

Assessment of Higher Education Learning Outcomes

Feasibility Study Report

Volume 1 – Design and implementation

Executive Summary

*This **first volume** will be followed by a second volume on Data Analysis and National Experiences at the end of February and a third and final volume on Further Insights (including the March Conference Proceedings) at the end of April.*

*The complete Volume 1 can be found on the AHELO website (www.oecd.org/edu/ahelo)
<http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf>*



In a complex, ever changing and growing higher education context, where a variety of rankings are often being used as the yardstick of academic excellence, there is a clear need for a way to effectively measure the actual outcomes of teaching and learning.

The OECD's Assessment of Higher Education Learning Outcomes (AHELO) aims to provide a tool for Higher Education Institutions (HEIs) to assess what their students know and can do upon graduation, on an international scale. This tool can play a key role in supporting of HEIs in their efforts to improve the learning outcomes for their students.

The feasibility study, which concluded at the end of 2012, was conducted to test the scientific and practical feasibility of such an assessment across diverse national, cultural, linguistic, and institutional contexts.

Participating countries and various international expert groups participated in the development of assessment instruments for the three strands chosen for the purpose of the feasibility study (generic skills, economics and engineering) and the contextual questionnaires which were then administered to students in participating institutions. This implementation of the test involved a total of 249 HEIs in the 17 countries and regions, and the instruments were administered to close to 4 900 faculty and some 23 000 students (close to the end of a Bachelor's Degree).

The first volume of the feasibility study report examines the process for the initial development of the feasibility study (why it was thought necessary to begin with, how it was developed, the challenges which had to be overcome, etc.) and the instrumentation and implementation of the survey.

Chapter 1 - The rationale for an AHELO: Higher Education in the 21st Century context

There is widespread recognition that skills and human capital have become the backbone of economic prosperity and social well-being in the 21st century. In this context, higher education plays a central role and has become increasingly important on national agendas.

Global trends in higher education

A number of global trends have shaped the development and wide-ranging mutations of higher education over the past half century. Once defined in terms of an "ivory tower" model of traditional research universities attended by the elites, higher education today is characterised by massive expansion and wider participation; the emergence of new types of institutions (vocationally-oriented and private providers); more diverse profiles of institutions, programmes and their students; broader adoption and more integrated use of communications and educational technologies; greater internationalisation, competition and signalling mechanisms; growing pressures on costs and new forms of financing; as well as new modes and roles of governance, including increasing emphasis on performance, quality and accountability.

The quality challenge and limitations of diverse attempts to fill the quality information gap

With the massive expansion of higher education systems and wider participation, there are persistent concerns on the quality and relevance of students' preparation for higher education in spite of the development of more sophisticated quality assurance systems in most OECD

countries since the early 1980s. OECD Education Ministers meeting in Athens in June 2006 committed their countries to the goal of raising the quality of higher education.

But despite quality assurance systems and the remarkable development of performance indicators worldwide, data on the learning outcomes of higher education remains scarce in many systems. The OECD Thematic Review of Tertiary Education recommended in 2008 to increase the focus on student outcomes to alleviate this problem and to address concerns about quality of learning outcomes in higher education.

The rationale for an AHELO

The proposition to explore the development of an international Assessment of Higher Education Learning Outcomes (AHELO) emerged during the Athens Meeting of OECD Education Ministers in 2006, at a time of great pressure to develop better performance metrics in higher education. This proposition is illustrative of a paradigm shift for higher education, which manifests itself in several ways:

- ***Move beyond collegial approaches to governance*** - Higher education has seen a shift of emphasis from collegial approaches of governance by communities of scholars towards models combining greater autonomy with increased transparency and accountability. This new paradigm has led to increased demands for institutions to engage in outcomes assessment.
- ***Growing focus on student learning outcomes*** - Another trend sees a shift away from inputs towards outcome-based notions of higher education throughput. This shift has been most evident with the Bologna Declaration which aimed at establishing a European Higher Education Area and to write all higher education programmes in terms of learning outcomes by 2010. This trend is becoming global with many countries aligning their systems to be Bologna-compatible.
- ***Emphasis on student centred learning and research on teaching-learning processes*** - The turn of the Century has also seen a shift in undergraduate education, from an "instruction paradigm" towards a "learning paradigm" in which the emphasis is no longer on the means but on the end. A corollary of this emphasis is to better understand the teaching-learning interplay. In this context, outcomes' assessments are important for the evaluation of instructional effectiveness.
- ***AHELO within the broader movement towards competencies and learning outcomes*** - While AHELO is the first international attempt at measuring learning outcomes across borders, languages and cultures, it is part of a broader context of distinct initiatives converging in their focus on performance, competencies and learning outcomes. These include initiatives collecting data to be used as indirect proxies of outcomes and performance, focusing on the definition of expected learning outcomes or attempting to measure learning outcomes achieved by students.

Together, these patterns have shaped the context of the OECD AHELO initiative. Learning outcomes are indeed a key factor of institutional performance but while indirect evidence can

be gained from proxy measures there are no instruments for international measurement. AHELO has the potential to fill this gap by directly assessing student performance and the feasibility study was designed to see whether it could be done.

Chapter 2 - The beginning of AHELO: decisions and challenges

Questioning feasibility: Key challenges in developing and operationalising an AHELO

The notion of exploring the development of an international AHELO generated much discussion throughout 2006-2008 in higher education circles. Some policy makers, experts and stakeholders embraced the idea enthusiastically and participating actively in its development, while others were more critical of the approach and vocal in pointing to its potential risks.

Overall, concerns focused on the following points:

- risk that AHELO data would be used:
 - as a ranking, despite the fact that AHELO was never envisaged as a ranking tool within OECD circles.
 - as a basis for (re-)allocating public resources.
 - as a basis for reallocating funding within HEIs towards teaching to the detriment of other missions.
- the complexity of engaging in fair comparisons of extremely diverse institutions in terms of their missions, profiles and student bodies. Detractors highlighted the limited information that standardised tests can yield for institutions and faculties and the risk of simplistic conclusions.
- the potential impact on institutional autonomy and academic freedom and fears that AHELO might be forced on institutions and could over time yield homogenisation and constrain academic freedom.
- the merits and applicability of AHELO's focus on generic skills given different academic traditions in different parts of the world and fundamental debates on the relevance of assessing generic skills independently from the disciplines.

Expert meetings framed a roadmap

Fully recognising the specificities of higher education and the particular challenges associated with developing an assessment of learning outcomes at that level, the OECD convened three international expert meetings throughout 2007 to provide recommendations on how to take the AHELO idea forward. The experts:

- identified considerable challenges to developing internationally comparative measures of higher education learning outcomes, although none considered the goal unattainable
- recommended establishing measures of learning outcomes at the level of institutions or departments, at least initially.

- advised to focus on students at the end of a three or four-year degree as a meaningful target population, and agreed that there was little point in reporting data at the level of individual students.
- acknowledged that there was no generally accepted definition of what higher education outcomes ought to be, and suggested that the feasibility study should encompass two strands – one on transversal higher-order competencies and another one on discipline-related competencies.
- stressed that a fully developed assessment was beyond the scope of this initial project, and instead suggested to draw upon existing assessments, flagging the CLA approach for the generic skills strand due to its high degree of face-validity.
- agreed that an assessment of subject-specific competencies would also be needed, and recommended conducting the feasibility study in subject areas that have a stable core of methodologies such as engineering or economics.
- suggested to draw on a diverse sample of countries and to involve a limited set of volunteer institutions for the feasibility study.
- recommended that the two purposes of the feasibility study should be to test whether an AHELO is scientifically and practically feasible. Given this purpose, the assessment for the feasibility study could be constructed from existing instruments.
- agreed that the feasibility study should cover a minimum of three countries and languages, and seek a cross-national consensus on desirable outcomes of higher education.
- recommended that the assessment should be computer delivered, and that two hours was a reasonable duration for individual students. In the opinion of the experts, the information that the feasibility study would provide should be an appealing incentive for institutions to participate, but they expected that it would be harder to motivate students in the start-up phase.

These insights and recommendations were subsequently shared with OECD education ministers during an informal ministerial meeting held in Tokyo in early 2008, in which Ministers welcomed the OECD initiative to assess the feasibility of an international study on assessment of learning outcomes.

Chapter 3 - Design and management of the feasibility study

Survey design

The AHELO feasibility study was envisaged, from the outset, as a research exercise rather than a pilot study for a fully-fledged AHELO survey. This had significant implications for the study design, and this approach:

- focused on gathering evidence in support of the AHELO concept by building as much as possible upon existing tools and instrument materials rather than developing new ones. This implied that the instruments chosen for the feasibility study would not in any way prejudge the instruments and tools of an eventual AHELO follow-up main survey.
- artificially broke down the work into several strands of work to examine different types of learning outcomes as well as different approaches to assessment and testing. This purely artificial approach would yield insights on the relative merits and drawbacks of various methodologies and assessment domains.
- guided the selection of disciplines. As stakeholders believed that developing an international assessment would be far more challenging in the social sciences than STEM, it was decided to focus on two contrasting disciplines from the STEM and social sciences areas.

The feasibility study design consisted of four distinct strands of work: an assessment in each of three domains: generic skills, economics and engineering and value-added measurement analysed from a research perspective (rather than direct measurement).

The feasibility study involved 17 countries or regions whose distribution across the three domain strands ensured a balance of geographic, linguistic and cultural backgrounds to take cultural and linguistic differences into account.

The study design also embraced some of the stakeholders' recommendations, including a Tuning-AHELO project to gauge whether consensus on learning outcomes was possible, a strong contextual dimension, a consultation mechanism with stakeholders and scrutiny of AHELO's suitability to diverse institution types.

While the three assessment strands were to be undertaken separately, the study design envisaged that they be carried out coherently in terms of processes, test administration and analysis, so as to maximise synergies across strands, streamline communications and generate economies of scale. To achieve such coherence, the three strands were brought under the umbrella of a single Consortium.

Constraints and implications on the survey design

The feasibility nature of the project meant that the study was to be funded through grants and voluntary contributions, as is commonly the case for new cutting-edge OECD projects. To secure financial support, the first step was to convene a critical mass of OECD countries to participate in the feasibility study. By December 2008, ten countries had committed to the study. However, the global financial crisis soon after the project was launched affected the private sector and philanthropic foundations and undermined fundraising efforts. This forced participants to extend or revise their arrangements.

While the feasibility study benefited from the continued commitment and support of its participants, the study design had to be adapted in the context of fundraising challenges. The OECD negotiated with the two main contractors to reduce their costs without sacrificing the study goals and widened country participation to diversify and broaden support, for the study

to proceed. In addition, the project design and timeline were reviewed and it was agreed to phase the work in line with funding availability, thereby allowing keeping momentum with the work, even if for more limited activities than initially envisaged.

Phases of work

The study thus proceeded in three phases.

- ***Providing an initial proof of concept*** by adapting and/or developing provisional assessment frameworks and instruments for an international context, and validating them across national, linguistic, cultural and institutional contexts through small-scale testing in participating countries.
- ***Implementing these instruments*** in the field to evaluate the scientific and practical feasibility of an AHELO, concentrating on the practical aspects of test administration, scoring of student responses, data analysis, and reporting on feasibility.
- ***Considering methodologies and approaches to capture value-added***, *i.e.* the contribution of institutions to students' outcomes after controlling for their incoming abilities. Indeed, the feasibility study was also tasked to provide insights on whether it would be feasible to measure growth in learning.

Study management and actors

The design and implementation of the AHELO feasibility study entailed collaboration among representatives of participating countries, an international Consortium and the OECD Secretariat. A number of expert groups and sub-actors were also involved in the conduct of the work.

Chapter 4 - Instrument development

The quality of the AHELO feasibility study results relies greatly on the quality of the instruments used to assess students' learning outcomes and to capture contextual background.

The instrumentation process

The development of assessment and survey instruments in cross-cultural educational settings usually involves a wide range of activities that can be grouped in five steps.

- ***Developing assessment frameworks*** to establish the purpose of the assessment and provide a clear definition of what is being assessed, a description of the items to be used and the basis for interpreting the test results. "Provisional frameworks" were developed through a process including audits of existing materials, review of draft framework specification by international expert groups, national consultations for validation, and final review. Fast-tracked procedures were used for the Generic Skills framework.
- ***Developing assessment instruments*** through the creation or selection of items to match the table of specifications of the frameworks. Two types of items were developed for each

assessment: constructed-response tasks in which students provide their own responses and multiple-choice items in which they choose the correct answer.

- ***Translating and adapting assessment instruments and surveys*** entails the national translation and adaptation of the assessment and survey instruments to ensure cross-language comparability. A set of quality assurance and quality control procedures was adopted.
- ***Small-scale validation of the instruments*** through pilot-testing the items developed with students similar to the target population. This provides an opportunity to assess the quality and appropriateness of items, and to collect feedback on the instrument length, difficulty and relevance.
- ***Final review of the assessment and survey instruments*** using results from small-scale validation activities, feedback collected from respondents, consultations conducted with stakeholders.

The Generic Skills assessment

The Generic Skills instrumentation did not follow the usual development process described above. The study design sought to adapt an existing instrument and did not include developing an international version of an assessment framework. Accordingly, work started with the adaptation of one component of the existing Collegiate Learning Assessment (CLA) instrument.

This constructed-response section was complemented by multi-choice questions drawing upon another existing generic skills test. As the instrumentation work evolved, it appeared that an international consensus on the generic skills to be measured was lacking. Work on the Generic Skills framework thus began afterwards, but time and budget constraints did not allow its oversight by a Generic Skills expert group.

The Economics and Engineering assessments

The development of instrumentation was similar for both discipline strands. The Economics and Engineering frameworks built upon the Tuning-AHELO frameworks and other background material on teaching in these two disciplines, and were both endorsed by international Economics and Engineering Expert Groups. They fulfil the requirement of well-designed frameworks by defining the domain to be tested, specifying the expected learning outcomes, and offering an overview of the instrumentation required to measure the competencies. Numerous cycles of consultation were undertaken.

International development and validation of the frameworks suggest it is possible to define discipline-specific learning outcomes internationally – an issue that was far from certain at the outset of the study, especially for Economics. Assessment instruments were developed in cooperation between test developers and expert groups, and were reviewed through cognitive labs and panelling. Qualitative validation was completed by focus groups of students in various institutions within each country.

The Contextual surveys

The development of the contextual framework and survey instruments followed a similar process and built upon work carried out by the OECD at the outset of the study. Three context survey instruments were developed to collect background information from students, faculties and institutions. Broad consultation took place to seek the opinion of a range of groups and individuals. In addition, focus groups were held to capture student and faculty insights into the range and characteristics of the items.

Localisation of assessment and survey instruments

All source instruments were prepared in English. For each field implementation, a number of distinct assessment and survey instruments required translation and adaptation. As in many international studies, a decentralised model was adopted in which National Centres were each responsible for localising materials for use in their respective systems while the AHELO Consortium guided and assisted them throughout the process. Three types of workflows were used. Overall, assessment instruments were localised using dual translation, reconciliation and verification (although with two distinct workflows), while centralised translation was used for the Generic Skills MCQs and contextual surveys due to time constraints.

Chapter 5 - Implementation

Higher education systems and volunteer institutions played a key role in assessing the practical feasibility of AHELO implementation, *i.e.* participants' selection, assessment delivery, response rates and scoring.

Management of AHELO field implementation in countries and institutions

Participating countries had to set up their national infrastructure and the National Project Manager (NPM) or National Centre coordinated all AHELO activities within the country, together with Institution Co-ordinators, Test Administrators and Lead Scorers. Communications were an important aspect of the practical implementation of the feasibility study. Training courses and instruction manuals also played a key role in describing explicit procedures for each distinct AHELO activity.

The timetable for field implementation was relatively short - from January to May 2012. Overall, most countries adhered to scheduled timelines although the timeframe posed challenges for some NPMs in defining student populations, finalising samples and scheduling testing with many students on internships, taking exams or working on research projects at that time. Overall, organisational arrangements worked well.

Selection of institutions and securing institutional engagement

NPMs were asked to recruit ten volunteer institutions per strand while trying to get a mix of institutions that would reflect the diversity of their higher education system. National Centres developed a range of institutional engagement strategies to recruit institutions, including websites to promote AHELO. All NPMs reported that invited institutions showed great interest in the study and only very few withdrew.

Sampling of students and faculty

Systematic equal probability sampling was used for the selection of student at each participating institution, unless there were less than 200 students, in which case, all students were included. This required each institution to provide a complete sampling frame (i.e. target population of those at the end of a 3-4-year undergraduate degree). This proved challenging as sampling has not been widely used in the higher education context and only three-quarters of them were able to provide the sampling frame information that met the technical standards.

For faculties, a systematic equal probability sample of 40 faculty had to be drawn from the lists of in-scope faculty (i.e. those teaching undergraduates). In many cases, institutions struggled to define in-scope faculty and only about half of them managed to provide the sampling frame information.

Electronic delivery

Student testing, data collection and scoring were all undertaken online through two different platforms (one for the CLA component of the Generic Skills assessment, and another one for all other instruments). Testing between both platforms was conducted to ensure a seamless transition.

The AHELO test system was used by students, faculty and institution coordinators between January and July 2012. Testing involved almost 23 000 students. Some institutions tested all students in one session, whereas others organised more than 20 test sessions. Generally, bigger testing groups and rooms were more challenging to coordinate than smaller groups of up to 20 students. The majority of testing sessions ran smoothly.

Response rates and incentives

A key challenge for the fieldwork was engaging and motivating sampled students and faculty to ensure they participated in the survey or test session. Student response rates varied quite dramatically across countries and strands, with higher response rates in the Engineering and Economics strands than in the Generic Skills strand, and for institutions using a census.

HEIs relied on a range of strategies of student recruitment involving communication, marketing and incentives. Cash or vouchers were the most common incentives, followed by a certificate of participation, prize draws or gifts, food or drinks, or academic bonuses in the form of increased grades. Some institutions also made participation in AHELO compulsory for students.

Overall, feedback from NPMs suggests that material incentives were not very useful to (significantly) increase student response rates, while institutional culture and strategies were of critical importance. As with students, faculty response rates were highest in the Engineering strand, and wherever a census was used. Overall, feedback from NPMs suggests that it was much easier to get faculty than students involved.

Scoring student responses

In each of the three strands, the instrument included both multiple-choice questions that were scored automatically and constructed response tasks which required human scoring by teams

of scorers within each country. Scoring quality is critical to the reliability of the interpretation of results and demands close adherence to standard scoring procedures, even more so in studies involving more than one language, culture or country. A Scoring Manual, strand-specific scoring guides, and face-to-face training of lead scorers were provided by the Consortium. All student responses in the Generic Skills strand and 20% in the discipline strands were double-scored to enable lead scorers to monitor the performance of their team.

Chapter 6 - Lessons learnt on design and implementation

The AHELO feasibility study was designed to establish whether an international survey could be developed to support better learning outcomes in higher education. Along the way, AHELO has also sharpened awareness of the importance of learning outcomes as a core part of higher education mission and learning outcomes have now moved to centre stage of the discussions about higher education performance. The feasibility study has also brought to the forefront many of the complexities and challenges of measuring learning outcomes.

Purpose of AHELO – to support improvement in learning outcomes

A key lesson from the feasibility study experience is the importance of establishing the purpose of AHELO and communicating clearly not only what AHELO is but also what it is not. Looking ahead to further AHELO development, it would be important to re-emphasise that AHELO is intended as a tool for institutional improvement and to ensure that future instrument(s) would actually help institutions achieve better learning outcomes. Engaging with key stakeholders is critical to achieving this.

Feasibility study design

The design of the feasibility study deliberately sought to confront the challenges associated with diversity and, as more countries joined, the final set of participants encompassed even more diversity than originally planned. Overall, this diversity proved to be a source of added richness to the feasibility study. Efforts to bring together diverse experts to define learning outcomes and develop the instruments paid off and the proof of concept was successful.

The feasibility study chose institutions as units of analysis. While this proved to be a reasonable approach and the volunteer institutions seemed highly motivated, a follow-up survey of participating institutions would be useful to probe deeply into which aspects and data are most useful and attractive to them.

The feasibility study process also affirmed the importance of seeking input from key stakeholders and the value of consultative processes. These consultations brought invaluable perspectives and any future development would only be enhanced by even greater involvement of all stakeholder communities.

The study design adopted an artificial distinction into three separate strands of work. In many ways, the discipline strands proved more straightforward to implement. However the relative merits of adopting a discipline-based *versus* generic skills approach in the future needs to

reflect further consideration of the relevant learning outcomes for different institutions and how they would want to use the results for improvement.

Management of the feasibility study

Uncertainties about funding were a key factor significantly impinging on the effective management of the feasibility study. A key lesson from this experience is the importance of securing early commitment from a critical mass of countries and sufficient funding for future development to warrant moving forward.

A feature of the study was the establishment of the AHELO Consortium to bring synergies and additional expertise to the study. While this approach worked well overall, it relied on effective collaboration between different partners within the Consortium, which was not always straightforward. This affirms the importance of ensuring that future contracting arrangements are as clear and straightforward as possible, communication channels explicit, and that the most comprehensive tendering processes are used.

Instrument development

A major challenge was to demonstrate that an assessment framework could be agreed upon across diverse country and institutional settings. This process went smoothly in the discipline strands, where it proved easier than expected to get agreement amongst discipline experts (including in Economics) on what AHELO should cover and measure. By contrast, the Generic Skills framework was only developed late in the process, thus limiting the scope for expert consensus. A lesson from this experience is that establishing international consensus on the assessment framework should be an essential upstream part of the instrument development process. Another lesson is that what might have seemed at the time as a reasonable short-cut may not have been the most cost-effective approach after all. This underlines the importance of developing completely new tailor-made instruments for any future AHELO.

Fieldwork revealed that institutions wish to provide students with feedback on their performance, which the use of matrix sampling (*i.e.* not asking all students to answer all questions) precluded. Trade-offs between the advantages of matrix sampling and being able to provide feedback to students should be reconsidered.

All assessment instruments combined MCQs and CRTs. A further lesson is to consider more deeply the relative appeal of both approaches for students and faculty. Finding the best mix of item types also needs to be linked more clearly to how institutions and faculty would want to use results. Finally, a well-designed contextual dimension is critical to any future AHELO.

Field implementation

Overall, field implementation proceeded very well and the use of electronic delivery was a major success, despite some glitches. However, timelines were generally too compressed and would need to be extended in any future AHELO, as well as be better aligned to each country's academic year.

For many countries and institutions, participating in the feasibility study meant establishing entirely new structures and processes to carry out the work. For the future, participation

readiness criteria could be drawn up to ensure that participants have all the systems and mechanisms needed to ensure that the results produced are reliable. This needs to be backed up by a strong external quality control function.

One of the biggest challenges in delivery was defining samples and obtaining adequate response rates. A number of different suggestions were offered to address this issue, making it clear that further research is needed into what measures would be most effective in raising response rates.

Reflections on teaching and learning

Finally, a number of countries felt they got an extra bonus out of participating in the feasibility study, because of the deep reflections that this induced about teaching and learning. This is perhaps the most important lesson from the feasibility study: that the assessment of higher education outcomes is not an end in itself, but rather a stimulus to deeper professional dialogue on desired learning outcomes and the teaching approaches needed to achieve them, for example:

“The students reported that curiosity of their own achievements was their main motivation for taking part in AHELO.” (Slovak Republic)

“For the generic skills strand, most students answered that they were unfamiliar with constructed-response tests, but they considered them engaging and challenging [...] they would like them to be used during the learning process.” (Mexico)

“The scorer training and scoring exercise proved to be eye openers [...] This exercise was instrumental in generating clearer understandings of the conceptual framework of competencies, and encouraged critical reflections on teaching practices.” (Japan)

“The framing of questions to make students ‘think like an engineer’ was innovative to some [faculty] – prompting them to [re-]consider how they taught their students, what they expected of them and how they were assessed.” (Ontario, Canada)