

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Ursin, Jani

Title: Finnish Experiences Of OECD's International Assessment Of Higher Education Learning Outcomes (AHELO)

Year: 2015

Version:

Please cite the original version:

Ursin, J. (2015). Finnish Experiences Of OECD's International Assessment Of Higher Education Learning Outcomes (AHELO). Hungarian Educational Research Journal, 5(3). <https://doi.org/10.14413/herj2015.03.03>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

FINNISH EXPERIENCES OF OECD'S INTERNATIONAL ASSESSMENT OF HIGHER EDUCATION LEARNING OUTCOMES (AHELO)

Jani Ursin, PhD
University of Jyväskylä

Abstract

The aim of this article is to describe the implementation of International Assessment of Higher Education Learning Outcomes (AHELO) in Finland and to highlight ethical considerations of large-scale international assessments. The Finnish results of the AHELO feasibility study show that a fully-fledged project is possible to carry if special attention is paid to the participation of students in test sessions, if enough time is reserved for the implementation phase and if scoring of open-ended questions is carefully carried out and monitored. It is, however, important that large-scale international assessments such as AHELO do not start to colonise and converge understandings of what is considered a desirable end result, promote conceptions of 'good' teaching and learning in higher education and hence become (un)intentionally a powerful political tool to steer and justify national educational reforms.

Introduction

Over the past two decades Organization for Economic Co-operation and Development (OECD) has become influential player in moulding education policies of nation states (Rinne 2006). This is owing to the large-scale international assessments of learning outcomes that it carries out on a regular basis in primary (Programme for International Students Assessment, PISA) tertiary (International Assessment of Higher Education Learning Outcomes, AHELO) and adult education (Programme for the International Assessment of Adult Competencies, PIAAC). These assessments have become globally important instruments to measure the competences of citizens of different ages. A need for the international assessments was established already in the 1990s when assessments were seen important tools for steering the education and social policies of nation states (Whitty 2010). Nowadays, many of the OECD's assessments have started to shape national decision making in the form of 'soft laws' which often have (in)direct influence on the national reforms of education systems (Kallo 2009).

A starting point for AHELO – which is the latest comer in OECD's assessments of learning outcomes – lies in the global trends, such as diversification of institutional profiles and student body (Teichler 2006), greater internationalization (Kehm & Teichler 2007) as well as growing emphasis of market forces in higher education (Exworthy & Halford 1999), which have profoundly changed national higher education systems. Traditionally the success of higher education institutions (HEIs) in the global educational markets has been measured by various league tables, such as Times Higher Education World University Rankings and Academic Ranking of World Universities by Shanghai Jiao Tong University, which are often biased towards inputs and research mission of universities (Hazelkorn 2011). AHELO was

produced as a counter effect to these input-based league tables in order to focus more on outputs of teaching and learning of HEIs without an intention to create any new ranking.

AHELO was conducted in 2010–2013. The aim of the AHELO feasibility study was to develop instruments – across different countries, languages, cultures and institution types – to measure tertiary level students’ competences, i.e. what they know and can do at the end of their undergraduate studies. The feasibility study investigated whether reliable cross-national assessments of HE learning outcomes are scientifically possible, and whether their implementation is feasible. Hence, the goals were both scientific and practical. Assessment instruments were developed for generic skills, economics, and civil engineering. Altogether 249 higher education institutions across 17 countries and regions participated in the study with about 23,000 students being tested. In Finland twelve HEIs participated in AHELO and 331 students took the test.

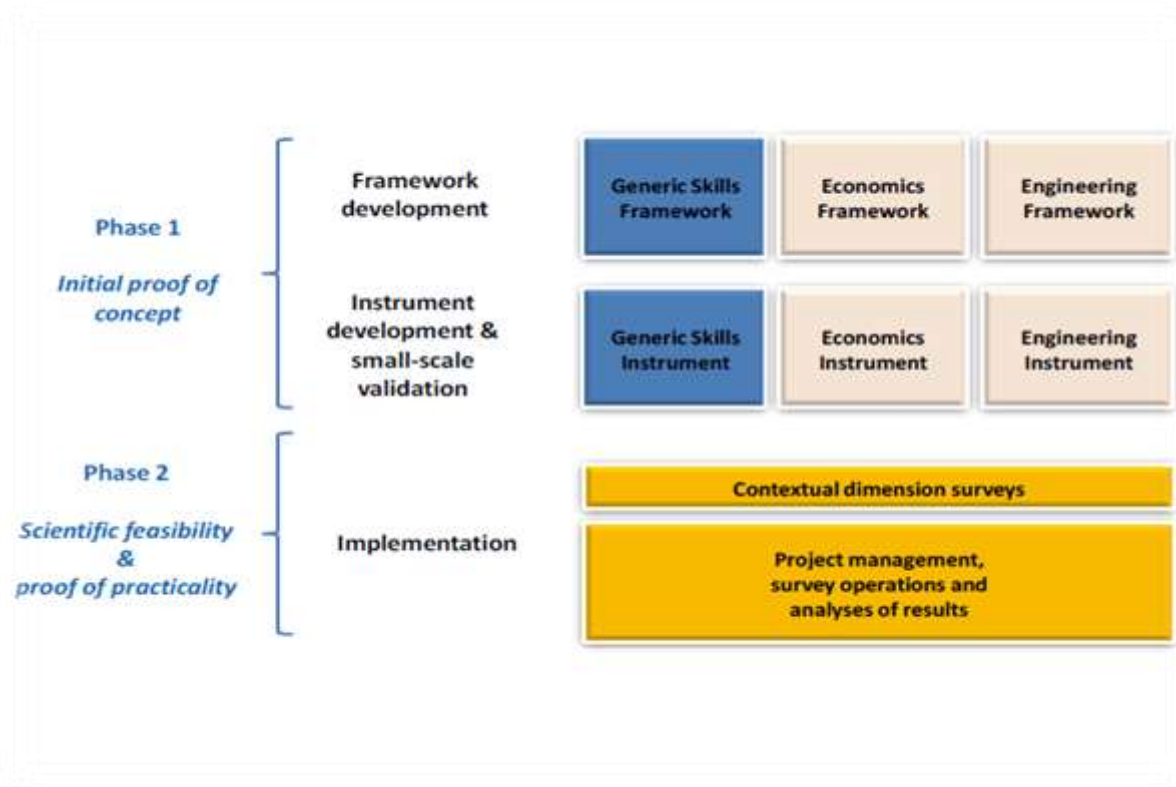
The aim of this article is to describe rationale, design and implementation of AHELO, discuss about the lessons learnt in Finland as well as provide some critical observations of the role of AHELO in developing national higher education systems. The article focuses mainly on Generic Skills strand in which Finland participated – and especially on the development of constructed response task (CLA instrument) of the generic skills.

Rationale, design and implementation of AHELO¹

AHELO was carried out in two phases (Figure 1). The first phase was about developing conceptual frameworks and instruments for all three strands of work and thus gain initial proof of concept. The second phase was the implementation through which scientific feasibility and proof of practicality were explored. The project also had three strands of work two of which were disciplinary (economics and engineering) and one about generic skills (critical thinking, analytic reasoning, problem solving and argumentative writing). Furthermore, each participating higher education institution, faculty member and undergraduate student filled in a background survey in order to help contextualize the results.

Figure 1. Design of AHELO feasibility study

¹ This chapter is based on the first volume of AHELO reports Tremblay, K., Lalacette, D. & Roseveare, D. 2012. *AHELO, Assessment of higher education learning outcomes. Feasibility study report, vol. 1. Design and implementation.* OECD. as well as Ursin, J. 2014. *AHELO – Korkeakouluopiskelijoiden oppimistulosten kansainvälinen arviointi.* [AHELO – International assessment of higher education students’ learning outcomes] *Kieli, koulutusjyhteiskunta.* Huhtikuu/2014. Verkkajulkaisu.



The idea of feasibility study was to include a variety of countries in terms of geographic origin, languages and cultures in order to ensure sufficient international variation in each strand. Altogether 17 countries or regions participated in AHELO which indeed presented a balanced picture of geographic, linguistic and cultural backgrounds (Table 1). In Finland 22 HEIs (out of 42) volunteered for AHELO of which six universities and seven universities of applied sciences (of which one later dropped out) were selected by the national steering group to participate in the project.

Economics	Engineering	Generic Skills	Contextual surveys
<ul style="list-style-type: none"> • Netherlands • Belgium (fl.) • Egypt • Italy • Mexico • SlovakRep. • Russia 	<ul style="list-style-type: none"> • Australia • Egypt • Japan • Canada (Ontario) • Mexico • SlovakRep. • Abu Dhabi 	<ul style="list-style-type: none"> • Egypt • Columbia • Korea • Kuwait • Mexico • Norway • SlovakRep. • Finland • USA (CT, MO & PA) 	All countries

Table 1. Participating countries in AHELO by strand of work

In the next chapters, I will concretize the framework and instrument development, validation and implementation from the point of view of Generic Skills strand as this was the strand that Finland participated. In generic skills – like in the disciplinary strands – two different types of instruments were developed: constructed-response tasks and multiple choice questions. The constructed-response task in generic skills was based on already existing instrument namely Collegiate Learning Assessment (CLA) developed by the Council for Aid to Education (CAE) from the USA. In AHELO, CLA measured three high-order cognitive skills: analytic reasoning and evaluation, problem solving and writing effectiveness. CLA is based on performance tasks which imitates real life problem solving situations. Two performance tasks were selected and adapted by participating countries in co-operation with CAE to ensure their cross-cultural appropriateness. The tasks were labelled as ‘Catfish’ and ‘Lake to River’.² Later on multiple-choice questions were added drawing from existing items to measure generic skills developed by the Australian Council for Educational Research (ACER). As both of the instruments were already pre-existing an assessment framework for generic skills was developed afterwards and not before the development of instruments as in economics and engineering which created a situation where not everybody was content with the framework.

Next the CLA instrument was translated and adapted for small-scale validation. The translation and adaptation followed an agreed localization process. A dual translation model was used with two translators working independently to provide a full translation from English to Finnish after which translations were reconciled by the third translator. Translation in Finland was a smooth process – there was some discussion about localization of names and places into Finnish context and about the level of formality of language used (in Finnish it is not so typical to use titles, for example). The final version was then verified by the national team and international consortium. In Generic Skills strand the validation of translated and adapted instruments included cognitive laboratory procedures and ‘think aloud’ interviews with student respondents. For the purpose of this validation in Finland twelve students were invited to take the test and while taking the test ‘think aloud’ that is to explain how they had constructed their answers and whether they had any difficulties in understanding the questions or task. This verbal probing method allowed identifying possible cross-cultural appropriateness issues which in Finnish case were very minor ones. Consequently, ‘think aloud’ method verified that the thinking elicited by the performance task was the thinking sought. Before the final testing the validated instruments were transferred to online platform and its functionality was tested.

Invigilated and online computer delivered test sessions were carried out in Finland in spring 2012. In each participating HEI 200 students at the end of their undergraduate degree were randomly sampled to take the test (N = 2400). The test lasted the maximum of 150 minutes where the first 90 minutes were reserved for CLA instrument and thereafter a student had 30 minutes for answering to 25 multiple-choice questions. At the end of the test student also filled in short survey collecting background information of a student such as gender, age and educational background of the parents. Participating HEIs were responsible of organizing the

²The ‘Catfish’ performance task is presented in the annex B of the first volume of AHELO reports, see <http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf>

test sessions for students. Given the tight time frame test sessions in Finland went well in all twelve participating institutes. There were only some minor issues, such as freezing of the computer during the test session, which were easily solved. However, the biggest challenge in Finland was to get students to take part in the test sessions. At worst less than ten students (out of 200) participated in the test and even in the best case a little more than 60 students took the test. This was the case even when all the participating HEIs arranged several test sessions at different times and days and also offered external incentives such as cell phone lotteries, free lunches and movie tickets. Overall the participation rate in Finland was 14 percentages which was one of the lowest among all participating countries. The main reasons for poor participation in Finland were timing of the testing which happened at the end of the spring term when students had already left the campus and lack of internal incentives like ECTS points.

In the test sessions a student had to fill in one of the performance tasks which computer had randomly assigned to a student. In order to answer a set of questions a student had to familiarize him/herself with materials in the online 'material bank'. These materials included, for example, scientific articles, email correspondences, radio interviews, graphs and newspaper articles. The materials contained relevant and irrelevant, reliable and unreliable and even deceptive information. A student had, for instance, to understand a difference between correlation and causality. Student had to base his/her answers – which length was not delimited other than 90 minutes time limit – on materials provided. Feedback received from Finnish students who took the test highlighted the challenging nature of the test as well as the pressures caused by the time limit. Nonetheless, many of those who gave feedback from the test also considered it to be interesting and useful exercise.

Once all the test sessions were completed the constructed response tasks were scored by trained scorers. Each performance task was double scored and the whole process was monitored by a Lead Scorer who also audited every fifth or so response. In Finland there were four scorers and a Lead Scorer. Each scorer gave marks from 0–6 to each assessment criteria (analytic reasoning and evaluation, problem solving and writing effectiveness) thus the minimum score being 0 and maximum 18 per constructed response task. The criteria for scoring were presented in a scoring rubric. The scoring of analytic reasoning and evaluation was based on, for example, how student can identify strengths and weaknesses of alternative arguments and how to distinguish reliable sources from unreliable. In scoring of problem solving attention was paid on how a student had utilized the documents in forming a decision. The main criterion for writing effectiveness was in the argumentativeness of writing; how logically and clearly the answer was written. There were few Finnish cases in which the two scorers' grades were different however the lead scorer was able to solve those cases. Number of scorers (four scorers and a Lead Scorer) also proved to be too few so in the future more scorers should be recruited.

Main lessons learnt in AHELO

The aims of the AHELO feasibility study were whether reliable cross-national assessments of higher education learning outcomes are scientifically possible, and whether their

implementation is feasible. All in all, AHELO was both scientifically and practically feasible; however, there are several issues that need to be solved before a full-scale AHELO can be implemented. The main scientific challenges relate to the reliability of test instruments and especially to the discussion on reliability and use of multiple-choice questions versus constructed-response tasks. Another important scientific issue is to find an international agreement of assessment frameworks which was not found in the feasibility study in terms of Generic Skills strand. Generic skills also raised a lot of discussion whether they should be measured as separate strand or as embedded in the disciplinary strands. Furthermore, participation rates need to be good enough in all participating countries in order to make reliable analyses and enough time should be reserved for student testing. However, these scientific challenges can be surpassed with careful planning.

The practical issues may be a bit more challenging to overcome. What is the value of AHELO for students, higher education institutions and decision makers? The students who participated in AHELO did not receive their own test results – which was agreed to be out of the scope of feasibility study. However, it is of crucial importance that students will get their own results if a fully-fledged AHELO will take place – as technically this can be easily done. It is also a way to motivate students to take the test – an issue which is important in countries like Finland. More complicated is to show the benefits of AHELO to HEIs and national decision makers, that is to the Ministry responsible for higher education. The unit of analysis in AHELO feasibility study was HEI which received a compact feedback report of the success in AHELO. However, based on the feedback received from HEIs they were unsure how to relate the results to, for example, their quality assurance practices and procedures as well as to internal development activities. Perhaps the results can be used for marketing purposes of undergraduate programmes but this would then change the original idea of AHELO of being more about developing than ranking institutions. How about the decision makers – what do they get from AHELO? Unlike in PISA or PIAAC in AHELO it is not justified to publish the results by country simply because the results cannot be reliably generalized to country level; if also in the future in AHELO only ten institutions per country are purposefully selected what do the results tell for example about the United States which has more than 7000 HEIs of different types and about thirteen million students? All in all, the benefits and value for money of AHELO for HEIs and nation states need to be sharpened. Another important aspect is to recognize ethical responsibilities of AHELO, an issue that I will tackle closer in the next chapter.

Ethical considerations of AHELO³

Although the benefits, such as benchmark information, of international large-scale assessments such as AHELO are important and indisputable, more attention should be paid to the (un)intentional consequences they produce. First, the fact that these assessments often colonialise and converge understandings of what is considered a desirable end result and promote conceptions of ‘good’ teaching and learning as well as students’ supposed role in the

³ This chapter is based on Ursin, J. International Comparative Assessments as Policy Devices – A Critical Perspective. Submitted to *Educatio* (R).

learning process will inevitably have an enduring effect on what is valued in education systems and policies globally (Riyad Shahjahan (2013; Shahjahan & Torres, 2013; Shahjahan & Madden, 2014; Shahjahan, Morgan & Nguyen, 2014). If we look more closely at the composition of the bodies responsible for the international large-scale assessments we can see that these undertakings are not only initiated and governed by the same international organisations but often the international consortiums are composed of the same Anglo-Eurocentric 'players' year after year and hence the basic formula and idea of carrying out these assessments remains more or less the same. This further consolidates the converging and colonialisng effects of international assessments with the consequence of setting at risk those countries that differ socially and culturally from the Anglo-Eurocentric views of measuring learning outcomes.

Second, international assessments have become a powerful political tool to justify national educational reforms. Even when there are plenty of critical studies and commentaries available especially about the PISA results the decision makers seem only be interested in the ranking position of their own country. It is very easily forgotten that actually only a very narrow proportion of success or failure of the system is manifested in the international assessments. If, for example, a country does well in PISA 2012 – which assessed student performance in reading, mathematics and science – what do these three domains actually tell about the whole system? What about things like students' and teachers' well-being, group sizes, school buildings, facilities, curricula, or teacher training as indicators of the quality of the educational system?

Third, the flip side of international assessments is that some educational systems are praised whereas others are doomed to live in a culture of blame. It can be questioned whether this kind of an approach is the best way to develop those education systems that do not perform particularly well in the assessments. Furthermore, as education systems are typically built on the needs of the nation states, is it fair to compare the systems with very different historical development and generate rankings based on this or try to imitate those well-performing systems? Hence, the ethical aspects of assessments, especially justice, equality, responsibility, integrity and tolerance, are easily forgotten. Hopefully these ethical considerations will be taken seriously in a full-scale AHELO which at the time of the writing of this article (summer 2015) is being planned by OECD and will be launched in 2016.

References

Exworthy, M. & Halford, S. (eds.) 1999. *Professionals and the New Managerialism in the public sector*. Buckingham: Open University Press.

Hazelkorn, E. 2011. *Rankings and the reshaping of higher education. The battle for world-class excellence*. New York: Palgrave Macmillan.

Kallo, J. 2009. *OECD education policy. A comparative and historical study focusing on the thematic reviews of tertiary education*. Finnish Educational Research Association. Research in Educational Sciences 45. Jyväskylä: Jyväskylä University Press.

- Kehm, B. M. & Teichler, U. 2007. Research on Internationalisation in Higher Education. *Journal of Studies in International Education*, 11(3–4), 260–273.
- Rinne, R. 2006. Like a model pupil? Globalisation, Finnish educational policies and pressure from supranational organizations. In J. Kallo & R. Rinne (eds.) *Supranational regimes and national education policies*. Turku: Finnish Educational Research Association. *Research in Educational Sciences* 24, 183–215.
- Shahjahan, R. 2013. Coloniality and a global testing regime in higher education: unpacking the OECD's AHELO initiative. *Journal of Education Policy* 28 (5), 676–694.
- Shahjahan, R. & L.E. Torres. 2013. A 'Global Eye' for teaching and learning in higher education: A critical policy analysis of the OECD's AHELO study. *Policy Futures in Education* 11 (5), 606–620.
- Shahjahan, R. & Madden, M. 2014. Uncovering the images and meanings of international organizations (IOs) in higher education. *Higher Education* (iFirst).
- Shahjahan, R., Morgan, C. & Nguyen D. J. 2014. 'Will I learn what I want to learn?' Usable representations, 'students' and OECD assessment production. *Discourse: Studies in the Cultural Politics of Education* (iFirst).
- Teichler, U. 2006. Changing structures of the higher education systems: The increasing complexity of underlying forces. *Higher Education Policy* 19, 447–461.
- Tremblay, K., Lalacette, D. & Roseveare, D. 2012. *AHELO, Assessment of higher education learning outcomes. Feasibility study report, vol. 1. Design and implementation*. OECD.
- Whitty, G. 2010. Marketization and post-marketization in contemporary education policy. In J. Kauko, R. Rinne & H. Kynkäänniemi (eds.) *Restructuring the truth of schooling – Essays on discursive practices in the sociology and politics of education. A Festschrift for Hannu Simola*. Jyväskylä: Finnish Educational Research Association, 23–41.