a small number of evaluative questions about each item. ETS will work with the Consortium to develop a set of item level questions that can be used for this evaluation.
- Summarising validity and usability information for each item in a final report.

4. Contribute to project management:

406. ETS will assist with broader aspects of the study’s planning, implementation, analysis and reporting. Specifically, ETS will work closely with the teams leading the other AHELO modules to maximise synergies across the different strands of work and ensure that the overall project is carried out on time and within budget.

5. Deliverables:

407. ETS will deliver:

- A provisional framework that is based on the ‘Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics’ and the ‘QAA Subject Benchmark Statement for Economics 2007’ and the consensus of view of the Economics Expert Group.
- A ‘mini assessment’ of Economics consisting of approximately 20 multiple choice items and one complex constructed response task with a scoring guide.
- A mapping of the items in the ‘mini assessment’ to the provisional framework.
- A summary instrument development report including details of the development process, the pretest outcomes and a mapping test items to the assessment framework for the Engineering instrument.

6. Milestone dates:

408. The table below outlines milestones, deliverables (shaded) and key dates for Module B.

Module	Milestone/Deliverable	Date
Module B	Begin framework and instrument development	June 2010
Module B	Contribution to study analysis plan	September 2010
Module B	Expert Group meeting (face-to-face)	October 2010
Module B	NPM meeting	October 2010
Module B	Cognitive labs with source version	November 2010
Module B	Produce source version of instrument	December 2010
Module B	NPM meeting	April 2011
Module B	Deliver framework and specifications	April 2011
Module B	Deliver verified translated instruments	May 2011
Module B	Deliver summary report	May 2011
Module B	Deliver report mapping items and framework	May 2011

Roles and Responsibilities

409. The ETS Research and Development staff members leading Module B will include:

- 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.

410. ETS staff will be available to attend meetings of the AHELO GNE, Module B Economics Expert Group, and of the NPMs from countries participating in the AHELO Economics assessment.

411. The Economics Expert Group includes:

- Cecilia Conrad, United States (Chair)
- William Becker, United States
- Fiorella Kostoris, Italy
- Maria de Lourdes Dieck-Assad, Mexico
- Henriëtte Maassen van den Brink, The Netherlands
- Tatsuya Sakamoto, Japan
- Vladimir Zuev, Russia

MODULE C WORK PLAN

Module Objective

412. This Annex presents a Statement of Work for the Australian Council for Educational Research (ACER) in collaboration with the National Institute for Education Policy Research in Japan (NIER) and the University of Florence School of Engineering (UFSE) in Italy, together labeled as the 'Module C team' to develop the assessment framework and instrument for the Engineering Strand (Module C) of the Assessment of Higher Education Learning Outcomes (AHELO) feasibility study.

413. ACER will lead the development of an assessment instrument to measure selected learning outcomes in Engineering that are 'above content'. The test will be evaluated qualitatively across those countries involved in the Engineering Strand of the AHELO feasibility study. The development of this instrument will involve establishing a provisional assessment framework, identifying existing and creating new items, and testing the instrument's validity and reliability.

Module Tasks

1. Framework development:

414. The Module C team will develop a provisional framework for the field of Engineering. The Engineering strand framework will be developed to include one of the three branches of Engineering articulated in the AHELO Feasibility Study Tuning Framework, namely civil engineering, as advised by the Engineering Expert Group and OECD. Framework development will involve:

- Utilization of the 'Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Engineering' as the basis of the framework development.
- Establishing and managing an Engineering Expert Group who will meet once face-to-face and additionally via teleconference and email to oversee the development of the Engineering Strand framework and establish international understanding, commitment and agreement with the approach.
- Create an assessment framework for Engineering that:
 - is informed by current thinking of experts in Engineering education who are listed as members of the Engineering Expert Group, and represents their consensus view;
 - takes into account the characteristics of the target population – in this case, final year 'first-cycle' (bachelor degree) Engineering students;
 - defines the subject domain in terms of features such as content areas, skills and processes, and background information about the discipline;
 - is specific enough to be useful in the instrument development process, but not so tightly specified that the opportunity for the assessment of integrated skills, conceptual understandings and 'above content learning outcomes' is removed; and
 - takes into account cultural and language differences of participating countries.

2. Instrument development:

415. Following the development of the provisional assessment framework for the Engineering Strand, ACER and Module C partners will develop an instrument to test one branch of Engineering (Civil Engineering). Instrument development will involve:

- Identifying relevant items from existing instruments used for similar populations to those targeted in the AHELO feasibility study.
- Developing new items for the instrument. These will cover different item types and may include multiple choice and short-constructed responses.
- Developing scoring/coding guides in consultation with the Module C Expert Group and National Project Managers (NPMs) to ensure that constructed response tasks measure higher order skills and can be scored reliably across cultural and linguistic groups.
- Mapping all potential items against the established framework, in consultation with the Expert Group.
- The instrument will consist of 15 to 20 minute testlets (groups of items based on a common engineering context) with a mixture of multiple choice and short text responses. We plan to develop approximately 180 minutes (between 9 and 12 testlets) initially to take to the item validation process with a view to reducing this to a recommended 6 to 9 testlets as the final recommended 90 minute instrument.

3. Item validation:

416. The Module C team will liaise with the Engineering Expert Group to establish the validity of items and decide on a suitable configuration of items that best meets the over-arching goals of AHELO. Maximising the validity of items will be based on:

- Mapping items to the conceptual framework;
- Ensuring that items reflect diverse modes of thought, and diverse cultural and national perspectives, experiences and priorities;
- Using drafts of English versions of the items as the basis for a small number of cognitive labs and think aloud interviews with Australian university seniors, which will be conducted and recorded.
- Pretesting the 'mini test' in-country with small groups of students in a convenience sample of institutions identified by AHELO National Project Managers and/or by the member of the Engineering Expert Group in each of the target languages. Staff in participating countries will replicate the earlier cognitive labs and think aloud interviews. Students will be asked to complete the items and then asked to complete a small number of evaluative questions about each item. ACER will work with the Consortium to develop a set of item level questions that can be used for this evaluation.
- Summarising validity and usability information for each item in a final report.

4. Contribute to project management:

417. The Module C team will assist with broader aspects of the study's planning, implementation, analysis and reporting. Specifically, the Module C team will:

- contribute to the development of a feasibility study analysis plan; and
- work closely with the teams leading other AHELO modules to maximise synergies across the different strands of work and ensure that the overall project is carried out on time and within budget.

5. Deliverables:

418. The Module C team acknowledge and accepts the deliverables stated in the AHELO Terms of Reference. The Module C team will deliver:

- The framework and test specifications for the Engineering assessment.
- The instrument for the Engineering assessment including the scoring/coding guides.
- A mapping of the items in the 'mini assessment' to the provisional framework.
- A summary instrument development report including details of the development process, the pretest outcomes and a mapping test items to the assessment framework for the Engineering instrument.

6. Milestone dates:

419. The table below outlines milestones, deliverables (shaded) and key dates for Module C.

Module	Milestone/Deliverable	Date
Module C	Begin framework and instrument development	June 2010
Module C	Contribution to study analysis plan	September 2010
Module C	Expert Group meeting (face-to-face)	October 2010
Module C	NPM meeting	October 2010
Module C	Cognitive labs with source version	November 2010
Module C	Produce source version of instrument	December 2010
Module C	NPM meeting	April 2011
Module C	Deliver framework and specifications	April 2011
Module C	Deliver verified translated instruments	May 2011
Module C	Deliver summary report	May 2011
Module C	Deliver report mapping items and framework	May 2011

Roles and Responsibilities

420. The core staff members (and organizational affiliations) involved in Module C team include:

- Australian Council for Educational Research (ACER), Australia:
 - Julian Fraillon – Director, Module C;

- Jacob Pearce and Ray Philpot – Engineering assessment leads;
- Daniel Urbach – psychometric support; and
- Roger Hadgraft (Associate Professor of Engineering, University of Melbourne) – Engineering content and education expert;
- National Institute for Educational Policy Research (NIER), Japan:
 - Shuichi Tsukahara – framework and item development;
 - Satoko Fukahori – framework and item development; and
 - Fumiko Yasuno – framework and item development;
- University of Florence School of Engineering, Italy:
 - Claudio Borri – framework and item development; and
 - Elisa Guberti – coordination.

421. The Engineering Expert Group includes:

- Professor Robin King, Australia (Chair)
- Professor Giuliano Augusti, Italy
- Professor Michael Hoffman, Germany
- Professor Kikuo Kishimoto, Japan
- Professor Johan Malmqvist, Sweden
- Professor Jim Melsa, United States
- Professor Lueny Morell, United States
- Professor Masuda Nobutoshi, Japan
- Professor Isao Satoh, Japan

MODULE D WORK PLAN

Forthcoming.

MODULE E WORK PLAN

Module Objective

422. This Annex presents a Statement of Work for Module E of the AHELO feasibility study. Module E covers the overall management of the project and its fieldwork. It involves a number of organisations. The Australian Council for Educational Research (ACER) will lead this Module and has overall responsibility for the project. Capstan will also be involved in Module E work and, with ACER, will form the 'Module E team'.

423. Module E 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 B and C to ensure clarity, coordination and consistency across the different strands of the project.

Module Tasks

1. Project Management:

424. ACER will be responsible for the oversight and coordination of Module E and therefore the whole AHELO feasibility study. This role will involve:

- Working and communicating closely on a daily basis or as needed with other Consortium members and the OECD Secretariat to ensure:
 - effective and efficient completion of implementation tasks;
 - the consistent use of the highest quality standards in the work completed;
 - openness and transparency in all dealings with each other, with AHELO feasibility study participants and with the OECD Secretariat; and
 - the application of strict financial controls.
- More specifically, in its overall management role, the ACER team will:
 - Develop a centralized communication tool to enable efficient management and sharing of documentation for all AHELO feasibility study participants.

- Coordinate the organization of the AHELO Technical Advisory Group (TAG), including one at least one face-to-face meeting.
- Assist countries establish a National Centre in each participating country. These Centres will play a vital role in supporting the development and implementation of the study.
- Assist participating countries to identify and select National Project Managers (NPMs) who will be located in the National Centres and will be the primary contact persons for dealing with participating countries. NPMs will lead work in their country and provide a channel through which national interests are represented in the implementation of the AHELO feasibility study.
- Coordinate and support countries to manage in-country item tests and student evaluations.

2. Assessment design and analysis plan:

425. All organizations in the Module E team will be involved in developing and delivering an integrated assessment design and an analysis plan. This work will take account of the AHELO feasibility study's aim: to assess whether it is possible to develop assessment frameworks and instruments for Economics and Engineering that are valid cross-linguistically and cross-culturally.

- The assessment design will specify the AHELO feasibility study's:
 - project management approach;
 - translation, adaptation and verification approach;
 - coding procedures;
 - analysis plan;
 - means of preparing data products and written reports; and
 - overall quality assurance and risk management processes.
- The analysis plan will be more specific and operational in nature. It will:
 - specify the research questions posed by the various strands;
 - outline the best methods and analyses to assess the cross-national and cross-cultural validity of the instruments being used in the various assessments and context surveys;
 - chart the quantifiable criteria to assess success in the various dimensions of the feasibility study; and
 - be structured to deliver data and insights that assist countries decide on next steps.

3. Translation, adaptation and verification:

426. Capstan will be in charge of linguistic quality control for the AHELO Feasibility Study. This will involve translation, adaptation and verification processes designed to maintain cross-national comparability of assessment materials (Modules B and C). The key facets of this task are:

- Quality Assurance involving:
 - a preliminary scrutiny of test materials to anticipate potential adaptation issues, ambiguities, cultural issues or item translatability problems;
 - the development of Translation and Adaptation Guidelines to assist national translation teams;
 - an appropriate localization environment proposed to participating countries;
 - the implementation of a double translation and reconciliation design to the extent possible; and
 - the assessment of the use of single translation of part of the materials if participating countries have resource problems in complying with double translation.
- Quality Control, involving:
 - verification of target versions against source versions;
 - effective reporting of residual errors and undocumented deviations;
 - expert advice in the case that corrective action is needed;
 - a final check procedure; and
 - standardised quantitative and qualitative reports on the extent to which the different target versions followed the translation and adaptation guidelines.
- Adaptations, many of which will be undertaken by national staff within participating countries with extended expertise in test development and in the domain assessed. However, the Module E team will facilitate adaptations and ensure accurate documentation of all adaptations implemented in national versions.

4. Data products and written reports:

427. 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 produce a draft AHELO Feasibility Study Report.

5. Deliverables:

428. ACER and partner organisations will deliver:

- AHELO Assessment Design, including the identification of the best methods/analyses to assess the cross-linguistic and cross-cultural validity of the results of the various instruments.
- AHELO Analysis plan, including specification of the research questions posed by the various dimensions of the AHELO feasibility study, proposed analyses to address them and the establishment of quantifiable criteria to assess success in these various dimensions.
- Translation/adaptation guidelines for NPMs and the report on the translation/adaptations issues.
- Draft AHELO Feasibility Study Report.
- The following management deliverables will be submitted:
 - a quarterly report on progress against the agreed project timetable; and
 - 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.

6. Milestone dates:

429. The table below outlines milestones, deliverables (shaded) and key dates for Module E.

Module	Milestone/Deliverable	Date
Module E	Establish study management	June 2010
Module E	Delivery quarterly progress report	September 2010
Module E	Deliver AHELO Assessment Design	September 2010
Module E	Deliver AHELO Analysis Plan	September 2010
Module E	Expert Group meeting	October 2010
Module E	Deliver translation and adaptation guides	December 2010
Module E	Delivery quarterly progress report	December 2010
Module E	Deliver Annual budgetary update	January 2011
Module E	Delivery quarterly progress report	March 2011
Module E	Delivery adaptation and translation verifications	May 2011
Module E	Expert Group meeting	April 2011
Module E	Delivery quarterly progress report	June 2011
Module E	AHELO Feasibility Study Report	June 2011

Roles and Responsibilities

430. 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 – research, coordination, document drafting
 - Siek Toon Khoo – technical advice, quality assurance; and
 - Tim Friedman – technical support.

- Capstan:
 - Andrea Ferrari – definition of procedures;
 - Steve Dept – operations director;
 - Raphaël Choppinet – assistant project manager;
 - Laura Wäyrynen – project manager for linguistic quality control; and
 - Shinoh Lee – administration and management;

431. The AHELO TAG includes:

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