

# Assessment of Higher Education Learning Outcomes Feasibility Study Report

## Volume 2 – Data Analysis and National Experiences

### Executive Summary

*This **second volume** will be followed by a third and final volume on Further Insights (including the March Conference Proceedings) at the end of April.  
The first volume on Design and Implementation was published in December 2012.*

The complete Volume 2 can be found on the AHELO website ([www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo))  
<http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume2.pdf>

## Chapter 7 - Validity and reliability insights on scientific feasibility from the AHELO feasibility study data

*This chapter was prepared on the basis of the information available at the time of publication. However the unavailability of certain information did not allow OECD analysts or external experts to replicate or complement the information and analyses the OECD has received. Also, because of the unavailability of some of the psychometric results, the inclusion of the associated conclusions in this report does not imply the OECD's endorsement of the conclusions.*

The scientific feasibility of AHELO rests on its capacity to produce valid and reliable results across different countries, languages, cultures and institutional settings. This chapter presents an overview of the data collected and analyses conducted in order to assess the scientific feasibility of the instruments that were used. These analyses and results presented should be interpreted in the “proof of concept” spirit of the feasibility study.

### **Validity and reliability concepts for assessing scientific feasibility**

**Validity** is a broad concept that involves making appropriate interpretation and uses of test scores. It requires that the purpose and inferences to be drawn from test scores be stated from the outset. The evaluation of instrument validity requires the collection of a variety of evidence to support different types of validity.

**Reliability** means that test results are consistent and stable across different testing situations. An instrument's degree of reliability can be affected by a number of different factors. Stable results suggest that the observed student scores are more likely to reflect true scores. Reliability of an instrument is classically expressed as the ratio between the true variance, i.e. the true ability, and the observed variance, i.e. the observed test scores that include random factors.

### **Evidence on scientific feasibility collected during the AHELO feasibility study**

Some individual test items may turn out to perform poorly and must be removed before validity and reliability can be assessed. Items not meeting psychometric standards are deleted from final analyses. Non-functioning items can be removed on a country basis.

The small number of items removed from the generic skills instrument indicates that it has good overall item quality while the relatively small number of items deleted from the economics and engineering instruments indicates they have sufficient overall item quality.

In an international study, it is critical to ensure that items have a similar level of difficulty across the different countries. Differential item functioning (DIF) analyses are conducted to further understand differences in performance of different student sub-populations. Most items showed no significant differences in performance between genders. For those who did, further analyses would be needed to identify the underlying reasons for these differences by gender.

Three institutional characteristics are used as a basis for comparison across the different types of higher education institutions. Results indicate difference in student performance across the different institution types, depending on the basis for comparison.

- Little difference in student performance is observed when using the institution size (small/medium/large).
- Differences in student performance are observed when compared on the basis of highest degree the institution offers (baccalaureate, master and doctorate).
- Differences in student performance are observed when using the institution emphasis on research and teaching (research, teaching, and research/teaching balance).

Many items do not function as expected for some countries. Further analyses are needed to identify the underlying reasons for these country differences. Many items do not function as expected for some languages. Constructed-response tasks in the generic skills strand show significant differences in student performance across languages. Further analyses are needed to identify the underlying reasons for **student** performance differences in the different languages.

### **Validity evidence**

Different types of evidence are collected throughout the feasibility study to determine the validity of instruments used.

The three assessment instruments display reasonable levels of construct validity evidence. Results indicate that the overall scale could be divided into complementary sub-scales.

Expert consensus provided evidence of content validity of the economics and engineering instruments but was not fully demonstrated for the generic skills instrument. Feedback from the generic skills cognitive labs showed that the constructed-response tasks were attractive to students. Students also reacted positively to the draft economics and engineering constructed-response tasks. Further content validity evidence for the two discipline instruments is still required to fully confirm content validity.

Face validity is assessed through several indicators.

- Students spent a good deal of time responding to the AHELO assessments. The low levels of non-response indicate good levels of student engagement with the instruments.
- Student reported putting a good deal of effort into the AHELO assessments. Self-reported effort by field of education for students participating in the generic skills strand also reveals limited variations across fields.
- Students' perceptions of the educational and professional relevance of the instruments vary across strands and also reveal some differences across fields of education.

Two indicators are used as criteria to provide concurrent validity evidence:

- Results show a correlation between students' AHELO test scores and their self-reported academic performance only for the engineering strand. The strength of the relationship between AHELO scores and self-reported academic performance varies across countries.
- The strength of the relationship between AHELO scores and students' overall education satisfaction varies across countries.

### ***Reliability evidence***

The feasibility study produced instruments with “acceptable” to “good” levels of overall reliability. Examination of reliability indices at the country level shows less reliable results for some of them.

Reliability analyses using data aggregated at the institutional level suggest “acceptable” to “good” levels of reliability in all three strands. Examination of reliability indices using data aggregated at the institutional level indicates less reliable results for some countries.

Inter-scoring reliability statistics for constructed-response tasks can be considered “fair” to “good” in all three strands. Scoring of student responses may vary across countries but their rank ordering is very consistent. Scoring of student responses is consistent across countries when considering the tasks total scores.

The correspondence between the item difficulty levels and the students' ability levels for the generic skills strand indicates that the instrument is well targeted to the student population. The distribution of student performance shows that the economics and engineering tests were too difficult. The large proportion of “zero” scores for the economics and engineering constructed-response tasks also indicates that the items were too challenging for students.

Effort seems to have a greater impact on constructed-response tasks than on multiple-choice items.

## **Conclusions**

### ***Overall item quality and functioning***

The AHELO feasibility study produced many items that functioned well.

### ***Overall assessment of validity***

All three instruments have achieved reasonable levels of construct validity. The evidence collected also suggests that the instruments have achieved reasonable levels of content validity in the disciplinary strands and suggests that the instruments have achieved reasonable levels of face validity in all three strands. Evidence on concurrent validity is less conclusive.

**Overall assessment of reliability**

The three instruments provided reliable results. Inter-scorer reliability can be considered “fair” to “good” in all three strands.

**Overall scientific feasibility**

The AHELO feasibility study demonstrated that it is feasible to develop instruments with reliable and valid results across different countries, languages, cultures and institutional settings.

**Chapter 8 – National Experiences**

Seventeen countries/economies took part in the AHELO Feasibility Study. We have asked them to reflect on the experience. Their feedback is provided in Chapter 8 of the Feasibility Study Report, country by country. The first page of each country’s contribution to the Report is the poster which was prepared for the AHELO feasibility study Conference. These posters are reproduced below.

[Abu Dhabi](#)

[Australia](#)

[Belgium - Flanders](#)

[Canada \(Ontario\)](#)

[Colombia](#)

[Egypt](#)

[Finland](#)

[Italy](#)

[Japan](#)

[Korea](#)

[Kuwait](#)

[Mexico](#)

[Netherlands](#)

[Norway](#)

[Russian Federation](#)

[Slovak Republic](#)

[United States](#)

## Abu Dhabi



# ABU DHABI

Economics  
 Engineering  
 Generic Skills

Abu Dhabi's vision to become a knowledge-based economy depends largely on the quality of its graduating higher education students, and the AHELO feasibility study represents a catalyst in being able to assess their knowledge and skills using a reliable, multifaceted, and internationally valid measure.

### Main Challenges

- ✓ Short timelines: the development of a communication and recruitment strategy had to be implemented in two weeks, the mobilisation of all stakeholders in less than 1 month and the set up of the national infrastructure in less than two weeks.
- ✓ Provision of training on the technical standards and operations guidelines, and for documenting processes to be aligned with 2 years of work accomplished by other countries.
- ✓ Recruitment of all in-scope students for the engineering strand on a census basis.

### Main achievements

- ✓ Assessment Culture: Promoted the culture of comparative assessment at the institutional level and added a new dimension to the QA framework in Abu Dhabi.
- ✓ Learning Outcomes: measured the learning outcomes of Abu Dhabi students in comparison to their international peers and informed institutional decision makers and leaders of the capacity of their education systems.
- ✓ Capacity Building: developed the capacity of a highly skilled team to lead and implement similar national projects in the future and paved the way for adopting similar assessment exercises at the national level.

### Main Lessons

- ✓ Success in implementing the project activities smoothly due to the rigorous selection criteria for the ICs, Scorers and TAs.
- ✓ The high response rate emphasized the importance of having a well designed communication strategy to explain the expected outputs of AHELO.
- ✓ More time should be allowed for scorer recruitment, training and activities.
- ✓ Exchange of information and collaboration between countries (on scoring or risk management) illustrated that AHELO would benefit from a larger institutionalized process involving all countries.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

## Australia



# AUSTRALIA

Economics  
 Engineering  
 Generic Skills

Participation in AHELO has shown that Australia is well equipped to participate in this type of international study, with interest from stakeholders in participating, and well established systems for implementation.

Main Challenges

- ✓ Motivating students to participate.
- ✓ Securing a representative sample.
- ✓ Highlighting that the main outcomes from this study relate to the processes generated and not the data.

Main achievements

- ✓ Beginning a conversation about learning outcomes in higher education and offering a tool for assessing them.
- ✓ Co-ordination and co-operation of institutions in the implementation of a large online assessment.
- ✓ Stimulating students through innovative forms of assessment.

Main Lessons

- ✓ Long term planning is key to successful student engagement in such assessments.
- ✓ There are very motivated and generous people in Australian higher education.
- ✓ Providing ongoing information, data and outcomes summaries is important for maintaining interest and motivation within institutions.

**Key message:** Australia is proud to be at the forefront of the development and implementation of new ways of evaluating quality and improving learning in higher education.




[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



Australian institution co-ordinators at AHELO Symposium, October 2011

## Belgium - Flanders



## BELGIUM (FLANDERS)

Economics  
 Engineering  
 Generic Skills

The AHELO feasibility study showed us that an international test to assess higher education learning outcomes can be developed.

**Main Challenges**



- ✓ Students were not very eager to participate. Some institutions couldn't find enough students to participate and had to stop the project. For the test we had to use a census methodology.
- ✓ The study was sometimes considered as another evaluation on top of the existing ones.
- ✓ Convincing university colleges to participate, next to research universities.


**Main achievements**

- ✓ Knowing that it's possible to develop an instrument for the international comparison of achieved learning outcomes.
- ✓ Learning about the do's and don'ts of developing this kind of assessment.

**Main Lessons**

- ✓ To find out what resistance exists among students and teachers towards these types of evaluation.
- ✓ Open ended questions should be kept to a minimum.
- ✓ It is better not to take a sample, but rather to use census methodology.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



## Canada (Ontario)



## CANADA (ONTARIO)

Economics  
 Engineering  
 Generic Skills

Participating in AHELO significantly furthered our understanding of learning outcomes assessment, both through our own experiences and those of our international colleagues.

### Main Challenges

- ✓ Student recruitment was a significant challenge requiring considerable time and effort by the institutions.
- ✓ Institutional concerns about student-level data collection meant we were unable to capture accurate information on the characteristics of the student sample.
- ✓ The timing of the assessment window was challenging as it occurred during a period when students were on study break, and writing final reports and exams.

### Main achievements

- ✓ Considerable 'buy-in' from institutions, with 9 out of 10 eligible institutions volunteering to participate because they wanted to know how their institutions compare.
- ✓ Faculty members took ownership of the project and supported student recruitment because they wanted to know how their students were doing.
- ✓ Scorers found value in the assessment and said that the experience with AHELO will change their future teaching and assessment methods.

### Main Lessons

- ✓ Engagement of key members within the faculty was central to successful implementation.
- ✓ Weekly Institutional Co-ordinator conference calls were valuable in keeping the project on track and created a supportive community.
- ✓ It would have been beneficial to work with institutional administration much earlier in order to ensure that all aspects of implementation were as successful as possible.

**Key message:** Taking part in the international assessment was particularly valuable for faculty members engaged in the implementation and scoring processes. It provided them the opportunity to reflect on their curriculum design and delivery, and on the assessment techniques they employ. Most significantly, it made faculty members question their own methods and re-evaluate what they require of students.



The Department of Civil and Environmental Engineering

Contact | Site Search | Login

Search Civil and Environmental Engineering

OECD-AHELO Test
UNDERGRADUATE
GRADUATE STUD

Chair's Message  
 News  
 Research  
 About  
 Prospective Undergraduate Students  
 Prospective Graduate Students  
 Information for Current Students  
 Course Outlines  
 Forms  
 Faculty  
 Staff  
 Job Opportunities  
 John Adjelesian Lecture  
 Contact Us

Home / OECD-AHELO Test

### OECD-AHELO test

Last name \*

First name \*

Student number (Please enter the last 3 digits only) \*

Do you plan to write the OECD-AHELO test? \*

Yes, I am eligible to graduate next June and will write the test

No, I will not write the test because I am not eligible to graduate next June

No, I am eligible to graduate next June but will not write the test

Please check all the times during which you can write the test: \*

Monday March 12: 10:30-12:30

Monday March 12: 5:30-7:30

Tuesday March 13: 12:00-2:00

Tuesday March 13: 5:30-7:30

Wednesday March 14: 10:00-12:00



www.oecd.org/edu/ahelo

## Colombia



## COLOMBIA

Economics  
 Engineering  
 Generic Skills

Taking part in an AHELO feasibility study has been of great significance for all of Colombia's different Higher Education stakeholders and in particular for ICFES, the Colombian Institute for Educational Evaluation. Important challenges were met and valuable lessons were learned along the way. All in all it was a great opportunity to take part in a discussion at the highest level about the technical and practical requirements of a Higher Education assessment.

### Main Challenges

- ✓ Short timelines both to organize the application and to complete the marking of constructed response tasks.
- ✓ Finding an application scheme to ensure the possibility of studying the relationship between SABER PRO and AHELO. Given Colombia's unique position in which all of its higher education graduates take an end-of-degree test, participating in AHELO provided a unique opportunity to compare the results of both tests and enrich the AHELO data with information collected for the national test.

### Main achievements

- ✓ Selection of participating institutions and programmes was successful: 25 of the 26 programmes approached decided to take part in the assessment.
- ✓ Nearly 4 000 students assessed in one day in 26 application sites across 18 cities.
- ✓ High students' response rates: due mainly to the strategy to couple the application of AHELO with that of SABER PRO: the median for Generic Skills was 95%, with minimum 91%; and for Engineering the median was 98% with minimum 79%.

### Main Lessons

- ✓ Allow more time for Faculty responses.
- ✓ Devote more work to discuss and adapt test items and marking grids.

**Key message:** Linking the results from AHELO, SABER PRO (including the socio economic data) and SABER 11 (end of high school exam) will help evaluate the possibility of producing higher education value added measures. Besides pursuing its own analysis, ICFES intends to make these data public and to provide support to interested researchers.





[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

## Egypt



# EGYPT

Economics  
 Engineering  
 Generic Skills

In light of the inspiring 25th of January revolution, the Egyptian people have expressed their desire for more effective reform, as well as greater expectations for better quality of service in all aspects of life, particularly education. The new era of democracy and transparency is in harmony with concepts such as self-assessment and the developments that a ground breaking reform project like AHELO targets.

### Main Challenges

- ✓ The rapid and radical changes that involved the whole Egyptian community.
- ✓ The repeated changes in the leadership of higher education institutions and management boards that altered the implementation schedule for the project's activities.
- ✓ The large Egyptian universities (80 000-250 000 undergraduate students) and their incomplete electronic databases.

### Main achievements

- ✓ High response rates for students (total number 4 212) and faculty (total number 877), representing 18.3% (students) and 18.2% (faculty) of total AHELO participation.
- ✓ Increased awareness of the academic societies regarding the importance of linking the intended outcomes of programmes with the labour market.
- ✓ Success of the first concurrent online testing in the participating universities.

### Main Lessons

- ✓ National governmental commitment and support are cornerstones for assuring success of such large scale research studies.
- ✓ Recruitment of students for participation in future studies entails innovative strategies.
- ✓ Test simulations using released test instruments should be considered, for training purposes, as well as for exploring pitfalls and how they can possibly be avoided.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



GS AHELO team in Translation Adaptation Training

## Finland



# FINLAND

Economics  
 Engineering  
 Generic Skills

AHELO gives important information to faculties, institutions and governments on how to develop educational activities to further promote students' learning in the era of globalised higher education.

### Main Challenges

- ✓ internationally unstable financial situation
- ✓ low participation rate of Finnish students, which jeopardises the whole idea of AHELO in helping HEIs to develop their teaching and learning activities
- ✓ tight schedule in the implementation phase

### Main achievements

- ✓ high interest in AHELO among Finnish higher education institutions
- ✓ completion of instrument development and implementation in the given timeframe
- ✓ firm governmental support and competent national organisation

### Main Lessons

- ✓ International financing of the project has to be fully secured before it can start.
- ✓ The international consortium, including all of the partners, has to have a solid and consistent understanding of what kind of instruments to develop and how to carry out a large-scale international comparative project such as AHELO.
- ✓ The implementation phase must be given enough time; more time is needed to motivate students, to train ICs and to organise test sessions.

**Key message:** In order to avoid problems in the management and steering of AHELO, it is of crucial importance to secure full financing for the project based on a realistic budget. This would also help the participating countries to plan and finance their activities more precisely and ultimately yield a more coherent and manageable project.



The screenshot shows the Finnish version of the AHELO website. The header includes the title 'Koulutuksen Tutkimuslaitos' and the AHELO logo. The main content area features a section for 'AHELO-FS hanke' (AHELO-FS project) with a description in Finnish. The website also displays logos for the OECD, the AHELO Feasibility Study, and the Association of European Universities (AEU).



## Italy



# ITALY

Economics  
 Engineering  
 Generic Skills

The implementation of the AHELO feasibility study in Italy was a positive and successful experience which has shown that universities want to be assessed and perceive exercises like AHELO as an opportunity and not as a threat.

Main Challenges

- ✓ Highly time consuming activities for all the individuals involved in the project at international and local level.
- ✓ Continuous necessity to adapt procedures during the implementation of the project due to the experimental nature of a feasibility study.
- ✓ A general lack of familiarity in the Italian culture to have students take standardised tests.

Main achievements

- ✓ Introduction of the first experience of learning outcomes assessment in the Italian higher education system.
- ✓ Successful implementation and local management of the entire exercise, proven by the high number of universities that applied and students involved.
- ✓ The implementation of a follow up at national level on the basis of the methodology tested with AHELO.

Main Lessons

- ✓ The IT assistance during the test administration should be managed at the local level in order to allow real time troubleshooting.
- ✓ This experience has shown that Italian universities want to be assessed and perceive exercises like AHELO as an opportunity and not as a threat.
- ✓ The translation of academic content from English to Italian is a delicate and challenging task.



ahelo  
Feasibility Study

[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

©Luigi Rosa/Tickr

## Japan



# JAPAN

Economics  
 Engineering  
 Generic Skills

For Japan, the AHELO feasibility study represented an exciting engagement in an international conversation on what engineering graduates are expected to know and be able to do in a knowledge based global society.

Main Challenges	Main achievements	Main Lessons
<ul style="list-style-type: none"> <li>✓ Involving faculties: taking faculty time away from teaching and research requires a very good reason, as well as a clear description of what kind of feedback they will be receiving.</li> <li>✓ Sustaining momentum: the prolonged planning period made it difficult to keep the higher education community interested and engaged.</li> <li>✓ Translating instruments according to protocols: achieving substantive equivalence requires some flexibility and extensive knowledge of the language and the subject matter.</li> </ul>	<ul style="list-style-type: none"> <li>✓ A tangible and substantive understanding of a conceptual framework of engineering competencies and learning outcomes that can be shared globally.</li> <li>✓ Concrete and innovative ideas for conceptualising and measuring competencies and learning outcomes.</li> <li>✓ A delightful experience working on an international team, learning from global partners, and being able to make unique contributions.</li> </ul>	<ul style="list-style-type: none"> <li>✓ An international assessment of higher education learning outcomes can become a useful tool for educators to globally benchmark and update their teaching practices.</li> <li>✓ Designing constructive response tasks to "measure" how students can "think" like an engineer requires a thoughtful balance between open-endedness and preciseness.</li> <li>✓ The exercise of scoring and modifying scoring rubrics by an international team of experts is extremely important to reach consensus on the scope and level of expected learning outcomes.</li> </ul>

**Key message:** AHELO can become a powerful tool for educational improvement, when instruments and scoring rubrics are made fully available to participating institutions, and when coupled with workshops that induce discussion about curriculum design and encourage innovation in teaching and learning.





[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

The Japanese Scoring Team (June 2012)

## Korea



# KOREA

Economics  
 Engineering  
 Generic Skills

The AHELO feasibility study represents a journey towards excellence in higher education.

Main Challenges	Main achievements	Main Lessons
<ul style="list-style-type: none"> <li>✓ Securing sufficient budget for AHELO at the national and institutional level.</li> <li>✓ Recruiting randomly sampled students across over 50 different departments at each HEIs!!</li> <li>✓ Ensuring quality in scoring student answers: the scoring work was very important for the success of test implementation</li> </ul>	<ul style="list-style-type: none"> <li>✓ Randomly sampled representative data that include students and faculty from over 50 different academic fields with little governmental support.</li> <li>✓ Established cooperative network among KMOE, experts of higher education, and members of Higher Education Institutions.</li> <li>✓ Drew interest and concerns of stakeholders on the ways in which that the current Korean university students are educated including the issues of goals, curriculum, and pedagogy.</li> </ul>	<ul style="list-style-type: none"> <li>✓ We have witnessed mounting demand for more relevant data on higher education learning outcomes, particularly in relation to Generic Skills.</li> <li>✓ We should consult with experts throughout the whole process of implementation and open communication opportunities with involved project members many as possible.</li> <li>✓ Rather than other forms of incentives, authority is one of the best motivators for participants to make a commitment to the AHELO project</li> </ul>

**Key message:** "Details, details, details": In order to bring about a successful "Further Studies", and serve the interests of stakeholders in the future, thorough and more elaborative schemes should be developed which include: objectives, implementation strategies, expected outcomes as well as its relevance to higher education.







[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



## Kuwait



# KUWAIT

Economics  
 Engineering  
 Generic Skills

For the State of Kuwait, the AHELO feasibility study represented an international team effort at enhancing national and institutional accountability and responsibility toward student learning and success.

### Main Challenges

- ✓ Translation and cultural adaptation of the Performance Task was difficult because the instrument chosen was based on a US tool which was not quite right to assess non-American modes of thinking or writing.
- ✓ The Kuwait National Team understood the difficulty associated with soliciting student participation on a national scale at the level of taking the assessment and with regards to the level of effort applied by the student being high enough that the instrument can be used as a measurement of degrees of learning accomplished by the student.
- ✓ Performance Task on-line platform: students were confused as to when the assessment had ended (when finishing a section for example). Thus, the assessment platform requires further development.

### Main achievements

Initiation of a more focused and expanded national and institutional conversation on the importance and benefits associated with:

- ✓ standardising assessment measures
- ✓ internationalising expectations with respect to student learning;
- ✓ internationalising benchmarks against which the quality of institutions, quality instruction, and quality learning can be measured.

### Main Lessons

- ✓ Need for a more comprehensive, unified and a more academically informed plan with respect to student incentive and participation, as well as to find ways to encourage "active" student engagement during testing (mandatory assessment, assessment for grades/credit).
- ✓ Need to internationalize the initial development of the Performance Tasks that more accurately reflect the linguistic nuances, cultural sensibilities and sensitivities, as well as student learning outcomes measures.
- ✓ National culture of assessment: need to mainstream standardised assessment on all institutional levels and encourage a "national culture of assessment."



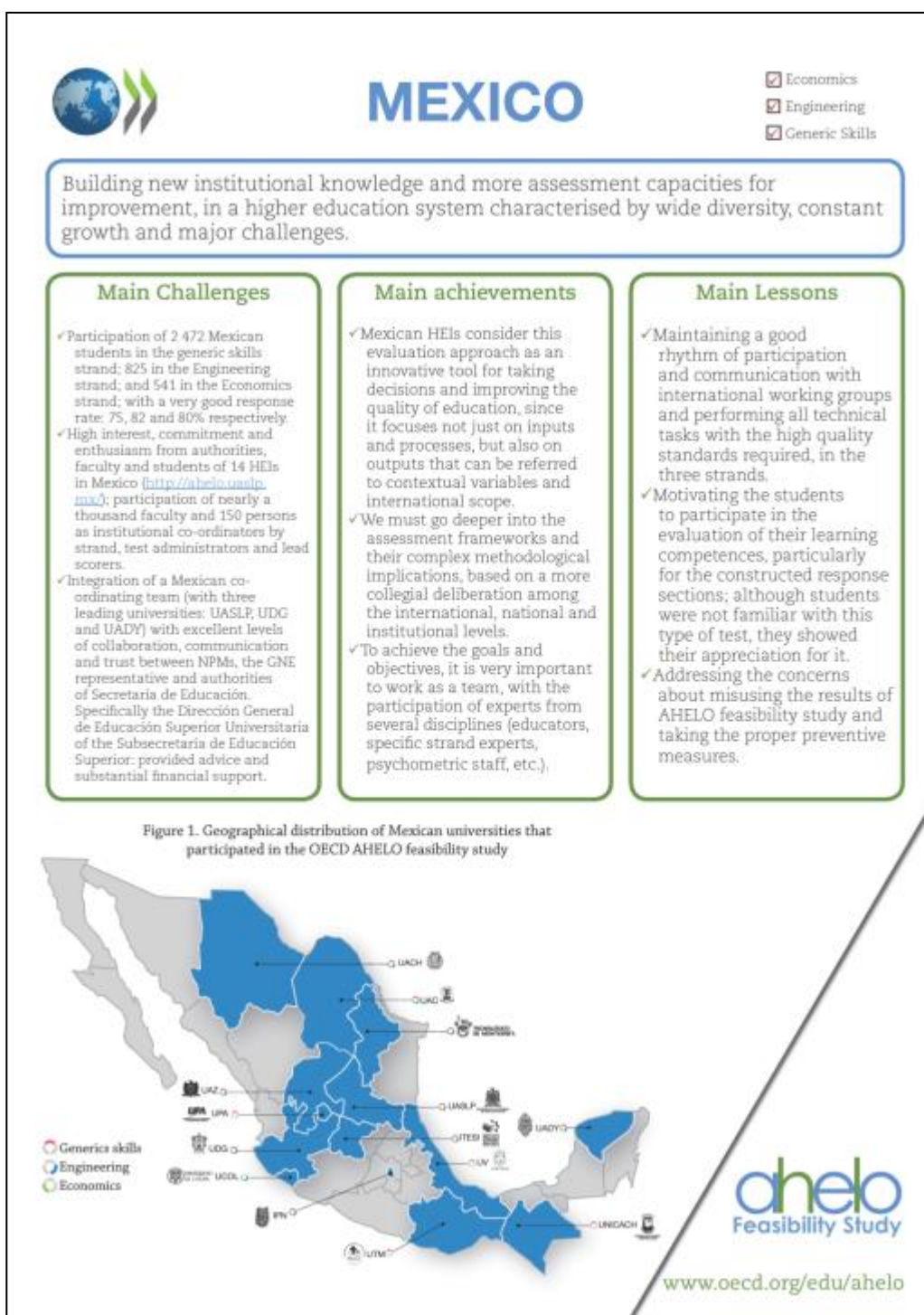
[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



Student taking AHELO test at Kuwait University



## Mexico



## Netherlands



## NETHERLANDS

- Economics
- Engineering
- Generic Skills

The AHELO feasibility study in the Economics Strand proved to be an interesting addition to the ways in which the quality of education is measured in the Netherlands, but challenging to organise on top of all other quality measures already being implemented (none of which allow for an international comparison though).

## Main Challenges

- ✓ Getting HEIs and students to participate without incentive, despite communication efforts and faculty involvement.
- ✓ Realising an international assessment in a very short timeline due to the uncertainty of the project and in a period when Faculty was very busy with other activities.
- ✓ Adapting the assessment of the economics strand to the Dutch economic curricula and the Dutch binary system.

## Main achievements

- ✓ Both types of Higher Education in the Netherlands represented in the field work (Research Universities and Universities of Applied Sciences).
- ✓ Despite the very short timeline a successful central management and implementation plan was set up.


## Main Lessons

- ✓ The items in the economics strand are more knowledge-based than would be expected and they did not cover the whole range of economic studies provided in the initial economics framework.
- ✓ In binary systems the development of assessment instruments should cater for the needs of different types of economics studies, in such a way that both research-oriented as well as applied-oriented types of education clearly recognise their own academic content and didactical approach.
- ✓ Incorporate the instruments more into the quality assurance measures already being implemented in Higher Education on a national and institutional level.
- ✓ Incorporate the main study instruments more into the academic programmes students participate in, or make sure that there is a clear (and equal) incentive offered to all students participating.
- ✓ The timing of the assessment should be considered.
- ✓ Make sure that in all parts of the academic community involved (i.e. Ministry, national organisations, institutions, faculties, teachers and students) there is enough support for an AHELO and carefully position NC and NPM in this field.

ahelo  
Feasibility Study

[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

## Norway



# NORWAY

Economics  
 Engineering  
 Generic Skills

The study offered a chance to learn more about the outcomes of higher education and the potential use of the generic skills tests. It showed us that engaging students and staff is a central challenge in such work.

**Main Challenges**


- ✓ Response rate: the most significant challenge by far was recruiting enough students, and the response rate was lower than hoped for.
- ✓ High demands on institutional resources: the complex scoring, substantial work to recruit for and run the survey and the need for large student incentives (prizes/giftcards) raised costs.
- ✓ Timing: the test period was in a semester when many students spend less time on campus and instead work on projects or a bachelor's thesis.

**Main achievements**

- ✓ Good co-operation within the Norwegian team: the planning, technical preparation and collaboration between the institutions and national team worked well.
- ✓ It showed that it is feasible to deliver electronic tests globally: aside from a handful of minor technical problems, the testing process ran smoothly.
- ✓ Trying out a variety of approaches to promotion and incentives: institutions tried a range of approaches to inform and attract students to this pilot.

**Main Lessons**

- ✓ The processes and collaboration required for such tests are in place, but the balance of costs and benefits for students and institutions needs to be considered carefully in the future.
- ✓ The institutions are interested in such tests, but the time and resources demanded in this case were high, especially considering the limited information offered by results (due to response rate).
- ✓ There is no obvious solution as to how students can be encouraged to take such an extensive test – a more targeted sampling approach may prove more practical and effective in Norway.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



## Russian Federation



## RUSSIAN FEDERATION

Economics  
 Engineering  
 Generic Skills

In the Russian Federation, the AHELO feasibility study is a large-scale bottom-up project, initiated and coordinated by universities (Higher School of Economics in cooperation with the Ural Federal University in Engineering) and accomplished with great support and a high level of participation of Russian universities, serving as a basis for the further exerts networking and cooperation in QA.

### Main Challenges

- ✓The nature of tasks for testing: theoretical tasks prevailed, the limited bank of tasks and a high risk of cheating.
- ✓The legal and practical difficulties of collecting student and faculty personal data for sampling.
- ✓The differences of curriculum between HEIs and difficulties for further comparability.

### Main achievements

- ✓Large scale of university participation, involvement of regional universities in this international research. High level of support from students, faculties, universities administration.
- ✓High response rate among students.
- ✓Elaboration of supplementary national instruments (online questionnaires for students and faculties) to get additional data about LO context.

### Main Lessons

- ✓Independent national supervisors are needed to assure compliance with the testing procedures and minimize the risk of cheating.
- ✓The bank of tasks needs to be larger and more diversified, it would be better to link the tasks with the real professional activities.
- ✓The HEIs are able to organize the assessment process following the international guidelines but national project management system is needed to assure their coordination and communication.

**Key message:** The AHELO project is in line with the national strategic objectives in the QA area relating to the creation of the independent quality assessment system in higher education and encourages such current national initiatives as a Federal exam for bachelors (since 2012). The AHELO feasibility study was the opportunity to have an experience of measuring higher education competencies on the national and international level, the possibility of comparison of educational quality level in different HEIs and possibility for federal and regional universities to participate in international research on the same level and the chance to bring students learning outcomes of different universities to the one scale.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)

© OECD 2013

## Slovak Republic



## SLOVAK REPUBLIC

Economics  
 Engineering  
 Generic Skills

For Slovakia, the AHELO feasibility study represented a light at the end of the tunnel in the endless debates on the quality of Slovak higher education.

**Main Challenges**

- ✓ Finding ways to motivate HEIs and students to participate in the AHELO project.
- ✓ Managing, in a short time and without any previous experience of this type, a project of such size and complexity in the HE environment.
- ✓ Taking the risk of a relatively high financial investment in a project for which the development and results were uncertain at the moment of making the decision.

**Main achievements**

- ✓ High participation of HEIs and high, in some cases even very high, participation of students in testing.
- ✓ Obtaining new information on the Slovak HEIs based on methodology guaranteed by and international consortium.
- ✓ Successful organisational and technical management, including good decisions regarding the technical arrangements (mobile computer labs), leading to smooth implementation.

**Main Lessons**

- ✓ It is possible to motivate the HEIs and students to take part in innovative activities and to engage them with the offer of obtaining information on their results in order to compare them with those of their peers.
- ✓ Success in a project of this kind requires the synergy of several factors: enthusiasm and professionalism of people; adequate funding; willingness to take risks; and the support of the Centre.
- ✓ In principle it is feasible to get external and internal comparisons of HEIs of different countries.

**Key message:** For a small country like Slovakia, the use of mobile computer labs operated by stable teams of well prepared experts considerably simplified the implementation of testing, decreased its dependence on local conditions, eliminated potential negative local influences and increased its technical reliability and smooth execution.



[www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo)



engineering test session at the Faculty of Civil Engineering, Technical University Kosice (April 24, 2013)

## United States





**Conclusions from Chapter 9 (by Peter T. Ewell, Chair of the TAG)****The TAG's overall assessment of the feasibility study**

The Technical Advisory Group (TAG) believes that the AHELO feasibility study constituted an unprecedented multi-national data collection effort at the higher education level. Data on student learning outcomes have been collected in three domain strands in 17 different countries or systems, using assessment instruments comprising both production-focused CRTs and forced-choice MCQs. Data have also been collected on a wide range of contextual factors by means of surveys completed by students, faculty members, ICs and NPMs.

Numerous implementation challenges including translation, contextualisation, sampling, electronic test administration, CRT response scoring, data cleaning, statistical analysis, and reporting have been met and successfully overcome. To be sure, some countries/systems experienced more difficulty than others and, because of this, levels of success varied.

Nevertheless, all participating countries reported they learned something from the experience and most would do it again. Just as important, the feasibility study generated a range of important findings about student learning at the higher education level, as well as dozens of lessons about how such a project should be implemented in the future.

That said, the TAG wishes to briefly point out a few things that went particularly well in the AHELO feasibility study and a few that did not go so well. Several of these have been touched upon in earlier sections of the report and most have implied lessons for any AHELO Main Study.

***What went well***

The TAG believes that the following were particular strengths of the feasibility study:

***Assessment administration***

Electronic administration of assessment on a global scale, and in multiple languages and jurisdictions, confronted the feasibility study with an enormous challenge. This challenge was met admirably. Only one significant failure in administration occurred over scores of testing sessions at hundreds of institutions. The technical infrastructure underlying this achievement, the thorough training regimens put in place for Institutional Co-ordinators, and the robust administration procedures established were all praiseworthy.

***Technical aspects of the data analysis***

The data yield of the feasibility study was large and complex, resulting from the administration of six different instruments to many different kinds of respondents. In the face of this, the Consortium's efforts to provide sound analyses were exemplary from a technical standpoint. The analysis plans were sound, the statistical techniques employed were proper and well executed, and appropriate and effective "work-arounds" were put into place when analytical problems (such as missing data or malfunctioning items) were encountered.

### *Instrument design for purpose-built instruments*

All of the instruments designed especially for the feasibility study were of exemplary technical quality including the MCQs and CRTs for Engineering and Economics and the three surveys comprising the Contextual Dimension. All were developed through reference to adequate and helpful Assessment Frameworks and were informed by knowledgeable expert groups (in the cases of Engineering and Economics) or considerable background work (in the case of the Contextual Dimension). Moreover, these instruments were produced quickly with little re-work, were designed to a high technical standard, and were piloted as well as could be expected in the short timelines available.

### *Overall co-ordination*

Management and co-ordination of an enterprise as complex as the AHELO feasibility study involved massive challenges of maintaining consistent procedures across five continents, 17 unique cultural-political contexts, and numerous time zones. The administrative arrangements established by the Consortium met these challenges with clear direction and minimum confusion. Where the inevitable problems were encountered, they were for the most part resolved quickly and smoothly.

### ***Things that did not go so well***

At the same time, the TAG believes that some aspects of the feasibility study did not go so well. As a consequence and as reflected in the TAG's recommendations for any AHELO Main Study, they constitute areas that must be particularly examined as the initiative moves forward.

### *Resources and time*

As the TAG pointed out repeatedly in the course of the feasibility study, the AHELO feasibility study was seriously under-resourced and was implemented on far too short a timeline. More resources and time could have enabled such important features as more cognitive interviews and pilots of newly-built instruments, full-scale field trials of administration and scoring arrangements, and more time for de-briefing and collective discussion of obtained results.

### *CRT difficulty and contextualisation*

While the CRTs used by the Engineering and Economics assessments were of high technical quality, they were simply too difficult for many students to effectively engage and perform well. At the same time, the CRTs used in Generic Skills based on the CLA proved excessively "American" in an international context. As above, more time for piloting and field trials might have revealed both of these situations at an earlier stage — in time for it to be rectified.

### *Reporting results*

While the TAG believes that the Consortium's analyses of the massive amount of data generated by the feasibility study were exemplary from a technical standpoint, the reporting of these results through the Consortium's final report was overly complex, and therefore difficult to understand. Most important, the report lacked clearly stated conclusions on which to make policy decisions for the future. Again, this was probably partly a result of time pressures, and



the reporting process would have benefitted from reflection and feedback from stakeholders after results were made available. Again, the March 2013 conference should prove useful in this respect.

#### *Contractual arrangements*

The AHELO feasibility study began with separate contracts between the OECD Secretariat and the two principal contractors — ACER and CAE. These independent contractual relationships resulted in poor communication among the contractors and occasional duplication of effort. Furthermore, no tendering process was used to procure or develop instruments for the Generic Skills strand — a fact that is highly unusual in international studies of this kind. By the time this situation was addressed by re-structuring contractual arrangements so that CAE was a subcontractor of ACER under the Consortium, a habit of independence — exacerbated by commercial rivalry—made it difficult for both parties to establish a culture of partnership.

#### ***Some additional lessons***

Finally, the TAG believes that the AHELO feasibility study offers several additional lessons that should be taken forward for any international assessment effort of this size and scale:

- ***There should be more opportunities for stakeholder participation in assessment design and in the analysis of assessment results.*** There were many points in the feasibility study at which the wisdom of practitioners and the national and institutional levels could have been better collected and used for improvement. While the many efforts to contextualise instruments and administration procedures were admirable and, for the most part, successful, a more collaborative approach might have yielded greater benefits.
- ***A full-scale try-out of all instruments and administration arrangements could enable stakeholder participation in a “design-build” process that would both pilot these designs and enable more stakeholder engagement in making them better.*** This is especially the case for reporting results and sharing data with countries and institutions. Many NPMs and ICs remain somewhat disappointed by the lack of attention to their needs for information resulting from the study — especially the provision of country-level data files that lacked the documentation needed for analysis.
- ***Any such study should be better located and integrated with the international scholarly community examining student learning outcomes and the policies and practices that support better learning.*** As pointed out in the rationale for AHELO, the past decade has seen a sharp increase in policy and scholarly interest in improve academic performance in higher education. Evidence of this can be seen in the Bologna Process and Tuning in Europe, the Spellings Commission and interest in accreditation in the United States, the rise of qualifications frameworks in many nations, and the emergence of multinational mapping and ranking initiatives like U-map and U-Multirank. AHELO represents an opportunity to better align the emerging scholarly and policy dialogue about quality.

- **All of this will require more time and adequate resources.** The TAG's conclusion in this regard remains unaltered: if the required resources and timelines needed are not forthcoming, a future study of this kind should not be undertaken.

On balance, the TAG believes firmly that the AHELO feasibility study was soundly executed and provided many lessons that will continue to inform international assessment efforts for many years to come. Among its most important contributions to the study were recommendations to ensure consistency of administration and scoring across contexts, steady reinforcement of the need for contextual data — especially at the beginning of the study, recommendations to reinstate an MCQ component in Generic Skills, and recommendations to the OECD Secretariat about how to prepare its final report. Members of the TAG all learned something important through their engagement in the study and congratulate the Consortium and the OECD Secretariat for a job well done.