Informing Universities for Educational Improvement: The AHELO Feasibility Study Experience in Japan, Canada, and Australia

Satoko Fukahori

National Institute for Educational Policy Research

1. Introduction

Assessment of Higher Education Learning Outcomes (AHELO) is an international initiative of the Organisation for Economic Co-operation and Development (OECD), aiming to assess the outcomes of university education using a common test across countries. A Feasibility Study was conducted between 2008 and 2012 to test whether it is possible to implement such an international learning outcomes assessment. With 17 participating countries and regions, the Feasibility Study investigated the cross-disciplinary strand of generic skills as well as the discipline-specific strands of economics and civil engineering. In addition to the tests in these fields, contextual dimension data were collected from all participating students. Japan took part in the Engineering strand together with eight other jurisdictions including Australia and Canada (Figure 1).

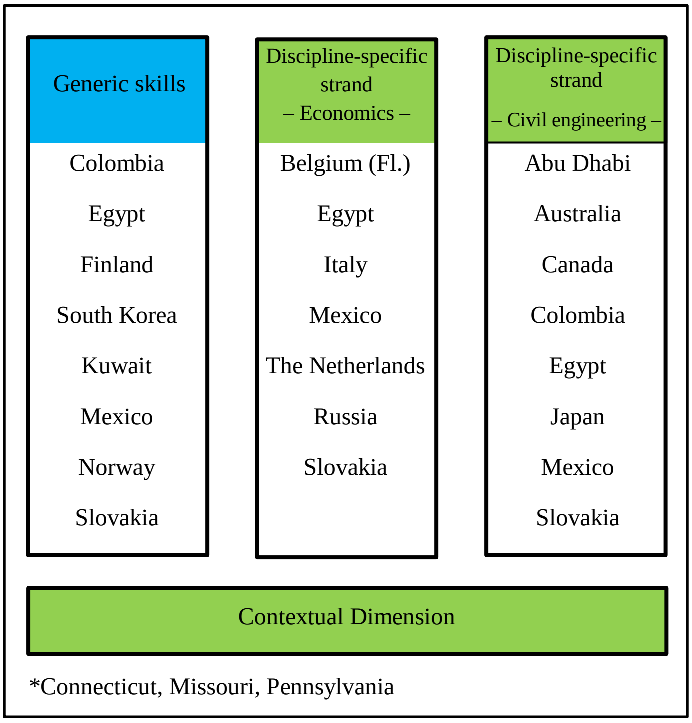

Figure 1. The AHELO Feasibility Study participating countries (by strand)

This paper will first outline the contextual background of the AHELO Feasibility Study, its purpose, and research design, and then describe the organizational structure and implementation process in Japan. Preliminary analysis of the study outcomes will then be presented, followed by a discussion on the significance of this international learning outcomes assessment.

2. Background, purpose, and research design of The AHELO Feasibility Study

What is expected of graduates with bachelor degrees in terms of their knowledge, skills, and attitude (learning outcomes of bachelor degree programs)? Are higher education institutions successful in enabling students to acquire these qualities? As a growing number of students enter universities and as entry level qualifications and paths after graduation diversify, a global trend is emerging to examine the educational quality of higher education institutions from the viewpoint of learning outcomes. Furthermore, the increasing international mobility of students and graduates is stirring an international interest in compatibility and comparability of academic credits and degrees between different countries.

The AHELO Feasibility Study was conceived in such a climate and developed with two goals in mind. The first goal was to verify the possibility of establishing an internationally comparable learning outcomes assessment by forming an international consensus on the knowledge, skills, and attitude students were expected to acquire through higher education, and to clarify whether it was possible to create assessment tools that validly and reliably measure students' levels of achievement.

The second goal was to determine whether it was possible at all to instigate participation from higher education institutions and students and conduct the assessment successfully. AHELO is not designed for uniformity or standardization of higher education. Nor does it expect to serve as an indicator for university ranking or governmental resource distribution. AHELO was an initiative aiming to improve the quality of higher education. The question was whether universities will find value in participating. Another question was whether it was possible, having explained the study's academic and political significance, to persuade students to partake in and undergo the assessment as seriously as they would in their actual courses knowing that the results would not be relevant to their personal academic records.

With these goals in mind, a research design was developed involving two phases. In Phase One, the study aimed to develop assessment tools and evaluate their validity. The plan was to develop cognitive instruments in "generic skills," "economics," and "civil engineering" as well as a questionnaire to gather "contextual" data from university students nearing graduation from bachelor degree programs.

A small-scale field study was conducted to qualitatively verify the validity of the instruments. It was designed to include approximately 100 volunteer students from 10 invited diverse universities in each participating country and to administer the test as well as a questionnaire inquiring the participants' opinions about the validity of the test, followed by group discussions (focus groups) with academic staff and students.

Phase Two involved a large-scale field study using revised instruments and questionnaires from Phase One. The goal was to conduct a quantitative analysis to establish the validity and reliability of the assessment, as well as to verify whether implementation was realistic. The large-scale examination was designed to include approximately 1,000 randomly selected students from 10 universities in each participating country invited for participation. Students were administered the instruments and contextual dimension survey. Furthermore, institutional administrators and faculty were also administered contextual surveys (OECD, 2012: 77-95).

As the sampling design suggests, the sample groups in the AHELO Feasibility Study do not represent the higher education system as a whole in each country. Hence, the data collected is not suited to be used for "cross-national" comparative analysis. Sometimes the AHELO Feasibility Study is dubbed the "PISA for higher education," but it should be noted that sampling design is different from PISA, which employed two-stage stratified random sampling to select nationally representative schools and students, with 15-year-old students as the target population. AHELO is designed to produce information for the improvement of higher education, focusing on universities as the unit of analysis. The Feasibility Study in turn sets out to investigate the validity and reliability of the instruments developed for AHELO.

3. Organizational structures and implementation process in Japan

3.1 Organizational structures in international and national domains

Participating in the engineering strand of the AHELO Feasibility Study, organizations were arranged to pursue respective roles in the study as shown in Figure 2. Descriptions of the organizational structures in international and Japanese domains are as follows:

Beginning with the international structure, the AHELO Feasibility Study had a multilayered decision-making system involving the OECD Education Policy Committee (EDPC), the Programme on Institutional Management in Higher Education (later renamed as OECD Higher Education Programme; IMHE Governing Board or "IMHE GB"), and a Group of National Experts (GNE). The Secretariat undertook actual operations. EDPC, composed of representatives of OECD member countries, considered relevant issues to education policies in each country. IMHE, represented by

IMHE GB, is a forum within OECD for discussions of issues regarding higher education, with higher education institutions and non-profit organizations dealing with problems in higher education as its members. The GNE is organized to discuss AHELO governance principles, joined by government and higher education institution representatives, mainly from participating countries of the Feasibility Study. As this organizational structure implicates, the Feasibility Study had to be conducted upon consensus among very diverse stakeholders associated with higher education.

The tasks of developing the instruments and questionnaire, managing implementation (examination) in participating countries, and analyzing the implementation data to verify validity and reliability were delegated to the AHELO Consortium, headed by the Australian Council for Educational Research (ACER) and comprising the following four groups: "Generic Skills" represented by the US Council for Aid to Education (CAE), "Economics" by the US Educational Testing Service (ETS), "Engineering" by ACER, the Japanese National Institute for Educational Policy Research (NIER), and University of Florence of Italy (Università degli Studi di Firenze, UF),[1] and the "Contextual dimension survey" undertaken by the Dutch Centre for Higher Education Policy Studies (CHEPS) and the Indiana University Center for Postsecondary Research of USA. Some technical specialists were also included in the AHELO Consortium; cApStAn Linguistic Quality Control Agency for translation, the International Association for the Evaluation of Educational Achievement (IEA) and the Data Processing and Research Center (DPC) for database development, SoNET systems for developing an online test system, and Statistics Canada for sampling. The Consortium held meetings regularly, and members pursued their own tasks on the basis of a shared understanding of AHELO principles and methodology as a rule.

Furthermore, a Technical Advisory Group (TAG) was established, consisting of experts in higher education evaluation and research, to provide GNE and the Consortium with third party expert advice (OECD, 2012: 96-100).

Regarding the organizational structure in Japan, the Office for International Planning, Higher Education Policy Planning Division, Higher Education Bureau of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) undertook administrative leadership in the AHELO Feasibility Study under the supervision of the AHELO working group, University Division of the Central Council for Education. NIER acting as the National Center carried out translation and validation of test questions as well as implementation of the assessment.

---

[1] University of Florence is a coordinator organization of the European and Global Engineering Education academic network (EUGENE) with 76 member states.

Furthermore, MEXT delegated to the Tokyo Institute of Technology a Leading University Reform Project “Research on Approaches to the Implementing the OECD-AHELO Feasibility Study” aimed at finding desirable ways of implementing the AHELO Feasibility Study in the engineering strand, and discussing what implications for university reform can be drawn from the experience. This research project proved to be instrumental in involving institutions to deliberate on the possibilities and challenges of AHELO in relation to higher education policies on quality assurance. It should also be noted that expert advice obtained from members of this research project enabled NIER to make proactive contributions to the AHELO Consortium in developing test items, as well as providing active feedback on the implementation process (Tokyo Institute of Technology, 2013).

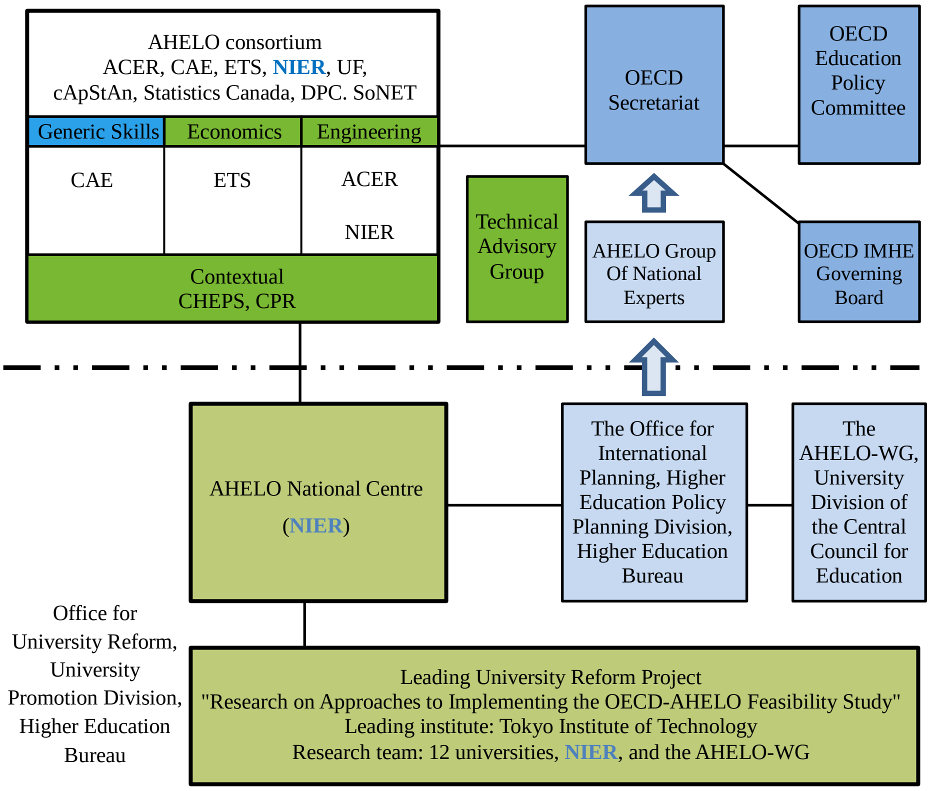

Figure 2. The AHELO Feasibility Study organizational structures (international and Japanese domains)

3.2 Development of competence framework in the engineering strand

What learning outcomes should the assessment measure? “A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering” was developed by the Tuning Association based on the Tuning approach (OECD, 2011a)[2], and approved by the GNE to serve as the competence framework to be referenced for instrument development (OECD, 2012: 121-128).

Tuning refers to the methods and processes of defining competence frameworks that would outline the knowledge, skills, and attitudes students are expected to obtain through their degree programs in respective specialized fields, as well as designing degree programs based on the defined competence frameworks. The Tuning approach is unique in that it requires faculty to develop competence frameworks in consultation with stakeholders, including employers, graduates, and students using language that can be understood by all parties. This approach allows faculty to sustain academic ownership but at the same time be responsive to societal needs. The Tuning approach is unique in that it proposes a common framework that are sufficiently abstract so that diverse universities can share, but explicitly requires institutions to substantiate the competences into concrete learning outcomes that are measurable and attainable within a given timeframe. Degree programs should be developed in a way that are responsive to student needs and are in alignment with institutional missions, but at the same time guarantee the attainment of pursued learning outcomes. Because degrees and credits can be conferred only when the learning outcomes defined by individual institutions but are based on a common competence framework have been attained, Tuning serves both as an internal and external quality assurance system. (Gonzales and Wagenaar: translated by Fukahori and Takenaka, 2012).

The Tuning Project was initiated in 2000 by European universities with the purpose of substantiating the Bologna Process, aimed at establishing the European Higher Education Area. It is administered on a voluntary basis while being financially supported by the European Commission. Competence frameworks were defined in nine subject areas (Business Administration, Chemistry, Education Sciences, Earth Sciences, History, Mathematics, Physics, Nursing, and European Studies) during the first and second phases of the Tuning Project (2000-2004). Since then a diverse array of disciplines has joined the list. Tuning has been adopted in institutions in Latin America, USA, Canada, ,Russia, Africa, Australia, Central Asia, Thailand, and China (Tuning Association, 2013).

In engineering, even before the launch of the AHELO Feasibility Study, a number of pioneering efforts have been taken to establish the international comparability of engineering degree programs through mutual recognition of accreditation and cross boarder accreditation. For example, the Washington Accord was signed in 1989 between signatory countries such as the United States, the United Kingdom, etc. for the mutual recognition of accreditation systems in engineering education.

[2]A similar conceptual competence framework based on the Tuning approach in the Economics strand (OECD, 2011b).

The Japan Accreditation Board for Engineering Education (JABEE) joined the Accord in 2005 (IEA, 2013). In Europe, the EUR-ACE was launched in 2008 as an overarching cross boarder accreditation system for engineering education. Authorized accreditation agencies award EUR-ACE labels to engineering degree programs in recognition of their quality in terms of meeting EUR-ACE standards (ENAEE, 2013). This trend is backed by an increasing number of engineers, whose activities extend beyond national borders, making indispensable a system to recognize the comparability of engineering qualifications in different countries.

In the AHELO Feasibility Study, the competence framework was identified by comparing the Washington Accord and EUR-ACE competence standards, and extracting their common features. As summarized in Table 1, the competence framework in civil engineering was categorized into the following five competence clusters: “Basic and Engineering Sciences,” “Engineering Analysis,” “Engineering Design,” “Engineering Practice,” and “Engineering Generic Skills.”

Table 1. Tuning-AHELO key competence framework in the engineering strand (Summary)

Competence cluster	Description
Basic and Engineering Sciences	Ability to apply knowledge in mathematics, science, and engineering.
Engineering Analysis	Ability to design and conduct experiments, and analyze and interpret data. Ability to identify, organize, and solve engineering tasks.
Engineering Design	Ability to design a system, factors, and processes to meet requirements under realistic conditions of economy, environment, society, politics, ethics, health, safety, production possibility, and sustainability.
Engineering Practice	Understanding of professional ethical responsibility. Knowledge of issues in modern society. Ability to utilize techniques and the latest engineering tools necessary for engineering practice.
Engineering Generic Skills	Ability to act as a member of an inter-disciplinary team. Effective communication skills. Extensive learning experience to understand the significance of engineering solutions in the contexts of international communities, economy, environment, and society. Positive attitude and capability to engage in lifelong learning.

The AHELO Consortium developed the engineering assessment instrument based on this competence framework. It was decided that the competence of “Basic and Engineering Sciences” would be measured by multiple-choice items suited to address the achievement of basic knowledge and skills, while the competences of “Engineering Analysis,” “Engineering Design,” “Engineering Practice,” and “Engineering Generic Skills” would be assessed by constructive response task items suited to address cognitive processes. Multiple-choice items were drafted based on the Civil Engineering Licensing Examination of the Japan Society of Civil Engineers (Organization for Promotion of Civil Engineering Technology, JSCE, 2013) and the First-Step Professional Engineer Examination by the

Institution of Professional Engineers, Japan (The Institution of Professional Engineers, Japan, 2013). The constructive response tasks were prepared by the Australia team, drafting items designed to assess abilities in analytical reasoning and problem solving in real life settings. The instruments were adapted and finalized by an international expert committee, consisting of prominent experts in the field of engineering education in various countries (Australia, Japan, Italy, Germany, Sweden, the United States, and Mexico). Instruments were accompanied by scoring rubrics that describe the required scope and level of learning outcomes to be demonstrated by the students (OECD, 2012: 252-268).

3.3 Implementation

As stated above, the implementation of the AHELO Feasibility Study assessment took place in two phases. Phase One took place between January 2010 and June 2011, and aimed to conduct a small-scale implementation to qualitatively investigate whether the assessment instrument developed by the AHELO Consortium adequately addressed what students in each country had learned at university, as well as whether translated instruments had the quality to measure learning outcomes equal to the original; that is, if there were any discrepancies in the items in terms of the clarity and the levels of difficulty.

Universities were given different sets of items, with each set containing 20 multiple-choice items and one constructive response task, all chosen from the pool of items. The test was conducted in 60 minutes. After completing the examination, students were asked to complete a questionnaire asking their opinions about the validity of the assessment, followed by a group discussion with faculty. In Japan, the small-scale implementation took place between 16 and 25 May 2011, with 75 students volunteering from the 10 universities invited by the National Center to participate.

Phase Two was undertaken between July 2011 and December 2012, using instruments that were revised based on the outcomes of Phase One. It aimed to investigate quantitatively the assessment instrument's validity (whether the instrument actually measured the intended learning outcomes) and reliability (whether the instrument could reproduce the same results if repeated) on a large scale. Furthermore, it aimed to verify the possibility of maintaining certain levels of response rates with the cooperation of higher education institutions and students.

The instrument consisted of 25 multiple-choice items and one constructive response task. The combinations of items were determined based on Item Response Theory, and a total of 18 sets were created. Each student was given a set, and completed it in 90 minutes. In Japan, the National Center invited 12 universities (eight national and four private), and 504 students participated (target population: all civil engineering students; response rates: 12-100%; average: 65%). The implementation took place between 23 April and 25 May 2012. In addition, surveys were conducted with the universities and their academic staff (196 faculty members) (OECD, 2012: 147-172).

3.4 Scoring

In assuring the assessment's validity and reliability, it is imperative that scorers have the same understanding of the target competences and learning outcomes to be assessed, and carry out scoring on the same understanding. In the AHELO Feasibility Study, emphasis was placed on defining the criterion and levels of the scoring rubrics of the constructive response task as well as scorer training.

The small-scale implementation in Phase one adopted the scoring rubrics (first edition) prepared by the AHELO Consortium. In Japan, six engineering experts marked the responses of 75 students. Problems found in the rubrics through the scoring exercise were reported to the AHELO Consortium. For example, they requested additions to the criterion where potentially correct responses could not be accounted for, asked for clarifications of unclear points in the rubrics, and requested re-weighing of scores so that they would better correspond to their levels of importance. Based on such feedback provided from participating countries, the AHELO Consortium prepared revised scoring rubrics (second edition)

International scorer training (two two-day sessions) was held to prepare for the large-scale implementation of Phase Two. Lead Scorers from participating countries gathered at the training sessions. They practiced marking on sample responses taken from small-scale implementations in Japan and Australia, based on the revised scoring rubrics. They carried out discussions until the results of scoring assumed uniformity, and amended the rubrics as necessary (the third and final edition of the rubrics). It should be noted that providing sample Japanese student responses for the international scorer training was instrumental in assuring that the scoring rubrics corresponded well with the performance of Japanese students.

The international scorer training was followed by domestic training sessions in each country, under the leadership of Lead Scorers. The AHELO Consortium developed an online scoring program for the purposes of training and actual marking, to ensure uniformity in the scoring process.

In Japan, 12 engineering experts, led by a Lead Scorer, engaged in a training and scoring exercise for three days. The training sessions were arranged in such a way that the Lead Scorer marked example answers beforehand, and every discrepancy in the marking results of the scorers was discussed among the scoring team members, until unified understanding of the scoring rubrics (final edition) was established.

The training yielded a high level of reliability in the actual scoring exercise. The scoring program was designed to have over 20% of the responses, randomly selected, to be assigned to two scorers, and this double-scoring yielded an 89.1% matching rate in the Japanese case. The discrepancies between two scorers were mediated at the discretion of Lead Scorers, while they provided the team with opportunities from time to time to ascertain the members' understanding of the scoring criterion and levels (OECD, 2012: 173-180; OECD, 2013a: 90-95).

4. Preliminary analysis of assessment results

4.1 Key concepts of assessment result analysis

As stated earlier, the AHELO Feasibility Study employed a convenience sample, where universities were invited by National Centers to participate. This being the case, the institutional sample is not representative of higher education systems in respective countries. As such, it is not appropriate to run a "cross-national" comparative analysis with AHELO data.

Furthermore, it is difficult, strictly speaking, to analyze the outcomes at the student level due to the fact that students took different sets of rotated items based on Item Response Theory. Although in principle, Item Response Theory presupposes the levels of difficulty and score distribution from existing data, enabling standardization and comparison between students answering different sets of items, because the AHELO Feasibility Study instruments are currently in the process of development, they lack the data that allow for such standardization. The decision to employ a research design based on Item Response Theory was a policy-related one, prioritizing the maximization of the number of items to be verified for validity and reliability while minimizing the burden on participating institutions and students, over drawing implications from the data analysis. It should be noted that employing a design that required sophisticated knowledge in psychometrics, resulted in making the instrument less accessible to content specialists, i.e. experts in engineering and education interested in interpreting the results of the tests.

The large-scale implementation was carried out with the purpose of quantitatively verifying the possibility of developing and implementing an internationally comparable learning outcomes assessment. From a psychometric analysis of the assessment outcomes, OECD reached the conclusion that the validity and reliability of the assessment were verified and that the assessment was implementable. OECD thus initiated discussions in preparation for the Main Study of the AHELO project (OECD, 2013b). However, no substantial proposal was made from the the AHELO Feasibility Study framework regarding what information AHELO could deliver that would effectively support higher education institutions in their efforts to improve education. In order to move on to an AHELO Main Study, it is imperative to gain the support from countries and higher education institutions. However, it has not yet been made clear as to what benefits the study will bring to stakeholders.

The research design employed in the AHELO Feasibility Study makes it inappropriate to conduct analyses on a "national" level, while it will take some time before sufficient information is accumulated so that AHELO may draw reliable estimates at the "student" level. What then can be said about the value in conducting a learning outcomes assessment for all the cost?

The following three major points were discussed in the aforementioned Leading University Reform Project "Research on Approaches to Implementing the OECD-AHELO Feasibility Study" regarding educational data that higher education institutions would expect from international learning outcomes assessments.

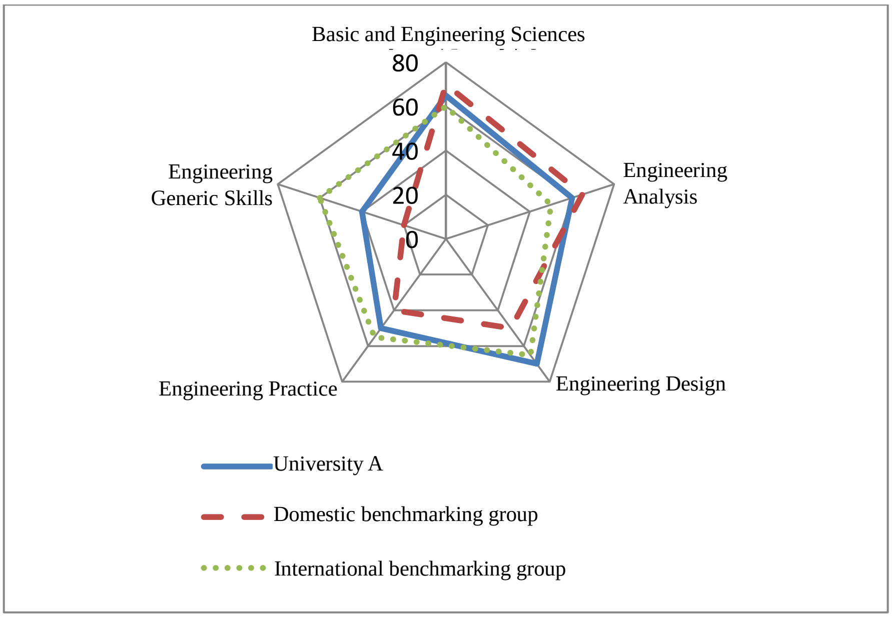

Figure 3. Competence profile (hypothetical)

The first point was about the need for data that served as international benchmarking. It enables universities to know how well their students perform in terms of their learning outcomes against an international standard. Their interests would not be in knowing the world ranking of Japanese universities "as a whole" or "on average." Their interest would be in knowing the levels of their students against the standard of a given group of universities. It was not an issue whether the university group participating in benchmarking was representative of national higher education systems. Rather, they would value information about which specific universities comprised the group.

The second point was about the need for data that informed universities of their strengths and weaknesses. Universities expected information on student performance on the five competence clusters that the AHELO Feasibility Study instruments had aimed to measure, namely, “Basic and Engineering Sciences,” “Engineering Analysis,” “Engineering Design,” “Engineering Practice,” and “Engineering Generic Skills.” Universities would wish to know in which areas their students as a group excel and/or struggle, so that they may better tune their programs to student needs.

Such information would be made possible if students’ competence profile was presented as a radar graph showing the level of performance for each competence cluster, with benchmarking information for the national and international groups, as shown in Figure 3.

The third point was about the need for data that informed universities on the relationships between students’ learning environment, engagement, and test scores. Higher education institutions would utilize such data to address ways to improve their educational programs so that students may more effectively achieve their learning outcomes.

To what extent can AHELO satisfy higher education institutions’ expectations vested in the international assessment of learning outcomes? It must once again be emphasized that the AHELO Feasibility Study was designed to investigate if it is possible to implement an international learning outcomes assessment, and not directly to draw implications about student performance based on data analysis. Although the ultimate purpose of AHELO is to generate data that would inform institutions about educational improvement, the Feasibility Study data does not allow for such analysis. In particular, data collected in the Feasibility Study did not contain sufficient information to conduct analysis on competence clusters. Therefore, the results of a preliminary analysis will be presented here, not for the purpose of producing information on associations, but for the purpose of illustrating the kinds of information a future AHELO might provide.

In the sections below, the relationships between student feedback on the assessment instrument, perceptions about their education, and test scores will be examined to discern the validity of the instrument. Then, the relationship between “student allocation of time,” “approaches to teaching and learning” at the university, and test scores will be examined. The scores used in the analysis are amalgamated scores of multiple-choice and constructive response tasks, which were obtained by estimating standardized scores based on results from multiple choice items that students have taken in common (average score 500, standard deviation at 100; 6,078 students from 70 higher education institutions in 9 countries).

4.2 Validity of instruments

To begin with, we will look at the validity of the instruments. Did the AHELO Feasibility Study successfully measure what it had intended to measure, namely what students learned through university education? We focus on whether the how students felt that the items were valid, and whether such perceptions correlated with test scores.

As shown in Figure 4, about 90% of students felt the instrument items were more or so “relevant to your current degree” (“Very much” 14%, “Quite a bit” 37%, and “Some” 37%), and the scores were higher among those who felt stronger compatibility. Similarly, 75% of students felt that the items asked questions “relevant to future professional practice” (“Very much” 7%, “Quite a bit” 24%, and “Some” 44%), and the scores were higher among those who agreed more with the statement. These results suggest that most students considered the AHELO Feasibility Study as a valid tool to assess the learning outcomes, and that the assessment was relatively successful in assessing the competences of students that were relatively well adapted to university education.

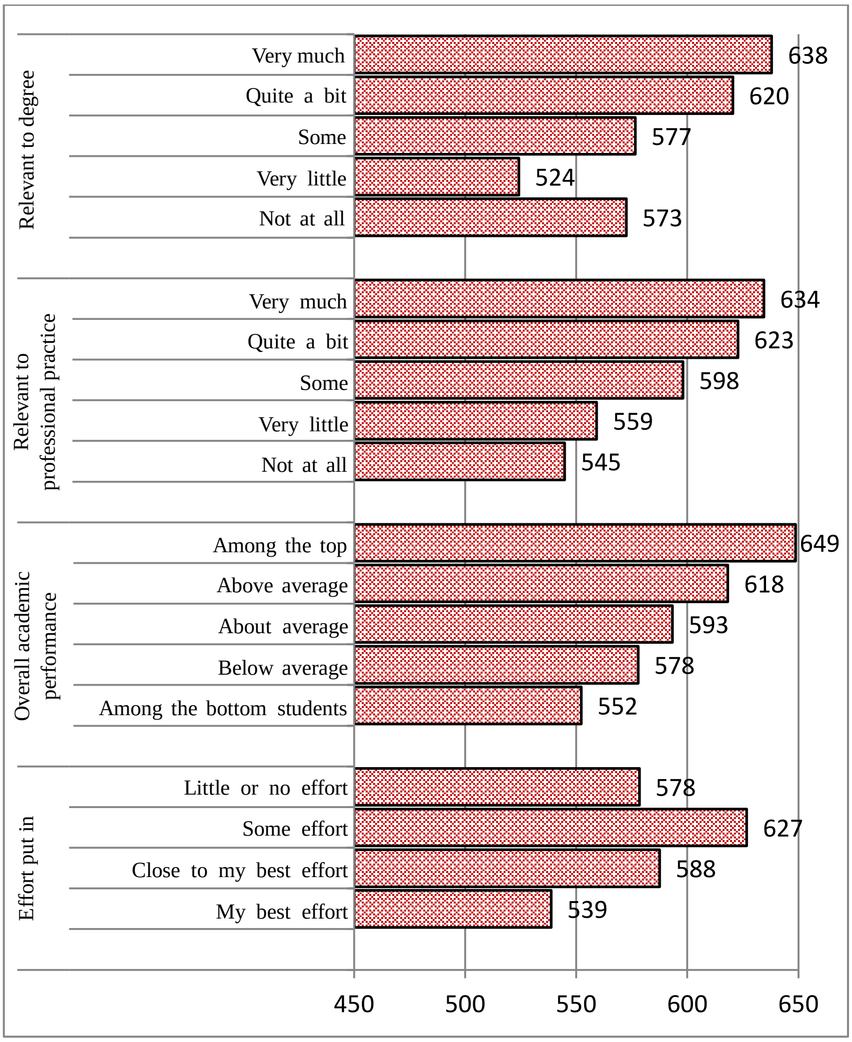

Figure 4. Validity of the instruments: students’ perception and test scores

The test scores did not contradict these students' overall academic performance, as indicated by self-reported academic grades compared with other students in the same degree program ("among the top" 13%, "above average" 25%, "average" 24%, "below average" 22%, or "among the lowest" 16%). Regarding "the amount of effort put into taking the test" ("best effort" 15%, "close to best effort" 45%, "some effort" 40%, "little or no effort" 1%), the students who said they needed greater effort, that is, those who found the questions difficult, had poorer score results. These results again support the notion that the AHELO Feasibility Study assessment was successful in evaluating competence of students who were well adapted to university education.

Now let us turn to the relationship between students' test scores and student evaluation of the quality of their educational programs. About 90% of students considered the curricula they pursued to be "helpful in developing knowledge and skills they expect to use in future professional and working life" ("very much" 6%, "quite a bit" 30%, "some" 50%), again correlating between the extent to which they so felt and their scores. Regarding the students' evaluations of "their entire educational experience so far," only half responded positively ("Excellent" 9%, "Good" 4%, "Fair" 41%, and "Poor" 6%), while the students with a better impression had higher scores. These results once again support the notion that the AHELO Feasibility Study assessment was successful in evaluating competence of students who were relatively well adapted to university education.

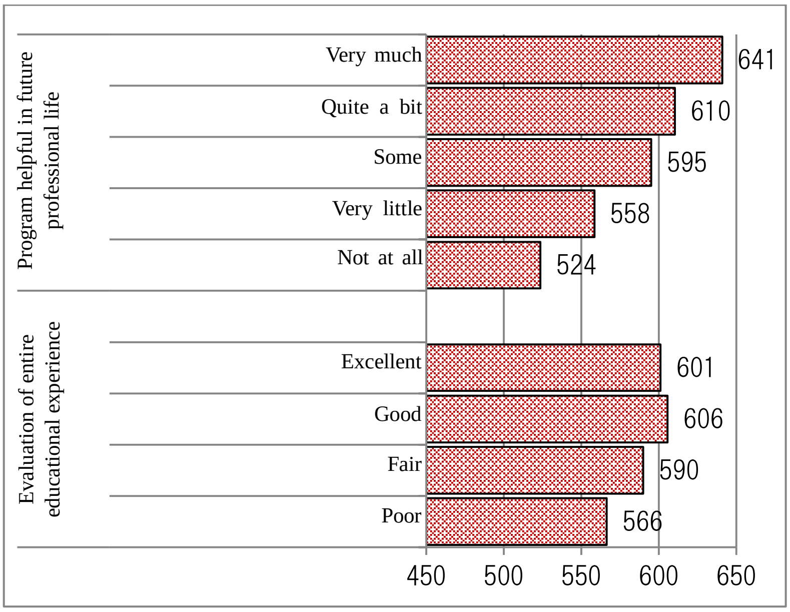

Figure 5. Validity of the instruments: evaluation of program and test scores

4.3 Student allocation of time

How can AHELO inform universities for educational improvement? In this preliminary analysis, let us explore the relationship between students' allocation of time and test scores.

Japanese higher education policies emphasize the importance of ensuring enough time for "preparation, course attendance, and post-class self-motivated learning" for students to accomplish high quality learning at HIGHER EDUCATION INSTITUTIONs (Central Council for Education, 2012). This comes from the awareness that the "total hours for learning" among Japanese students are no more than 4.6 a day on average (Center for Research on University Management and Policy (CRUMP), The University of Tokyo, National Survey of University Students).

In the AHELO Feasibility Study, students were asked their allocation of time during a typical mid-term week (7 days). The results were as follows (figures are weekly average): 18.9 hours for "attending courses (lectures, tutorials, seminars, etc.)," 5.7 hours for "engaging in practical exercises (laboratory work, fieldwork, etc.)," 6.8 hours for "course preparation (reading, doing homework, rehearsing presentations, etc.)," 0.5 hours for "part-time job relevant to their reading subjects," 10.5 hours for "part-time job irrelevant to their reading subjects," and 7.7 hours for "participating in non-academic pursuits at university (on-campus journalism, students union, club activities, etc.)." The average time for these students engaging in academic pursuits amounts to 31.4 hours, or 4.5 hours a day. This figure matches the data from the research cited above.

Meanwhile, the students spend 18.2 hours a week for non-academic activities such as a part-time job irrelevant to their studies, and co-curricular activities on campus. It is notable that Japanese students rarely experience part-time jobs that are relevant to their studies.

To look at the relationship between the ways students allocate their time and their test scores, "attending formal classes" indicates that those who attend the formal courses for less than 10 hours (15%) tended to score low. This may suggest that this student group may be in need of attention for learning support. In terms of the "preparing for class," the "0-5 hours" group, accounting for 56% of the total, had low scores, while the "11-15 hours" group accounting for 10% of the total achieved highest scores. Spending longer than 16 hours does not seem to contribute to better test scores. These results suggest that policies guiding students in the "0-5 hours" group to increase their self-initiated study to the "11-15 hours" level may be beneficial.

On the other hand, engaging in part-time jobs irrelevant to their reading subjects has a negative impact. The longer hours students engaged in such part-time jobs, the lower their test scores. This tendency is starkly apparent for those who spend longer than 10 hours a week. It may be worthwhile to investigate the reason for students spending extremely long hours in occupations unrelated to their majoring studies and to adopt appropriate measures to extend student financial support. In contrast, research in Australia has found a positive impact on students' learning outcomes from engagement in part-time jobs relevant to their majors. Organizing opportunities for students to be involved in part-time jobs that are more relevant to their studies may be an option worth investigating.

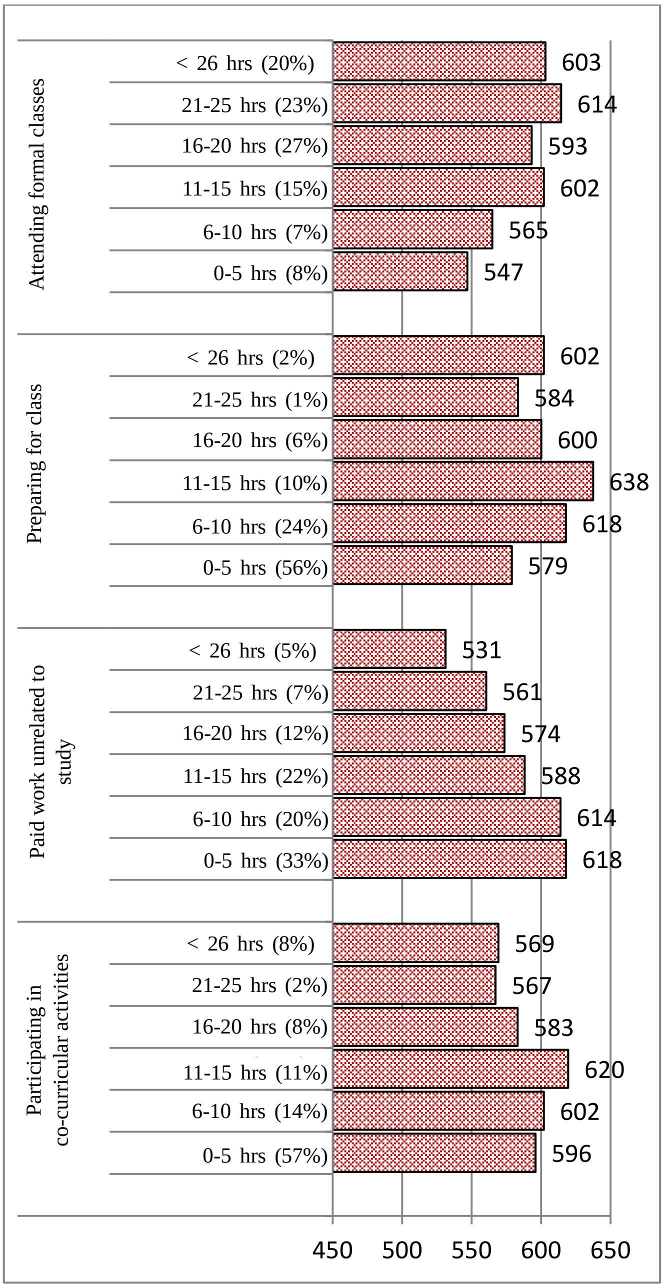

Figure 6. Allocation of time and test scores

Lastly, let us look at the time used for co-curricular pursuits. More than half of the students hardly take part in co-curricular activities, but high-achieving students are mostly found in the "11-15 hours" group (11%). Given that universities serve as a venue for learning as well as for socializing, it may be hypothesized that participation in non-academic on-campus activities brings about positive impact on students' learning outcomes through integration into a comprehensive university experience.

4.4 Approaches to teaching and learning

We now turn to the relationship between approaches to teaching and learning and students' test scores. In terms of the proportions of different course types involved in formal coursework by April of the fourth year, the average proportions of "lectures, where mainly the instructor speaks," "seminars or tutorials, where students are encouraged to discuss subject matter," and "group work, where students work together under supervision/facilitation from an instructor" were 74%, 20%, and 25%, respectively. Students experienced mostly lecture-based courses, and participatory-type courses such as seminars and group work were not prevalent.

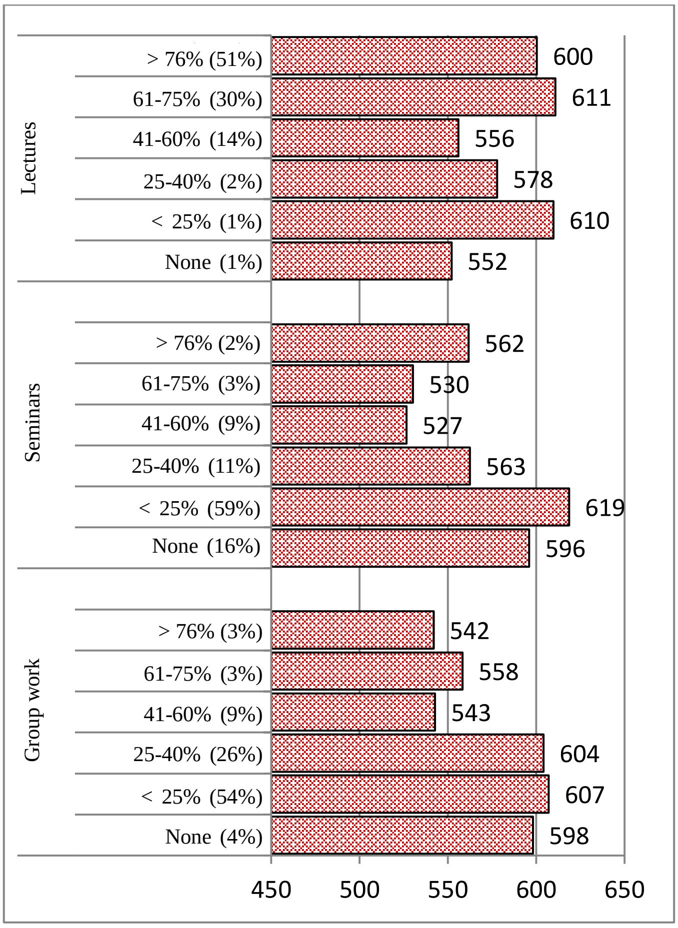

Figure 7. Approaches to teaching and learning and test scores (total)

As shown in Figure 7, cross-tabulations of course types and test scores show that students that take more "lectures" (61% or more of all courses, accounting for 81% of students), fewer "seminars" (less that 25% of all courses, accounting for 75% of students), and fewer “group work” (40% or less of all courses, accounting for 84% of students) type courses tend to have higher test scores.

It is thus possible to draw the following two hypotheses. First, while universities with self-motivated, high-achieving students mostly keep the perpetual lecture-style without making much effort to change, ones with less academically motivated students are more creative in introducing student-centered participatory approaches, including seminars and group work.

Second, the educational impact of participatory approaches such as seminars and group work are difficult to assess. Encouraging self-motivated learning requires a well-developed curriculum design, thorough preparation, and excellent teaching skills on the part of teachers, the conditions for which may be premature in Japan as “lecture-style” courses are still dominant. Alternatively, it may be that by nature, the educational impact of participatory courses is difficult to translate into test scores.

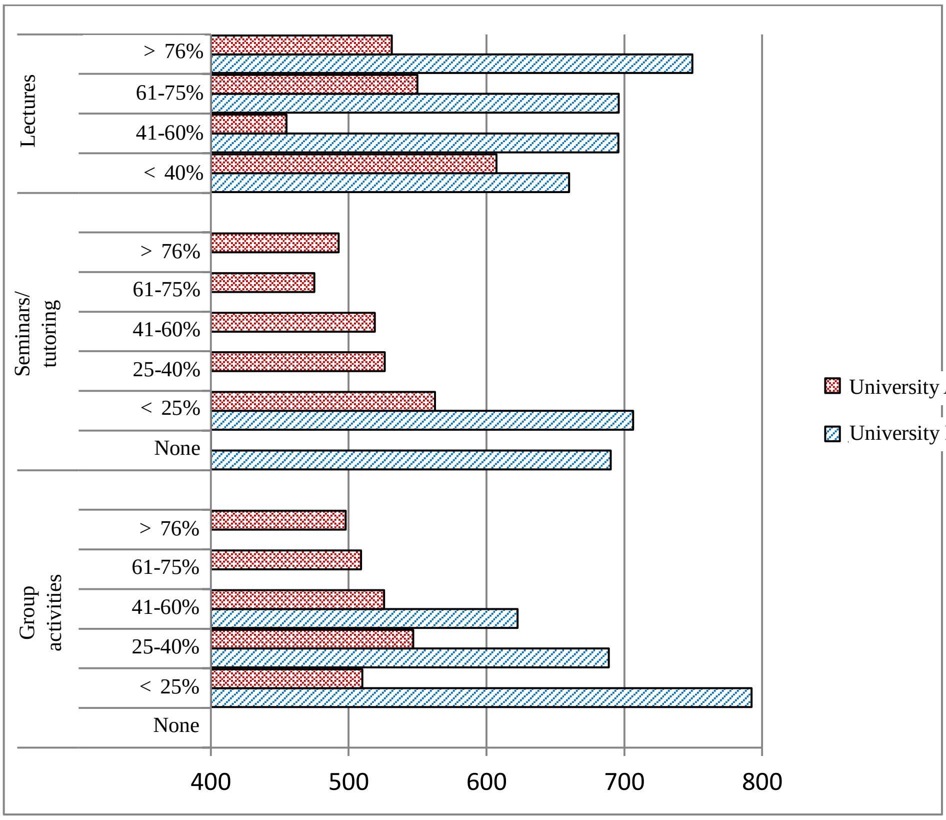

Figure 8. Test scores by approaches to teaching and learning (Universities A and B)

Figure 8 illustrates the difference between two universities, A (n = 45) and B (n = 23), in students’ scores by approaches to teaching and learning. Due to the small sizes of the sample groups, the analysis is far from conclusive, but the observation seems to support the above two hypotheses.

University B, the higher achiever of the two, adopts more “lecture-style” classes and a fewer “seminars” and “group work.” Their students seem to adapt to this style relatively well. University A, on the other hand, provides fewer “lectures” and adopts “seminars” and “group work” in larger proportions. In University A, students taking 40% or less of their entire program had relatively higher test scores. However, no clear tendency is observed showing that students who take more seminars and group work perform better.

4.5 Informing universities for educational improvement

Examples from the preliminary analysis presented above show that the AHELO Feasibility Study has strong potential for providing information universities can utilize for educational improvement. By producing data on achievement levels by competence clusters benchmarked against national and international peers, and by showing the relationships between students’ learning environment, engagement, and test scores, AHELO can inform universities about their strengths and weaknesses and support strategic plans for improvement.

In order to fulfil this potential, AHELO would need to tackle the challenge of producing internationally benchmarked competence profiles of universities. Due to limitations in the data, the AHELO Feasibility Study analyses did not manage to create competence profiles indicating the strengths and weaknesses of universities. Further expert consideration will be necessary to re-design the assessment instrument so that this information will be made possible in a future AHELO.

As for international benchmarking, the National Centers in Japan, Australia, and Canada (Ontario) signed an agreement on data sharing, and have been conducting analyses using combined data. The agreement provides that the three parties mutually respect the confidentiality of the data that belong to each National Center, that the data should not be used for purposes other than for educational research, and that any of the data or analyses thereof should not be disclosed without written consent of other parties. In accord with this agreement, the three National Centers have endeavored to provide participating higher education institutions with international benchmarking data. This effort is expected to become an important precedent for protocols on international data sharing.

5. The significance of international assessment of learning outcomes

The AHELO Feasibility Study has been a significant step forward in that it successfully accumulated information necessary to initiate substantive discussions on the benefits the an international assessment of higher education would bring to stakeholders including, governments, higher education institutions, students, etc., and on the challenges that need to be addressed in order bring about those benefits.

In addition to this accomplishment, our participation in the engineering strand, and the experience gained from the participation in terms of instrument development, implementation, and scoring proved to be fruitful in itself. To close this report, three major accomplishments will be summarized here.

First, participation in the Feasibility Study provided us an opportunity to witness the actual process of internationally shared understandings on expected learning outcomes in the engineering strand develop, as experts from different countries and universities worked together. The assessment instruments and scoring rubrics were embodiments of the consensus. Active dissemination of information on what was learnt through this experience should guide future discussions on enhancing international comparability of Japanese education in engineering

Second, the fact that the Civil Engineering Licensing Examination of the Japan Society of Civil Engineers and the First-Step Professional Engineer Examination by the Institution of Professional Engineers Japan were adopted for an AHELO Feasibility Study multiple-choice items, and the fact that OECD has endorsed their validity and reliability for the international assessment of higher education learning outcomes signify the international comparability of these instruments. These facts signify international recognition that the Japanese standard for the “Basic and Engineering Sciences” competence are compatible with the international standard.

Third, participation in the Feasibility Study has given us updated knowledge about other components of the competence framework in civil engineering, namely “Engineering Analysis,” “Engineering Design,” “Engineering Practice,” and “Engineering Generic Skills.” Although there has been awareness of the importance of these competences, they have not yet been substantiated and widely adopted in educational programs in Japan. By shedding focus on measuring these competences through the AHELO Feasibility Study initiative, we have been able to send a clear message about their importance. The experience has also provided faculty and students concrete ideas about how to measure these competences.

Examples of student comments on the constructive response tasks are shown in Box 1 below. They symbolize the powerful “awareness” that the AHELO Feasibility Study experience has brought about among participants regarding their university education. There are strong expectations for this international initiative to have far-reaching impact on teaching and learning approaches at universities.

Box 1. Students' feedback on the written test

- Student A: "I found it good that the questions were on a practical issue, as we deal with more theoretical, abstract questions more often at university. I enjoyed the ethical question, too."
- Student B: "It was a well-thought-out question in that it made us think of the cause and then respond to it. I think it will be useful in practice in the future."
- Student C: "If what was tested in the written test is important, I feel we should have more group discussions and case studies in our classes. We haven't had them so far, so we haven't had the opportunity to learn approaches to problem-solving."

References

- Central Education Council (2012), *Towards the Qualitative Change of University Education for a New Future, Tokyo: Central Education Council*.
- Tokyo Institute of Technology (2013), *Research on the Implementation of OECD AHELO Feasibility Study (final report)* (Leading University Reform Project sponsored by the Ministry of Education, Culture, Sports, Science and Technology).
- Center for Research on University Management and Policy (CRUMP), The University of Tokyo (2008), *"National Survey of University Students"*.
- Julia Gonzales and Robert Wagenaar eds. (2012). *"Tuning Educational Structures in Europe, Universities' contribution to the Bologna Process: An introduction"*, translated by Satoko Fukahori and Toru Takenaka, Akashi Shoten.
- ENAEE (European Network for Accreditation of Engineering Education) (2008), *EUR-ACE (European Accreditation of Engineering Programmes) Framework Standards for the Accreditation of Engineering Programmes*. (http://www.enaee.eu/wp-content/uploads/2012/01/EUR-ACE_Framework-Standards_2008-11-0511.pdf)
- IEA(International Engineering Alliance) (2013), Graduate Attributes and Professional Competencies (http://www.ieagreements.com/IEA-Grad-Attr-Prof-Competencies.pdf)
- OECD (2011a), "A Tuning-AHELO Conceptual Framework of Expected Desired/Learning Outcomes in Engineering", *OECD Education Working Papers*, No. 60, OECD Publishing. (doi: 10.1787/5kghtchn8mbn-en) (http://www.oecd-ilibrary.org/education/a-tuning-ahelo-conceptual-framework-of-expected-desired-learning-outcomes-in-engineering_5kghtchn8mbn-en)
- OECD (2011b), "Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics", *OECD Education Working Papers*, No. 59, OECD Publishing. (doi: 10.1787/5kghtchwb3nn-en) (http://www.oecd-ilibrary.org/education/tuning-ahelo-conceptual-framework-of-expected-and-desired-learning-outcomes-in-economics_5kghtchwb3nn-en)
- OECD (2012), *Assessment of Higher Education Learning Outcomes Feasibility Study*

Report Volume 1 – Design and Implementation.
(http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf)

Websites

- Organization for Promotion of Civil Engineering Technology, JSCE. *Japan Society of Civil Engineers qualification examination for certified civil engineer system - past examination papers*. (http://www.jsce.or.jp/opcet/02_testQ.shtml). (Accessed on November 01, 2013)
- The Institution of Professional Engineers, Japan. *Qualification examination and registration—information*. (http://www.engineer.or.jp/c_categories/index02019.html). (Accessed on November 01, 2013)
- JABEE. (http://www.jabee.org/). (Accessed on November 01, 2013)
- Tuning Association. *Tuning Educational Structures in Europe*. (http://www.unideusto.org/tuningeu/).(Accessed on November 01, 2013)