

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

Group of National Experts on the AHELO Feasibility Study

AHELO FEASIBILITY STUDY ANALYSIS PLAN

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

JT03289934

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



**AHELO Feasibility Study
Analysis Plan**



INTRODUCTION

1. The purpose of the AHELO feasibility study is to assess whether it is possible to measure at the international level what undergraduate degree students know and can do, in order to provide better information than is currently available to higher education institutions, governments, and other stakeholders including students and employers. This entails evaluation of the scientific feasibility of undertaking an international assessment of higher education learning outcomes at the end of a bachelor degree programme, as well as gauging the feasibility of its practical implementation.

2. The AHELO Analysis Plan is derivative of the broader AHELO Assessment Design. It details an evaluation design that specifies research questions posed by the various strands, outlines the best methods of analysis to assess the cross-linguistic, cross-cultural and cross-institutional validity of the instruments being used in the various assessments and context surveys, and charts the sources of evidence to assess the various dimensions of the study. The Analysis Plan guides a formative evaluation of the study which is structured to deliver data and insights which will assist countries to decide on next steps.

3. In summary, evaluation of evidence from the AHELO feasibility study will help determine the scientific and practical feasibility of conducting an international assessment of higher education learning outcomes on a larger scale. The evaluation is designed to synthesise a vast and diverse amount of evidence into a finding on whether the AHELO Feasibility Study has provided a proof of concept. While the collection of evidence will be undertaken throughout the study, the evaluation will largely be of a summative kind. It will consider the extent to which AHELO has produced a feasible assessment.

EVALUATION DESIGN

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

5. The AHELO Feasibility Study has scientific and practical dimensions. There are two research questions:

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

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

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

7. This AHELO Analysis Plan presents these questions in an operationalisable framework. The framework provides a guide for the research, and helps review the outcomes and success of the study. The framework is presented in Table 1 and Table 2 below, each of which contains columns headed:

- Research questions:

The AHELO research questions map out the high-level issues that are the focus of the study. There are a relatively small number of these, and they cover scientific and practical factors of broad concern in the study. In simple terms, answering ‘yes’ to these questions means that AHELO is a success.

- Feasibility indicators:

The AHELO feasibility indicators operationalise the research questions in terms of more specific objectives. These indicators focus attention on contexts, processes and resources that will be used to evaluate the study’s success. Most of these indicators are described in a positive direction (such that ‘more’ is ‘better’), however some are expressed in a reverse format.

- Evidence sources:

To enhance the rigour of the process and value of the outcomes, various evidence sources will be consulted. The main sources (along with their abbreviation used in Table 1 and Table 2) include:

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

8. A range of evidence will be collected to 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 also be sourced.

9. Where possible the feasibility indicators have been specified to be analysed in quantifiable ways. Examples of the forms of quantitative evidence to be consulted include:

- results of psychometric and statistical analyses of test and context data (the most significant evidence);
- results from surveys conducted with various stakeholders during the analysis phase; and
- counts of various events that arise during the study.

10. Given the complex nature of many facets of the work it is often necessary to defer to qualitative evidence. Typical forms of qualitative evidence include:

- feedback and notes from meetings;
- feedback from formal consultations; and
- informal observations arising during design and implementation.

11. Each of the derivative research questions touches on aspects of scientific and practical feasibility. Rather than separate them along these lines, therefore, Table 1 presents material relevant to instrumentation, and Table 2 presents material relevant to implementation.

12. This document covers the entire study, and the capacity to undertake the progress the evaluation hinges on the focus, scope, approach and scheduling of related facets of work. It is important to note that the evaluation of feasibility will be framed by the constraints imposed on the feasibility study. Certain facets of a full-scale study have been deliberately excluded from the feasibility study. For instance, only provisional frameworks and partial instruments will be developed for the Economics and Engineering strands.

ANALYTICAL FRAMEWORK

Table 1: Instrumentation considerations

Research question	Feasibility indicator	Evidence sources
1. Was the Generic Skills assessment framework reflective of an international consensus about the areas that are important to assess?	<ul style="list-style-type: none"> • Achieved consensus among experts on the domain to be assessed • Achieved consensus among experts on the balance given to elements in the domain • Degree to which framework accounts for institutional, language and cultural differences 	TAG, NPM
2. Was the instrumentation developed on the basis of the Generic Skills framework faithful to the spirit and intent of the framework?	<ul style="list-style-type: none"> • Construct, content and face validity of instrument considered appropriate • Instrument has appropriate levels of reliability • Differential item functioning within acceptable limits 	AC, FR,SR, TAG

Research question	Feasibility indicator	Evidence sources
3. Was the provisional Economics assessment framework reflective of an international consensus about the important learning outcomes in Economics?	<ul style="list-style-type: none"> • Achieved consensus among experts on the domain to be assessed • Achieved consensus among experts on the balance given to elements in the domain • Degree to which framework accounts for institutional, language and cultural differences • Acceptance by a broader group of stakeholders and the community 	TAG, EG, NPM
4. Was the instrumentation developed on the basis of the Economics framework faithful to the spirit and intent of the framework?	<ul style="list-style-type: none"> • Construct, content and face validity of instrument considered appropriate • Instrument has appropriate levels of reliability • Differential item functioning within acceptable limits • Acceptance by the broader economics community 	AC, EG, FR, SR, TAG
5. Was the provisional Engineering assessment framework reflective of an international consensus about the important learning outcomes in Engineering?	<ul style="list-style-type: none"> • Achieved consensus among experts on the domain to be assessed • Achieved consensus among experts on the balance given to elements in the domain • Degree to which framework accounts for institutional, language and cultural differences 	TAG, EG, NPM
6. Was the instrumentation developed on the basis of the Engineering framework faithful to the spirit and intent of the framework?	<ul style="list-style-type: none"> • Construct, content and face validity of instrument considered appropriate • Instrument has appropriate levels of reliability • Differential item functioning within acceptable limits • Acceptance by the broader engineering community 	AC, EG, FR, SR, TAG
7. Was the provisional Contextual Dimension conceptual framework reflective of an international consensus about the important contexts that shape higher education learning outcomes?	<ul style="list-style-type: none"> • Achieved consensus among experts on the domain to be assessed • Achieved consensus among experts on the balance given to elements in the domain • Degree to which framework accounts for institutional, language and cultural differences 	TAG, EG, NPM, SCG
8. Was the instrumentation developed on the basis of the Contextual Dimension framework faithful to the spirit and intent of the framework?	<ul style="list-style-type: none"> • Construct, content and face validity of instrument considered appropriate • Instrument has appropriate levels of reliability • Differential item functioning within acceptable limits 	AC, EG, FR, SCG, SR, TAG

Table 2: Implementation feasibility considerations

Research question	Feasibility indicator	Evidence sources
9. Was the AHELO Assessment Design valid and feasible?	<ul style="list-style-type: none"> • Relevant scope for an international assessment of learning outcomes • Degree of change to design during implementation, analysis and reporting • Degree to which design provides basis for full-scale extension 	TAG, AC, EG, GNE, NPM, OECD, SCG
10. Was the study implemented in a methodologically rigorous fashion?	<ul style="list-style-type: none"> • Count of significant variations between design and operationalisation • Positive feedback on technical design and operationalisation • Degree to which the methodology advances higher education research 	TAG, AC, EG, NPM, OECD, SCG
11. Was the study feasible from a practical perspective?	<ul style="list-style-type: none"> • Study delivered within agreed timeframe and budget • NPMs deliver within agreed timeframe and budget • ICs deliver within agreed timeframe and budget • Positive feedback from institutions on the value of the study • Extent to which communication processes enable the smooth flow of work • Extent to which information technology led to management efficiencies • Count of significant practical implementation problems encountered? 	TAG, AC, GNE, IC, NPM, OECD, TA
12. Was the study successfully generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally?	<ul style="list-style-type: none"> • Count of errors in verifier reports • Count of translation and adaptation errors remaining after verification • Count of translation and reproduction errors in final instrumentation • All items translated and verified in time for implementation • Broad acceptance of instruments in each country • No linguistic bias evident in assessment data 	TAG, AC, EG, FR, IC, NPM, SR
13. Was it possible to engage systems and institutions in the study?	<ul style="list-style-type: none"> • Count of countries involved compared with expected • Count of institutions per country compared with expected • Count of institutions in which assessment implemented successfully • Count of institutions that discontinued involvement during study 	TAG, AC, GNE, IC, NPM, OECD, SCG

Research question	Feasibility indicator	Evidence sources
14. Was it possible to engage faculty respondents in the study?	<ul style="list-style-type: none"> • Count of institutions in which target population defined according to international specifications • Number and percentage of faculty involved in the assessment • Faculty ratings of quality and significance of assessment • Percentage of context assessments completed within timeframe • IC report of positive engagement among faculty 	AC, FR, IC, NPM, TA, TAG
15. Was it possible to engage a random sample of student respondents in the study?	<ul style="list-style-type: none"> • Count of institutions in which target population defined according to international specifications • Number and percentage of students involved in the assessment • Faculty ratings of quality and significance of assessment • Percentage of context assessments completed within timeframe • IC report of positive engagement among students 	AC, IC, NPM, SR, TA, TAG
16. Was the instrument delivered successfully?	<ul style="list-style-type: none"> • Count of disruptive information technology problems during design phase • Count of disruptive information technology problems during administration • Count of information technology problems influencing analysis 	TAG, AC, EG, IC, NPM, TA
17. Were survey operations successfully managed by countries and institutions?	<ul style="list-style-type: none"> • Training of NPMs implemented in timely manner • Count of major fieldwork errors that caused disruption to project schedule • Count of major deviations from quality assurance criteria • Number of testing sessions that did not adhere to prescribed conditions • Number of security breaches • Satisfaction with NPM training captured via surveys of various stakeholders • Number of manuals that could be re-used in a full-scale replication • Extent of serious problems with data quality 	TAG, AC, EG, IC, NPM, TA
18. Was it possible to train people in different countries to score tasks in cross-linguistically and cross-culturally generalisable ways?	<ul style="list-style-type: none"> • Evidence of any significant difference in conceptualisation of coding guides • Number of coders trained in face-to-face and online workshops • Extent of psychometric difference in application of coding guides between scorers 	TAG, EG, IC, NPM
19. Were statistical reports of assessment results of value to systems and institutions?	<ul style="list-style-type: none"> • Number of participating systems and institutions that would participate in study again • Degree to which results are seen as having value for policy development, quality assurance and continuous improvement 	TAG, EG, GNE, IC, NPM, SCG, OECD