

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

Group of National Experts on the AHELO Feasibility Study

AHELO FEASIBILITY STUDY TECHNICAL STANDARDS

9th meeting of the AHELO GNE

Paris, 19-20 March 2012

This document was prepared by the ACER Consortium.

The AHELO GNE is invited to COMMENT and AGREE on the technical standards.

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

JT03317583

Complete document available on OLIS in its original format

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



TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS 3

INTRODUCTION 4

 Overview 4

 Underpinning principles 4

 Application of the standards 5

STUDY MANAGEMENT 6

 International methodology and management 6

 National management 7

 Communication between the NC and AHELO Consortium 8

 International training and meetings 8

 Implementation of national options 8

 Confidentiality and security 9

 Quality assurance 9

 Timeliness 9

 National and stakeholder involvement 9

INSTRUMENT PRODUCTION 10

 Assessment frameworks 10

 Instrument development and validation 10

SELECTING PARTICIPANTS 12

 Target population specification 12

 Sampling and recruitment procedures 12

TRANSLATION AND CULTURAL ADAPTATION 13

FIELD OPERATIONS 14

 General conditions 14

 Scoring constructed response tasks 14

DATA AND REPORTING 15

 Data management and analysis 15

 Reporting and interpretation 16

RESOURCES 16

ACRONYMS AND ABBREVIATIONS

AHELO	Assessment of Higher Education Learning Outcomes
ATAV	AHELO Translation, Adaptation and Verification
EG	Expert Group
GNE	Group of National Experts
HEI	Higher Education Institution
IC	Institutional Coordinator
LS	Lead Scorer
NC	National Centre
NPM	National Project Manager
OECD	Organisation for Economic Cooperation and Development
TA	Test Administrator
TAG	Technical Advisory Group

INTRODUCTION

Overview

1. The Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study (henceforth 'AHELO') seeks to determine a robust approach to measuring learning outcomes in ways that are valid across cultures and languages, and across the diversity of institutional settings and missions.
2. This document provides the OECD AHELO Technical Standards. The standards are designed to ensure processes that yield trustworthy data which can be used to make meaningful comparisons between higher education institutions (HEIs). It is essential that processes and outcomes are robust.
3. These AHELO Technical Standards form part of a suite of resources used for the AHELO Feasibility Study. Detailed operational and technical information about AHELO can be obtained from National Project Managers (NPMs) in participating countries. Public information is available from www.oecd.org/edu/ahelo.
4. AHELO is a large and complex project. These standards have wide scope and are relevant to all facets of AHELO. The AHELO Technical Standards establish agreed on methods in which there are mutual accountability among and between participating higher education systems, HEIs, and the AHELO Consortium.

Underpinning principles

5. These standards are not designed as universal standards for testing, even in the context of higher education. Rather, they have been developed through research, consultation and experience to provide foundations for OECD AHELO. They are intended to be relevant, succinct, measurable and enforceable. Hence this document is deliberately brief, practically focused and contextualised. Where possible and relevant, the AHELO Technical Standards rest on and reference broader standards produced for educational measurement, assessment and evaluation. Key reference documents are listed in the resources section of this document. In particular, readers are referred to various International Test Commission standards (see: ITC, 2000; 2005; 2010; 2011), the IEA standards (Martin, Mullis, Caldwell, Foy, Sibberns, Adams & Rust, 1998) and the OECD PIAAC standards (OECD, 2010).
6. The standards have been developed with major and inter-related principles in mind:
 - a. Relevance: Technical work must be well positioned within salient educational, policy and practical contexts. Technical work must be designed and conducted to enhance the relevance of AHELO. AHELO is an international policy initiative intended to provide HEIs and other stakeholders with information about the educational achievement of undergraduate students which will inform higher education policy. It is imperative that technical work be appropriate and set to the highest standards possible.

- b. Consistency: Data should be collected in an equivalent fashion in all HEIs and systems, using equivalent assessment materials. A comparable sample of the student population should be assessed under test conditions that are as similar as possible. The use of consistent data collection methods will ensure that assessment results can be compared between HEIs, and that the performance of students in the sample will provide a statistically valid representation of the performance of students in the target population with similar characteristics. Equally, consistent data collection will ensure that differences in student performance are not caused by factors which are un-related to individual abilities.
- c. Precision: Data collection and submission practices should leave as little room as possible for variation or error, whether systematic or random. This includes errors which could be caused by variations in testing environments for different groups of students, and errors which could occur during data preparation. The greater the level of precision in data collection and preparation, the higher quality of results and the more reliable data analysis and reporting. It is also imperative that AHELO results are reported and interpreted in appropriate ways.
- d. Generalisability: Data are collected from specific individuals, in a particular situation, and at a certain point in time. The selection of individuals, test materials, tasks and all other testing conditions should be highly standardised. This will ensure that conclusions reached from the analysis of data from the sample will not only reflect the circumstances in which the data was collected, but will also be accurate more generally, and hence relevant for the target population as a whole.
- e. Timeliness: Relevance, consistency, precision and generalisability of the data can be obtained in a number of ways. In AHELO it is imperative that all activities are conducted within given timelines in order to meet HEI, national and international schedules, and reporting deadlines.

Application of the standards

7. The AHELO Technical Standards act as a benchmark of good practice. As such, the standards are designed to assist National Centre (NCs) and the AHELO Consortium by explicitly defining expectations for study design, development and implementation. NPMs should also refer to all documents contained in the AHELO NPM Manual for guidance on the implementation of different aspects of the AHELO. Particularly relevant documents include the AHELO:

- a. Analysis and Reporting Guidelines;
- b. Analysis Plan;
- c. Assessment Design;
- d. Assessment Frameworks (for various strands);
- e. Data Access, Use and Reporting Policy;
- f. Data File Specifications;
- g. Focus Group Guidelines;
- h. Institution Coordination Manual;
- i. International Scoring Manual and strand-specific Scoring Guides;
- j. International Training Meeting Agendas and Presentations;
- k. NPM Manual;

- l. Sampling Manual;
- m. Technical Standards;
- n. Test Administration Manual; and
- o. Translation and Adaptation Guidelines.

8. Where this document indicates it may be possible for countries to vary from a technical standard this is premised on agreement between countries and the AHELO Consortium. Wherever this occurs, NPMs are asked to initiate the process of negotiation with the International Project Director on behalf of the AHELO Consortium, and to undertake everything possible to facilitate an agreement. The OECD will adjudicate any issues resulting from non-compliance with the technical standards that cannot be resolved between participating countries and the AHELO Consortium.

9. The AHELO Technical Standards are designed primarily for people and organisations involved directly in AHELO. The standards pertain to administration, development, implementation, analysis, reporting and review activities related to AHELO. The standards do not cover all potential applications and use of AHELO activities, data or results.

10. System and HEI data will be recommended for inclusion in the international data files when countries are able to demonstrate to the AHELO Consortium's satisfaction that standards have been fully met. When countries are not able to demonstrate that standards have been fully met an adjudication process will determine the extent to which the quality and international comparability of the data have been affected and whether the data should be included in files and any publications.

11. Given the importance of complying with these AHELO Technical Standards, NPMs should report any potential issues to the AHELO Consortium as soon as these arise, to enable corrective action to be taken.

STUDY MANAGEMENT

International methodology and management

12. AHELO should be conducted in ways consistent with methods and practices developed in the latter part of the last century for international educational assessment. While mostly formed within contexts of compulsory education, many technical standards can be transferred to large-scale studies in higher education. Important facets for alignment include the coordination of management and support personnel, provision of manuals and support resources, conducting of testing sessions, scoring of constructed responses, verification and processing of data, and analysis and reporting.

13. The Assessment Design and the Analysis Plan (the 'Analysis Plan' is AHELO's evaluation plan) are documents that detail the methodological, practical and evaluative contexts for AHELO. As a large international assessment AHELO should experiment with and evaluate innovative research approaches where possible.

14. The AHELO Assessment Design and Evaluation Plan is operationalised in terms of international manuals and guidelines that specify numerous technical and operational requirements. These AHELO

Technical Standards are designed to be consistent with these manuals and guidelines, but the Technical Standards will prevail in the case of any ambiguity or disagreement.

15. Formal governance and project leadership arrangements are established and supported by OECD. International technical and operational work is managed by collaborative international teams (the ‘AHELO Consortium’) which are expert in large-scale educational assessment and higher education. Advisory and stakeholder groups have been established—notably a Technical Advisory Group (TAG) and an Expert Group (EG) for each strand of testing, if not the TAG. Structures are provided to manage the involvement of systems and HEIs.

National management

16. Participating countries must establish a National Centre, appoint a NPM, recruit Lead Scorers (LS) and a national scoring team, and engage HEIs, Institutional Coordinators (ICs) and Test Administrators (TAs):

- a. A NC must be established in each participating system. The NC is the organisational group that manages AHELO in the system.
- b. The NPM is responsible for the implementation of AHELO at the national level. The NPM has overall responsibility for ensuring that all required tasks are carried out on schedule and in accordance with the AHELO Technical Standards and operations guidelines, and for documenting processes implemented at the national level. The NPM must have the capacity to represent the system internationally, and also to manage national stakeholders and work.
- c. In each system a Lead Scorer (or Lead Scorers, if a system is participating in more than one strand) is responsible for the scoring of all student responses. Lead Scorers are supported by a national scoring team.
- d. Each participating HEI must nominate an IC. The IC represents the HEI in AHELO and works closely with the NPM.
- e. TAs work closely with ICs to administer tests to students within HEIs. TAs must have experience in the management and supervision of testing and examination in university settings. They must be familiar with the use of computer-based assessment delivery systems. TAs should not have a direct personal or professional relationship with any of the students in the testing sessions.

17. NCs and NPMs must be sufficiently resourced to enable them to participate in the study. In particular the AHELO Consortium will provide guidance to ministries and national agencies on resources required to establish a NC, and to NPMs on resources required to managing international and national facets of the work. National resource constraints must be made clear and discussed with the AHELO Consortium when participation commences.

18. Organisations and personnel participating in AHELO must have appropriate levels of expertise, training, support and monitoring to fulfil project duties. Appropriate professional capability and experience is particularly important for people in leadership and management roles such as NPM and LS. As ICs and TAs play a central role in fieldwork, NPMs must give special consideration to the training of these personnel, ensuring that as little variation in the data as possible is caused by random or systematic variation in test administration activities.

Communication between the NC and AHELO Consortium

19. Delays in communication between NCs and the AHELO Consortium should be minimised given the tight schedule of AHELO. NCs need continuous access to the resources provided by the AHELO Consortium.

20. The AHELO Consortium ensures that qualified staff are available to respond to requests by NCs during all stages of the project. The qualified staff:

- a. are authorised to respond to NC queries;
- b. acknowledge receipt of NC queries within two working days;
- c. respond to scoring queries from NCs within two working days; and
- d. respond to other queries from NCs within five working days or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.

21. The NPMs ensure that messages from the AHELO Consortium are responded to promptly during all stages of the project by:

- a. responding to requests for information by any deadline specified in written communication from the AHELO Consortium or, when no deadline is specified, within ten working days; and
- b. acknowledging receipt of AHELO Consortium queries within two working days.

International training and meetings

22. Communication and discussion of assessment (or educational) contexts, methods and practices plays an essential role in AHELO. Accordingly, national staff, including NPMs and LS, are expected to attend international meetings and take part in international conference calls.

23. AHELO involves the design, development, adaptation and application of a wide range of complex assessment methods and practices. Accordingly, it is essential given that people with coordination or management roles (e.g. NPMs, LS, ICs, TAs) participate in prescribed online and face-to-face training.

Implementation of national options

24. Variations made at the national level must not affect the data used for any international comparison. Data must be collected consistently across HEI and systems, and potential unintended effects like test fatigue, or learning effects from national test items, are precluded.

25. Only national options that are agreed on between the NC and the AHELO Consortium are to be implemented. Such requests tend to relate to population definition, student or faculty sampling, instrumentation (particularly the omission of items on grounds of relevance, or inclusion of additional context items), language for testing, and structuring and management of national operations.

26. Any national items or instruments are administered after all of the international test and context instruments have been administered.

Confidentiality and security

27. AHELO materials designated as secure must be kept confidential at all times. Secure materials include all test materials, data, management materials and technical documents. All secure materials must be clearly and visibly identified as confidential.

28. Formal confidentiality agreements must be in place for all people with operational, technical or advisory roles in AHELO. No one except approved people and participating students during the test session is able to access and view test material. No one other than approved people will have access to secure AHELO data and embargoed material.

29. Any breach in security must be reported immediately to the AHELO Consortium.

Quality assurance

30. The AHELO Consortium will ensure that all operational documentation emphasises quality control as an integral part of every activity. Quality assurance contexts and processes must be embedded within international and national methods and practices as specified in AHELO guidelines and manuals, including these AHELO Technical Standards.

31. The AHELO GNE and the OECD have ultimate responsibility for the quality of AHELO inputs, processes and outcomes. The AHELO Consortium and nominated EGs (notably, the TAG) provide international oversight of quality assurance. NPMs are responsible for the quality of fieldwork in their system. AHELO test administration is monitored by trained TAs and ICs.

32. To assist with quality assurance, international and national managers must document AHELO administration, designs, development, validation, implementation, data and file management, analyses, reports and interpretations. This documentation must be provided to the AHELO Consortium to assist with quality improvement (training and engagement) and assurance (monitoring and review) activities.

Timeliness

33. The timely progression of the project within tight schedules depends on diligent and responsive management, effective fieldwork and scoring procedures, and clear communication and collaboration. The AHELO Consortium needs to render effective support to NCs, and to receive materials from NCs and NPMs by deadlines specified by the AHELO Consortium.

National and stakeholder involvement

34. It is critical that NPMs take active steps to localise AHELO within their national/regional contexts. Where relevant, national stakeholder engagement strategies must align with international recommendations and practices. Policies and practices associated with learning outcomes assessments are relatively new in many systems, requiring NPMs to be particularly active in their promotion and engagement activities.

35. Feedback from NCs, HEIs and other national stakeholders is important in maintaining the dynamic and collaborative nature of AHELO. Such feedback will cover broad areas such as the positioning and relevance of AHELO, framework and instrument development, validation and operationalisation, implementation, and analysis and reporting. Feedback ensures that materials, processes and outcomes achieve international, cross-cultural, cross-linguistic and cross-institutional validity. It also promotes the inclusion of the interests and involvement of national stakeholders.

INSTRUMENT PRODUCTION

Assessment frameworks

36. Assessment frameworks are produced for all AHELO assessments—test and context instruments alike. These frameworks must contextualise the assessment among related activities, define the domain being measured, offer a clear conceptual structuring of the domain, and provide details of how the domain is to be operationalised for the purposes of measurement. The framework must reflect a synthesis of international research, be reviewed and agreed on by relevant national stakeholders and experts, and reflect an international consensus about the areas that are important to assess—through being formally approved by the relevant EG. Following item development, a mapping should be produced which links the assessment framework and instrument.

Instrument development and validation

37. Test and context items and instruments must include materials or feedback from national representatives and domain experts. Item content and format must be accessible to all populations, and typically be sourced from a number of cultural contexts. The linguistic complexity of items should be pitched at an appropriate level for the target population. To the extent possible, item difficulty should be a function of subject matter rather than linguistic complexity.

38. Before being used in fieldwork, draft items must be reviewed and approved by domain and assessment experts, typically but not limited to the EG for each strand of work and/or the TAG.

39. Scoring guides and rubrics must be developed for constructed response tasks. These materials must describe and include examples of student responses for each score level. The guides and rubrics must be approved by the relevant EG and (as possible) LS from participating countries.

40. Before being used in fieldwork draft items and rubrics must be subjected to qualitative analysis and review. At a minimum, the qualitative review should involve review by students and faculty from the target population in each participating system. Feedback questionnaires should be deployed to collect student and faculty views on the items.

41. Before being used in fieldwork draft items must be subjected to pre-testing with individuals from the target population, and quantitative analysis and review. To support psychometric analysis, the quantitative validation should involve the collection and analysis of at least 100 responses for each item from as many participating countries as possible. Rubrics for constructed response tasks should be tried out in each country. Psychometric analyses must be conducted to produce final source versions that exceed requirements set out in these Technical Standards.

42. The final test design must take into account constraints surrounding administration (notably time and resources), computer availability, content coverage, item sampling and rotation (to minimise order and fatigue effects), respondent sampling, linking and equating needs, and analysis and reporting contexts and methods.

43. When generalised across HEI subsample replications, the item clusterings in the test instruments must exceed prescribed thresholds for internal consistency. In particular, any item clusterings that will be used for reporting must have a reliability estimate of at least 0.80 at the individual level.

44. The test and context instruments must have well defined content validity. This validity must be established by using the assessment frameworks to situate the instruments in appropriate applied and research contexts. The content validity of the items must then be confirmed by mapping the items on the clear conceptual structuring of the domain given in the framework.

45. The test and context instruments must have well defined 'face' (henceforth 'above content') validity. Above-content validity refers to whether or not the instrument looks like it is pitched at a sufficient level of generalisability to enable international use. Above content validity must then be confirmed by feedback from EGs, TAG, HEIs, stakeholders.

46. The test and context instruments must have acceptable levels of construct validity. Psychometric and statistical procedures should be used to analyse the dimensionality of the measured construct, to confirm divergent and convergent validity of the measured constructs, and to define any differential item functioning. Replication analyses should be undertaken across contexts to affirm that alternate versions of the test instrument have equivalent psychometric properties. Construct validity can be referenced by these criteria:

- a. Standardised item loadings from exploratory/confirmatory factor analyses above 0.4;
- b. Items should not be too easy or too difficult, should show discrimination between 0.3 (and preferable 0.4) and 0.8, should have an average discrimination around 0.5, and should spread out across the range of the target construct;
- c. Goodness of fit indices from confirmatory factor analyses should be within acceptable ranges (e.g. GFI or TLI above 0.95; RMSEA below 0.08);
- d. Point biserial correlations for keyed response should be above 0.3, and point biserial correlations for distractors should be negative;
- e. Mean square residual (fit) statistics for each item should be close to 1.00 (specifically, above 0.8 and below 1.1);
- f. The centre and distribution of item demand estimates should align with the centre and distribution of student ability estimates;
- g. Item demand estimates for different subsample calibrations should be within 95% confidence limits; and
- h. Correlations between 'like/convergent' constructs should be larger than correlations between 'dissimilar/divergent' constructs, as is possible to determine given available data.

47. A report must be produced describing the test development processes and outcomes for each strand. This report must detail framework development, item development. In particular it should include information on item sourcing and drafting, panelling, rubric and scoring guide production, qualitative and quantitative validation, review processes, translation, adaptation, graphical design, operationalisation, and proofing.

SELECTING PARTICIPANTS

Target population specification

48. Meeting the specified sampling standards ensures that students and faculty tested and surveyed come from the comparable target population in every system, and are in a nearly equivalent stage in their first-cycle or bachelor-degree program.
49. Specifications for the definition and sampling of HEIs will be provided by the OECD or AHELO Consortium.
50. The student and faculty target populations are agreed on through negotiation between the NPM and the AHELO Consortium, within constraints imposed by the AHELO Sampling Manual. As with other national options, variations must be discussed and agreed on with the AHELO Consortium.
51. Exclusions are defined in the AHELO Sampling Manual. The proportion of excluded individuals, and the reason for their exclusion, must be documented. Only eligible students participate in the test.
52. NPMs must document the sampling design and frame that has been used and provide this documentation to the AHELO Consortium.

Sampling and recruitment procedures

53. Students and faculty are sampled using agreed on, established and professionally recognised principles of random sampling as documented in the AHELO Sampling Manual. Using probabilistic sampling guarantees that HEI estimates are comparable across institutions even though the strategy employed within each may differ.
54. Unless a census is to be used, it is important that the sample be large enough to achieve a certain precision of measurement. Thus, the minimum number of participating students is specified. The student sample size is a minimum of 150 assessed students for AHELO participants or 75 per cent of the entire AHELO target population where the AHELO target population is below 200.
55. The student and faculty samples must be drawn by the NC/NPM using procedures specified in the AHELO Sampling Manual, then verified by the AHELO Consortium. NPMs must supply the AHELO Consortium with unit-record data files and appropriate syntax files that enable the sample to be validated and replicated. The national sampling frame must be approved by the AHELO Consortium before fieldwork commences.
56. Unless otherwise agreed on, the testing period is no longer than the internationally specified period in duration, and does not coincide with summative final-year assessments.
57. NCs must also develop appropriate mechanisms to promote participation, effective implementation, and dissemination of results amongst all relevant national stakeholders. NPMs and HEIs (ICs and TAs) must align with any defined international methods developed to recruit HEIs, faculty and students into the study. NPMs must consult with the AHELO Consortium regarding the appropriateness of any incentives to be provided to induce student participation in AHELO, and must record incentives offered by participating HEIs in national context instruments.

58. The threshold student and faculty response rate is set at 75 per cent of the sampled student and faculty in each participating HEI. An adjudication process will be launched where it is unclear that outcomes have not met standards.

TRANSLATION AND CULTURAL ADAPTATION

59. Translation, adaptation and verification procedures will be used to ensure the generalisability of assessment materials and instruments across contexts of development, validation, use and reporting.

60. While AHELO assessment and support materials may be sourced from a range of countries, all are translated as required and produced in international English as the primary and sole source language.

61. AHELO test instruments are administered to students in the language of instruction provided by the sampled institution to that sampled student in the majority of her/his study. If the language of instruction in the major domain is not well defined across the set of sampled students then, if agreed on, a choice of language can be provided with the decision being made at the student, HEI or NC level. Agreement with the AHELO Consortium will be subject to the principle that the language options provided should be languages that are common in the community and are common languages of instruction in HEIs. In all cases the choice of test language(s) used in the test is made prior to the administration of the test.

62. A set of linguistic translation, cultural adaptation and independent verification guidelines are to be prepared that specify acceptable variations that may be made by NPMs (the Translation and Adaptation Guidelines). NPMs will be trained by the AHELO Consortium in the use of these guidelines, and in ATAV processes.

63. To ensure psychometric equivalence to the source version all test and context instruments are translated and adapted into the language(s) of instruction by national translation teams convened by the NC, or by professional translators approved by NPMs which have capability in the translation of educational assessments. Two independent translations/adaptations are produced, and these are reconciled. Detailed transaction-level records of translation/adaptation processes are recorded in a log book for management, analysis and audit purposes.

64. To ensure psychometric equivalence to the source version all test and context instruments are subjected to independent verification arranged by the AHELO Consortium. This verification is conducted by teams consisting of a linguist verifier (with experience in verifying international assessment materials) and a domain specialist (with teaching experience in the relevant country) fluent in both English and the testing language. Detailed transaction-level records of the verification processes are recorded in a log book for management, analysis and audit purposes.

65. Test and context instruments must be graphically designed by the AHELO Consortium. NPMs review the designed test and context instruments to validate layout. The design/approval process is repeated until the NPM provides formal sign-off that the instruments are ready for fieldwork.

66. Support materials such as explanatory scripts, rubrics and manuals that are required for administration are translated and adapted into testing languages. Verification may or may not be required. The NPMs provides formal sign-off that the support materials are ready for fieldwork.

FIELD OPERATIONS

General conditions

67. AHELO fieldwork must be conducted in ways consistent with processes and conditions stipulated in the AHELO NPM Manual. To ensure that the data is collected consistently and in a comparable fashion for all students and HEIs it is very important to keep the chain of action in the data collection process as consistent as possible. Procedures given in the Test Administration Manual are particularly important.

68. OECD AHELO is designed according to principles of human research ethics. NPMs and ICs are responsible for ensuring that systems and participating HEIs satisfy relevant requirements.

69. An important part of the testing situation relates to the relationship between TAs and test participants. The relationship between TAs and participating students must not compromise the credibility of the test session. In particular, the TA should not have a direct personal or professional relationship with any student in the assessment sessions he or she will administer for AHELO. Any personal interaction between a TA and students either in the past or in the testing situation counteracts the goal of collecting data in a consistent fashion across HEIs.

70. The nature and purpose of AHELO must be made clear to participants prior to testing sessions. In particular, participants must be provided with a plain language statement about AHELO. The AHELO Consortium will provide a draft statement to NPMs for them to translate and adapt. Where necessary, the AHELO Consortium and NPMs must ensure that students are familiar with the rationales of the assessment, with item formats, and with test conventions.

71. AHELO is a computer-delivered (typically online) assessment and it is essential that international and HEI technology and physical resources meet appropriate technical standards. It is essential that tests are delivered in standard environments such that computer delivery systems do not compromise student participation. Computer-based platforms must have customised interfaces, required management functionality, and the capacity to be deployed and supported internationally.

72. Any AHELO test session may be audited by an independent quality monitor appointed by the AHELO Consortium.

Scoring constructed response tasks

73. To ensure comparability of data from all test participants the scoring scheme is implemented according to instructions in the International Scoring Manual and associated strand-specific Scoring Guides. All scoring procedures must be as standardised across systems and strands as feasible.

74. LS and the national scoring team are recruited, trained and managed following agreed international procedures, and as specified in the International Scoring Manual. Representatives from each NC—typically NPMs and LS—must attend international scorer training sessions.

75. Both the single and double/multiple scoring procedures as specified in manuals, or an agreed on variation thereof, are implemented. At least 20 per cent of responses to constructed response items should be double scored.

76. Indices of inter-rater agreement/reliability should reach 85 per cent as defined by internationally defined reliability metrics.

77. Ongoing quality control must be embedded into each system's scoring operations through NPMs and LS:

- a. Working closely with the international resources and supports—the AHELO Consortium, NPM Manuals, research studies, etc.—to ensure compliance with international standards and practices;
- b. Establishing national scoring operations—facilities, personnel, training, supervision, communication channels, quality controls, etc.—to exceed minimum standards specified in the International Scoring Manual; and
- c. Training, supporting, harmonising and supervising the national scoring team and its individual members.

78. Scoring processes and results from constructed response tasks are affirmed by validity studies conducted by the AHELO Consortium. These studies should investigate comparability between the distribution of scores across countries, rubrics, inter-rater reliability, and (if possible given resources) the consistency in which benchmark scripts are scored across countries.

DATA AND REPORTING

Data management and analysis

79. The AHELO Consortium is responsible for producing high-quality cleaned, verified and adjudicated international data files. These data files will be produced to internationally defined data standards, classifications and specifications as documented by the AHELO Consortium.

80. NPMs and the AHELO Consortium must agree that national data can be included in AHELO international data files.

81. AHELO data must be analysed using accepted psychometric techniques for a study of its kind. Analysis must be conducted in ways consistent with AHELO rationales and contexts, test and context instruments, data properties and constraints, and international methods as detailed in the AHELO Technical Manual. Item response modelling must be used for scaling and validation. Scaling methods must enable potential equating across time and contexts. Supplementary alternative analyses may be undertaken.

82. AHELO results must be fed back to participating HEIs in prescribed metrics, as defined in international AHELO reports. Information on variability (arising from measurement and sampling error) should be reported along with point estimates.

83. Sampling weights will be prepared by the AHELO Consortium to the most detailed extent feasible given methods, processes and data constraints. Standard errors should be produced using appropriate technical procedures.

84. AHELO data must be analysed using appropriate statistical techniques for a study of its kind. Statistical analyses must use relevant weights, take account of estimate properties and sampling errors, and make adjustments where necessary for subgroup differences.

85. Certain AHELO data may be released for information or research purposes. Third-party access to confidential/unreleased AHELO data and files is governed by the OECD and AHELO GNE policy on data access, use and reporting.

Reporting and interpretation

86. AHELO results must be reported in ways consistent with data, analyses, intended audiences and reporting rationales. Results should be presented and explained in ways consistent with institutional reports. Appropriate uncertainty estimates (confidence bands/standard errors) must be reported.

87. A series of international reports will be produced: AHELO Database and Codebooks, Compendia, AHELO Technical Report; AHELO Institution Reports; and AHELO Report.

88. Interpretation of AHELO results must take into account the purpose for which the assessment was conducted, the contexts of interpretation. The scope, constraints and limitations of results must be reported, with particular reference made to bias and lack of precision.

89. Access to AHELO data products and reports is set in the AHELO Data Access and Reporting Guidelines.

90. The AHELO Consortium will maintain an electronic archive of all assessment materials, manuals and scoring guides as prescribed by agreements with OECD.

91. Unless otherwise requested, NCs will archive all materials relating to administration, instrumentation and implementation until publication of the international report.

RESOURCES

American Educational Research Association, American Psychological Association and the National Council on Measurement in Education (AERA, APA and NCME) (2005). *Standards for Educational and Psychological Testing*. Washington: AERA.

Arora, A., Foy, P., Martin, M. & Mullis, I. (2009). *TIMSS Advanced 2008 Technical Report*. Chestnut Hill: TIMSS & PIRLS International Study Center.

Borden, V.M.H. & Owens, J.L.Z. (2001). *Measuring Quality: Choosing among surveys and other assessments of college quality*. Washington, D.C.: American Council on Education.

Carstens, R. & Pelgum, W. (2009). *SITES 2006 Technical Report*. Amsterdam: IEA Secretariat.

Darling-Hammond, L. (2002). *Standards, Assessments, and Educational Policy: In pursuit of genuine accountability*. Princeton, NJ: ETS.

- Gregory, K.D. & Martin, M.O. (2001). *Technical Standards for IEA Studies: An annotated bibliography*. Amsterdam: IEA Secretariat.
- Hembry, I. & Fina, A. (2010). *Resources for Assessment Development: A bibliography for the assessment community*. Accessed 2 February 2012, from: <http://ncme.org/linkservid/983E4748-1320-5CAE-6EFB9C4290DCE72E/showMeta/0>.
- International Test Commission (2000). *International Guidelines for Test Use*. Accessed 2 February 2012, from: <http://www.intestcom.org>
- International Test Commission (2005). *International Guidelines on Computer-based and Internet Delivered Testing*. Accessed 2 February 2012, from: <http://www.intestcom.org/guidelines>
- International Test Commission (2010). *International Test Commission Guidelines for Translating and Adapting Tests*. Accessed 2 February 2012, from: <http://www.intestcom.org>
- International Test Commission (2011). *ITC Guidelines for Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*. Accessed 2 February 2012, from: <http://www.intestcom.org>
- Joint Committee on Testing Practices (2004). *Code of Fair Testing Practices in Education*. Washington, DC: National Council on Measurement in Education.
- Martin, M.O. Mullis, I.V.S., Caldwell, N., Foy, P., Sibberns, H., Adams, R.J. & Rust, K. (1998). *Technical Standards for IEA Studies*. Unpublished draft.
- Martin, M.O. Rust, K & Adams, R.J. (1998). *Technical Standards for IEA Studies*. Amsterdam: IEA Secretariat.
- National Health and Medical Research Council, Australian Research Council, and Universities Australia (NHMRC, ARC, UA) (2007). *National Statement on Ethical Conduct in Human Research*. Canberra: Universities Australia
- OECD (2004). *Technical Standards for PISA 2006*. Paris: OECD.
- OECD (2007). *Technical Standards for PISA 2009*. Paris: OECD.
- OECD (2010). *AHELO Analysis Plan*. Paris: OECD.
- OECD (2010). *AHELO Assessment Design*. Paris: OECD.
- OECD (2010). *AHELO Focus Group Guidelines*. Paris: OECD.
- OECD (2010). *AHELO Translation and Adaptation Guidelines*. Paris: OECD.
- OECD (2010). *PISA 2012 Technical Standards*. Paris: OECD.
- OECD (2010). *TALIS Technical 2008 Report*. Paris: OECD.
- OECD (2011). *AHELO Focus Group Guidelines*. Paris: OECD.
- OECD (2011). *AHELO Institution Coordination Manual*. Paris: OECD.

- OECD (2011). *AHELO International Scoring Manual*. Paris: OECD.
- OECD (2011). *AHELO Sampling Manual*. Paris: OECD.
- OECD (2011). *AHELO Test Administration Manual*. Paris: OECD.
- OECD (2011). *Quality Framework and Guidelines for OECD Statistical Activities*. Paris: OECD.
- OECD (2012). *AHELO Analysis and Reporting Design*. Paris: OECD.
- OECD (2012). *AHELO Data Access, Use and Reporting Policy*. Paris: OECD.
- OECD (2012). *AHELO Economics Assessment Framework*. Paris: OECD.
- OECD (2012). *AHELO Economics Development Report*. Paris: OECD.
- OECD (2012). *AHELO Economics Scoring Guide*. Paris: OECD.
- OECD (2012). *AHELO Engineering Scoring Guide*. Paris: OECD.
- OECD (2012). *AHELO Engineering Assessment Framework*. Paris: OECD.
- OECD (2012). *AHELO Engineering Development Report*. Paris: OECD.
- OECD (2012). *AHELO Feasibility Study Country Data File (CDF)*. Paris: OECD.
- OECD (2012). *AHELO Feasibility Study Faculty Data File (FDF)*. Paris: OECD.
- OECD (2012). *AHELO Feasibility Study Institution Data File (IDF)*. Paris: OECD.
- OECD (2012). *AHELO Feasibility Study Student Data File (SDF)*. Paris: OECD.
- OECD (2012). *AHELO Generic Skills Assessment Framework*. Paris: OECD.
- OECD (2012). *AHELO Generic Skills Multiple Choice Question Development Report*. Paris: OECD.
- OECD (2012). *AHELO Generic Skills Scoring Guide*. Paris: OECD.
- OECD (2012). *AHELO NPM Manual*. Paris: OECD.
- OECD (2012). *AHELO Technical Standards*. Paris: OECD.
- OECD (2010). *PIAAC Technical Standards and Guidelines*. Paris: OECD.
- OECD (2012). *PISA 2009 Technical Report*. Paris: OECD.