Contents

Opening Remarks  (Haruki Ozaki  Director General, NIER) ........................................ 129

Section 1: Presentations

Keynote Speech ................................................................. 133
  How has the current assessment scheme been introduced and how has it been received by HEIs?
  (Tsutomu Kimura)

Presentation 1 ................................................................. 141
  TUNING Competences Frameworks: Key Tools for Educational Reform and the Assessment
  and Comparison of Learning in Global Perspective
  (Robert Wagenaar)

Presentation 2 ................................................................. 160
  The AHELO Feasibility Study: Study Results and the Conclusions of the Technical Advisory Group (TAG)
  (Peter Ewell)

Presentation 3 ................................................................. 170
  AHELO in Japan and Its Prospects
  (Kikuo Kishimoto)

Section 2: Panel Discussion

Report ................................................................. 181
  Informing Universities for Educational Improvement
  : The AHELO Feasibility Study Experience in Japan, Canada, and Australia
  (Satoko Fukahori)

Panel Discussion ................................................................. 199

Closing Remarks  (Tsuyoshi Sugino  Deputy Director-General, NIER) ........................................ 213

Appendix ................................................................. 215
Opening Remarks

Haruki Ozaki*
Director General, NIER

N.B.
• The * mark indicates that the original language of the speech was Japanese and that the transcript is a tentative translation based on the simultaneous interpretation provided during the symposium.
• The transcripts include changes made after the symposium for purpose of publication.
• The affiliations and professional titles of the speakers are as of December 10, 2013.
Hello, ladies and gentlemen. On behalf of the National Institute for Educational Policy Research (NIER), the organizer of the International Symposium on Educational Reform 2013, I would like to say a few words. This symposium has been held yearly since 2001, with the aim of learning from the experiences of countries by inviting the experts who are in the forefront of educational reform, and utilizing the knowledge gained for Japan’s educational reforms. This year’s theme, as you can see above the stage, is Tuning-AHELO Global Quality Assurance through Sharing Competence Frameworks and Degree-Level Specifications. Twenty embassies here in Japan and around 400 participants expressed interest in this theme. Thank you very much for your participation. AHELO (Assessment of Higher Education Learning Outcomes) is a study conducted by the OECD (Organisation for Economic Co-operation and Development) for university students in the final year of a bachelor’s degree program. Its aim is to measure, through a common global examination, the kind of knowledge and skills they have acquired and the level of learning outcomes they have achieved through university education.

A feasibility study to review the possibility of an international learning outcome survey has been conducted in three domains: generic skills, economics, and engineering, with the participation of 17 countries. During the informal OECD meeting of education ministers held in Tokyo in January 2008, Mr. Tokai, Japan’s Minister of Education, Culture, Sports, Science and Technology at that time, expressed his intention to participate in the feasibility study. He decided to participate in the domain of engineering after discussions were held in the working group regarding the OECD assessment of learning outcomes in higher education in the university subcommittee of the Central Council for Education.

As for the National Institute for Educational and Policy Research, we put our efforts into developing examination questions as a member of the AHELO’s international consortium, as commissioned by the OECD. At the same time, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) designated the Institute as the national center for conducting the examination, and we acted as the secretariat for that purpose. “Tuning” represents the method individual universities use to define the competence of a student—i.e., his or her knowledge and skills, in the context of the subject areas—and build a degree program by sharing the competence framework.

Since its birth in Europe in 2000, Tuning has expanded to South America, Russia, North America, Africa, Australia, Canada, Central Asia, Thailand, and China. In the economics and engineering disciplines of AHELO, the definition of the competence framework has been set through the Tuning Method, and test questions have been created. That is why the theme of this symposium is Tuning-AHELO. The efforts of Tuning and AHELO include an extremely crucial insight into Japan’s higher education policies. As you know, in Japan, in order to cultivate human resources that will drive our globalizing knowledge-based society, high expectations are placed on university education. On the other hand, the issues related to the Quality Assurance of University Education are gaining prominence, as Japan’s university population is expanding. At the same time, the 18-year-old population is declining. The Japanese
educational policy has pursued the path of diversity and individuality since the end of the 1980s. However, in recent years, the weight has been shifting to the creation of a common framework that will become a platform of quality assurance.

The shift was evident in the Central Council for Education’s so-called “report on quality transformation” that came out in 2012. The focus is on the necessity of establishing a degree program and setting up education management properly. The environment of Tuning in AHELO is the specific method for sharing competence frameworks and degree level specifications—important themes that are very timely.

Today’s symposium has invited seven experts from overseas and Japan. I would like to extend my heartfelt welcome to all experts who are participating in this symposium despite their busy schedules. The first keynote speech will be given by Mr. Tsutomu Kimura, present advisor to the Minister of Education, Culture, Sports, Science and Technology, and former president of the National Institute for Academic Degrees and University Evaluation, a position he held for many years. He will be speaking of current issues regarding university assessment systems. The first presentation of this section will be made by Prof. Robert Wagenaar, Director of Undergraduate and Graduate Studies of the University of Groningen, who has been involved in important initiatives to harmonize European higher education, such as the ECTS (European Credit Transfer and Accumulation System), framework for qualifications of the European higher education area, and European qualification framework for lifelong learning. He will share with us the role that Tuning and competence framework should play in the global measuring and comparison of educational reform, and learning.

After Prof. Wagenaar, Dr. Peter Ewell, vice president of the National Centre for Higher Education Management Systems and chair of the AHELO Technical Advisory Group (TAG), will speak about the results of the AHELO feasibility study and the conclusions made by the technical advisory group. Next, Prof. Kikuo Kishimoto from the Tokyo Institute of Technology will talk of AHELO in Japan and its prospects. An expert in AHELO’s engineering discipline, he has played a central role in the development of globally acceptable test questions and has also pushed for the implementation of the study in 12 universities in Japan.

In the panel discussions in Section 2, the facilitator will be Prof. Motohisa Kaneko, a technical advisory group member of AHELO like Dr. Ewell. Prof. Kaneko has reviewed the feasibility study. We also welcome two other panelists: Ms. Mary Catharine Lennon, a Senior Research Analyst at the Higher Education Quality Council of Ontario, and Dr. Daniel Edwards, a Principal Research Fellow at the Australian Council for Education Research.

Before we start the discussions, Dr. Satoko Fukahori, our Institution’s senior researcher, will report on the AHELO feasibility study in Japan, Australia, and Canada, from the viewpoint of utilizing study results; she will also join the panel discussions. (Dr. Fukahori was the emcee at the start of the symposium.) The AHELO feasibility study was concluded last December and is currently being reviewed by OECD as to whether or not it will evolve into a main study. Regardless of its direction – for Japan’s higher education policies that aim for a systematic degree program and for the universities working on specific majors – the insights from this Tuning-AHELO symposium will be extremely important. I expect the symposium to contribute tremendously toward the improvement of Japan’s university education. With this, I would like to conclude my speech on behalf of the organizer. Thank you for your kind attention.
Section 1: Presentations

Keynote Speech
Tsutomu Kimura
Advisor to the Ministry of Education, Culture, Sports, Science, and Technology

How has the current assessment scheme been introduced and how has it been received by HEIs?

Presentation 1
Robert Wagenaar
Director of Undergraduate and Graduate Studies, University of Groningen

TUNING Competences Frameworks: Key Tools for Educational Reform and the Assessment and Comparison of Learning in Global Perspective

Presentation 2
Peter Ewell
Vice President, National Center for Higher Education Management Systems

The AHELO Feasibility Study: Study Results and the Conclusions of the Technical Advisory Group (TAG)

Presentation 3
Kikuo Kishimoto*
Dean, Graduate School of Engineering, School of Engineering, Tokyo Institute of Technology

AHELO in Japan and Its Prospects

N.B.
• The * mark indicates that the original language of the speech was Japanese and that the transcript is a tentative translation based on the simultaneous interpretation provided during the symposium.
• The transcripts include changes made after the symposium for purpose of publication.
• The affiliations and professional titles of the speakers are as of December 10, 2013.
Good afternoon, distinguished guests, ladies and gentlemen. I’m extremely honoured to be here as the keynote speaker of this very ambitious international symposium organised by National Institute for Educational Policy Research. Before beginning my presentation, I would like to express my sincere appreciation towards the participants who have flown a long way from different parts of the world to Tokyo. I have to say this because I have been working as the chairperson of advisory board for National Institution for Educational Policy Research. In fact, I was in Brussels last week. I just got back to Tokyo Saturday last week, which is only 3 days ago. So I’m still suffering from very bad jet lag, which is a usual sort of a routine for me in these years, probably because of old age. I hope none of our foreign colleagues are in that condition.

My presentation is centred around the current assessment scheme going on in this country. It is slightly away from the topic of Tuning and AHELO, but I believe that this type of assessment is one step before tuning. As many other countries in the world, Japan has been trying very hard to achieve higher education or university reform for the past, say, 20 to 30 years. The current assessment scheme is certainly one of the products of that effort. My presentation consists of two parts. In the first part, I would like to describe how the idea of implementing new scheme has been developed in the course of higher education reform in this country. In the second part, I would like to show you how the current assessment scheme has been received by higher education institutions, by showing the results of post-event surveys carried out by National Institution for Academic Degree and University Evaluation to which I used to belong.

The majority of the proposals in the line of higher education reform in this country came from the University Council. This University Council was established on the basis of the request put forward by Emergency Council for Education. The Emergency Council for Education was very powerful. I would dare to say it is the most important inquiry committee in the history of discussions at all levels of education in this country. This Council for education was established by then Prime Minister Nakasone who was a very far sighted Prime Minister. The Council identified the different types of problems at all levels of educations. Particularly with respect to higher education, the Council pointed out that higher education is one of the weak points of Japan. This message came out around late 1970’s to 1980’s. At that time, Japan was boasting its prospective economy and unheard-of prosperity in many sectors. All the Japanese were very confident about everything. “Japan as number one” is the phrase which described Japan at that time. Therefore this type of critical message sent out by the Emergency Council for Education has met fierce resistance from various sectors. But I still think that the indication by the Council was exactly right. Receiving the criticism from the Council, Ministry of Education set up the University Council immediately. It was in 1987. The Council continued to exist until 2000. This Council was very active and worked very hard. The Council has produced more than 20 reports, I believe. The number of proposals they hammered out has exceeded one hundred. Amazing
activities!

The first report came out in 1988. The Council said that this country has to diversify the missions of PhD courses. Traditionally, PhD courses in this country were considered to be organizations or institutions to produce high quality researchers, or high quality academicians. But the Council bravely said that this mission should be diversified so that PhD courses can also foster highly professional human resources who can work at non-research sectors of the government, and even at industry. On receiving this recommendation, this country has carried out some improvement, but still we have a big problem in this line. Majority of the graduates of PhD courses still tend to work in research-related areas.

In 1991, a very important comprehensive report came out. I myself call this the 1st highlight report because of its comprehensiveness. This report includes many important proposals. For example, for the first time in the history of discussions related to higher education, the Council pointed out the importance of teaching at university levels. Before this time, Japanese university used to think that only mission they had was to carry out first class research. But the Council clearly said that the universities have to recognise the importance of teaching. With respect to teaching reform, the Council put forward a proposal to deregulate university curricula. This has given a big impact on the Japanese university. Immediately after the last war, this country carried out a drastic reform in curricula construction. As a result Japanese universities came to have two groups of disciplines. One is the general or liberal arts education group and the other is specialised education group. At most universities, general education was provided in the first or second year after the students’ entering a university. Then specialised education follows. In the beginning, this system worked well. However as the time goes on, a very serious discontinuity between general education and specialised education began to appear. Facing this situation, the Council said that this dual system should be abolished. The Council proposed that the freedom to design the curriculum should be given to each university.

The most important proposal in this report is perhaps the indication of importance of self-check and self-assessment. I can say that this is the first sign of recognising the importance of assessment in this country. Before this, this country was not aware of the importance of assessment in higher education.

Another important proposal is the expansion of graduate programmes. Traditionally, Japanese’s higher education sector placed higher emphasis on undergraduate courses. The Council said that priority should be shifted from undergraduate to graduate programmes as in other developed countries in the world. In 1993, another report came out and the Council says that all the universities in this country have to open the gate wider so that they could accept part-time students. Accepting part-time students into a Japanese university was unheard-of before this time.

Then in 1998, another very comprehensive report was put forward. I named this the second highlight report. This report also included many important proposals. It pointed out the necessity of re-examining undergraduate teaching, together with promotion of faculty development. Another important proposal is the last one shown in this slide. For a long time, Japanese universities were considered to be so-called an ivory tower. Even the government and politicians were not able to say anything critical about their performance and management. The Council was very brave in saying that officers such as the presidents and deans have to have strong leadership to realize efficient management. This indication has given a strong impact on the management system of Japanese universities. This again is the right indication. The most important proposal is perhaps the indication of importance of setting up of third party evaluation system. This report came out in 1998. As I explained earlier, we had the first sign related to the importance of assessment in 1991. So, seven years after the first sign, the Council proposed to establish the third party evaluation system. Receiving this proposal, Ministry of Education moved very fast. In the year 2000, National Institution for Academic Degrees which is a Japanese version of CNAA in UK, was reorganized so that they can take the responsibility of newly established accreditation and evaluation scheme. The name of the institution was changed to National Institution for Academic Degrees and University Evaluation(NIAD-UE). The other two agencies were also recognized as the official organizations. The official name of these organizations is certified evaluation institution.
This second one is Japan University Accreditation Association of which abbreviation is JUAA. The membership of this organisation comes mainly from private universities with long history. The third one is Japan Institution for Higher Education Evaluation abbreviated as JIHEE. Its membership comes mainly from private universities founded in recent years. These three organisations were certified as the official evaluation institutions for four-year universities. Each university can choose one out of these three. The first implementation was carried out in the year 2004.

Now I’d like to summarise what I have introduced to you up to this point. Keywords are these three: diversity, individuality and third-party evaluation. The University Council kept saying that each university should have its individuality. The Council stressed the importance of introducing the third party evaluation system. The official name of the third-party evaluation was set as “certified evaluation and accreditation”. It is a rule now that all the universities in this country, nearly 750 of them, have to go through this evaluation and accreditation once in seven years. Seven years is one cycle. National Institution for Academic Degrees and University Evaluation finished the first cycle in 2011. I can say that, this new scheme has gone fairly well. But there is one problem. In this scheme research assessment is not compulsory. Only teaching performance is looked at. A question arises. Can we compete with universities in other countries in the world by only carrying out assessment on teaching performance? The answer is definitely no. So what the Ministry of Education did was to devise new funding schemes, such as promotion of CEO (centre of excellence) programmes, promotion of WPI (world premier institute) and many others. Through these funding schemes, research oriented universities are now able to obtain sizable amount of research fund to keep their international level research going. What the Ministry of Education has been doing is combining these two schemes. Teaching side is strictly assessed by newly introduced evaluation and accreditation system, and by introducing the competitive funding system, the Ministry tries to enhance the quality of research.

Now I will move on to the second part. Here I would like to show you results of post event questionnaire survey carried out by National Institution for Academic Degrees and University Evaluation. Different types of question were given to the universities which had received assessment by NIAD-UE. The first category of questions is on assessment criteria set up by NIAD-UE. The first question is; “Are criteria set up by NIAD-UE appropriate in order to assure the quality of your research and teaching activities?” Many questions of this type were given. By the way, Y05 represents the year 2005 and 11 is for year 2011. This is for the first cycle. This T indicates a total number of higher education institution which have received the assessment by NIAD-UE; 5 in year Y05 and 38 in year 09. The universities were asked to give responses in five grades. Five for very positive, four for positive, three for neutral, two for somewhat negative, one for very negative. There’s one thing we have to be very careful about. Culturally perhaps, Japanese are quite reluctant to give negative answers to this type of official surveys. So three, although it is neutral, we have to interpret this as slightly negative. I want you to remember this.

Let’s have a look at this one. Overall responses are very good, so they are very satisfied with this. What about next question? Responses are very positive. However, when it comes to the question of, “If this criteria help your university gain societal support for their research and teaching activities?”, responses are not very positive. They are doubtful about its effect. This shows that NIAD-UE has to revise the assessment criteria in this line. The next one is on the quality of self-assessment report. Each university has to write out a self-assessment report in order to receive the assessment. A question was asked; “Do you think that the self-evaluation report you wrote was easy to understand?” Slightly negative response. Some universities are not very confident about the quality of their own report.

Then next. The effect of the final assessment report prepared by NIAD-UE. The response is good. But when it comes the story of understanding and the support of general public, the response is generally negative. There are some very negative responses. The next question; “Were you able to obtain new perspectives for the activities of research and teaching through the final report?” The responses are fairly negative. Some are very negative. So we can say that the final report is very good for reviewing what universities did in the past, but that it cannot give useful suggestions to the future direction.
Section 1: Presentations

The next is on the impact of carrying out self-assessment. The response is good, but for the question: “Did, the majority of your staff come to understand the importance of carrying out self-assessment?” The responses are very negative. This means that universities are having a problem in persuading their members to understand the merits of this assessment scheme. We see a similar situation for the question: “Did the majority of your staff come to understand the importance of carrying out research and teaching activities systematically all across your institution?” The response conjures up a problem at Japanese universities. Traditionally, Japanese universities have a very strong autonomy particularly at the faculty level. In order to establish efficient management within a university, we have to eliminate excessively strong autonomy of each faculty. Probably, the responses we see here may have been put forward by officers of universities such as Presidents, Deans, committee chairpersons, and so on. That is a reason why responses are very negative. This shows that officers are having a difficulty in making the general staff mobile towards the direction taken by the country.

The next question; “Were you able to improve the quality of management of your institution as a whole through the final report?” Again responses are not that positive. What about the development of new challenges in research and teaching activities? Very negative. They say that certainly for the past achievements, the indications of the final report are useful, but not for the enhancement of future or new challenges.

Now I’ll come to a conclusion. Generally speaking, the current assessment scheme has been well-received by higher education institutions. But it seems to be difficult for higher education institutions to convince their staff of the merits of the assessment, although the officers have understood them. I feel that the assessment criteria need to be modified so that research and teaching activities of higher education institutions can gain understanding and support from the students as well as general public. Further efforts are necessary to enhance unique and innovative future challenges of each institution. Future is the keyword.

This is the final slide. In 1998, the University Council put forward a report titled “How should higher education institutions in Japan be in 21st century?” The report stressed the importance of individual institution’s effort to generate its uniqueness and individuality and of carrying out research and teaching activities in a competitive environment. These are mottos set for the university reform in this country. With respect to competitive environment, I must confess that we do not have it yet. Has the current assessment scheme I just explained created a competitive environment in higher education sector in this country? The answer is “No” . I always compare our current assessment scheme with the UK research assessment exercise and teaching quality assessment. There are big differences in the two countries’ scheme with respect to creation of competitive environment. What this country has to do is to hammer out a new assessment scheme, which enables to build up a competitive environment. This is the key to success for universities reform in this country. This is my personal view. One of the solutions may be to carry out discipline-based assessment like the one UK has been doing. The key element is the quality of research. We haven’t had a look at it yet. Also students learning outcomes are important. These two have to be incorporated into a new assessment scheme in order to create a competitive environment. I was in UK when they had a big fuss about the results of research assessment exercise. I think it’s happened in the field of history. The story was Oxford Brookes University versus Oxford University. Oxford Brookes is a small university. The social pecking order for Oxford Brookes is not that high, but in that year, Oxford Brookes got higher mark than Oxford University. That was a big news in all the media in England. A system like this will certainly create a competitive environment. I would like to see this happen also in this country.

Thank you very much indeed, for your attention.
How has the current assessment scheme been introduced and how has it been received by HEIs?

Advisor, Ministry of Education, Culture, Sports, Science and Technology
Tsutomu Kimura

1. Proposals related to university reform in recent years
   —From proposals of University Council—

   Emergency Council for Education
   「Higher education is one of the weak points of Japan」
   1987 University Council founded
   (1987 ~ 2000)

Outlines of proposals
1) 1988.12
   • Fundamental issues related to graduate courses
   • Diversification of the missions of PhD courses
     Fostering highly-professional human resources
     (Not only academics with high-profile)

2) 1991.2 : The 1st highlighted report
   • Drastic improvement of teaching
     ~ Emphasis on teaching for the 1st time
   • Deregulation on curricula
     → Abolishing the difference between general education and specialized education
     → Creating unique curriculum by individual university
   • Self-check and self-assessment ~ The 1st sign of importance of assessment

3) 1993.9
   • Acceptance of part-time students

4) 1996.10
   • Introduction of limited-term employment for academic staff ~ Increasing mobility of academic staff

5) 1998.10 : The 2nd highlighted report
   • Restructuring of undergraduate teaching F D
   • Advancement of teaching and research at graduate schools
   • Setting up of professional graduate schools
     Law schools to start with
   • Introduction of efficient management system
     Leadership of president and deans
     Setting up of advisory board
Section 1: Presentations

1. Introduction of multi-faceted university evaluation system
   → A third-party evaluation system recommended.

2000 National Institution for Academic Degrees
   → National Institution for Academic Degrees and University Evaluation (NIAD-UE)
   Japan University Accreditation Association (JUAA)
   Japan Institution for Higher Education Evaluation (JIHEE)
   ~ Certified evaluation institutions

2004 The introduction of the certified evaluation and accreditation system.

5. Half-way summary
   Diversity, Individuality, Third-party evaluation

Certified Evaluation and Accreditation (7 years’ cycle)
Research evaluation ~ Not compulsory
Others (e.g. teaching, management) ~ Absolute: In reference to each institute’s own missions

Can we compete with universities in other countries? Definitely no!

Introduction of competitive research grants such as COE programme or World Premier Institute

6. Results of post-event questionnaire survey
   by NIAD-UE

1. Assessment criteria set up by NIAD-UE

   • Appropriate in order to assure the quality of R&T activities

   V 05 1 3 1 0 0 5
   V 06 1 8 1 0 0 11
   V 07 4 19 14 2 0 39
   V 08 1 9 5 0 0 15
   V 09 1 21 15 1 0 38
   V 10 4 15 11 0 0 30
   T 13 58 30 4 0 143

2. Quality of self-assessment report

   • Was able to produce a report for general public easy to understand

   V 05 2 2 1 0 0 5
   V 06 0 9 2 0 0 11
   V 07 5 27 6 1 0 39
   V 08 11 0 2 0 0 13
   V 09 9 23 8 2 0 38
   V 10 6 20 4 0 0 30
   T 21 2 6 0 0 143

10. 2. Results of post-event questionnaire survey
    by NIAD-UE

   • Appropriate in order to accelerate improvement of the quality of R&T activities

   V 05 1 3 1 0 0 5
   V 06 2 8 1 0 0 11
   V 07 5 31 3 0 0 39
   V 08 1 11 1 0 0 13
   V 09 4 28 5 1 0 38
   V 10 7 9 3 0 0 30
   T 20 108 11 1 0 143

11. 2. Results of post-event questionnaire survey
    by NIAD-UE

   • Appropriate in order to gain societal support for R&T activities

   V 05 1 3 1 0 0 5
   V 06 2 8 1 0 0 11
   V 07 5 31 3 0 0 39
   V 08 1 11 1 0 0 13
   V 09 4 28 5 1 0 38
   V 10 7 9 3 0 0 30
   T 20 108 11 1 0 143

12. 2. Results of post-event questionnaire survey
    by NIAD-UE

   • Was able to produce a satisfactory report

   V 05 2 2 1 0 0 5
   V 06 0 9 2 0 0 11
   V 07 5 27 6 1 0 39
   V 08 11 0 2 0 0 13
   V 09 9 23 8 2 0 38
   V 10 6 20 4 0 0 30
   T 21 2 6 0 0 143
Generally speaking, the content of the report is satisfactory. The majority of the staff came to understand the importance of carrying out self-assessment, which was useful in improving the quality of R&T. Adequate in order to assure the quality of activities related to R&T. The impact of carrying out self-assessment includes being able to understand issues to be solved in R&T activities, useful enough in order to improve the quality of R&T, adequate in order to assure the quality of activities related to R&T, and improved the quality of management of the institution as a whole. Staff's consciousness to engage in R&T activities has increased, improved the quality of management of the institution as a whole. Contents and quality of final assessment report prepared by NIAD-UE.

| 05 | 06 | 07 |
| 08 | 09 | 10 |
| 11 | 12 | 13 |

| 05 | 06 | 07 |
| 08 | 09 | 10 |
| 11 | 12 | 13 |

| 05 | 06 | 07 |
| 08 | 09 | 10 |
| 11 | 12 | 13 |

| 05 | 06 | 07 |
| 08 | 09 | 10 |
| 11 | 12 | 13 |

| 05 | 06 | 07 |
| 08 | 09 | 10 |
| 11 | 12 | 13 |
Section 1: Presentations

5. Outcomes of this assessment if it continues

- Will assure the quality of R&T activities
  Y 2005 0 2 2 0 0 1
  Y 2006 1 2 2 0 0 1
  Y 2007 0 1 1 0 0 1
  Y 2008 1 1 1 0 0 1
  Y 2009 0 1 1 0 0 1
  Y 2010 0 1 1 0 0 1
  Y 2011 0 1 1 0 0 1

- Will be able to obtain understanding and support of general public for R&T activities
  Y 2005 0 2 2 0 0 1
  Y 2006 1 2 2 0 0 1
  Y 2007 0 1 1 0 0 1
  Y 2008 1 1 1 0 0 1
  Y 2009 0 1 1 0 0 1
  Y 2010 0 1 1 0 0 1
  Y 2011 0 1 1 0 0 1

- Will promote new challenges in R&T activities
  Y 2005 0 2 2 0 0 1
  Y 2006 1 2 2 0 0 1
  Y 2007 0 1 1 0 0 1
  Y 2008 1 1 1 0 0 1
  Y 2009 0 1 1 0 0 1
  Y 2010 0 1 1 0 0 1
  Y 2011 0 1 1 0 0 1

Summary

Generally speaking, the current assessment scheme has been well received by HEIs, but it seems to be difficult for HEIs to convince their staff of merits of the assessment. Assessment criteria need to be modified so that R&T activities of HEIs can gain understanding and support from the students and general public. Further efforts are necessary to enhance unique and innovative future challenges in R&T activities of HEIs.

Thank you so much for your kind attention

In 1998 University Council put forward a report titled “How should HEIs in Japan be in 21st century? — Proposals of reform plans—.” The report stressed the importance of individual institution’s effort to generate uniqueness and individuality in its R&T activities in competitive environments.

Has the current assessment scheme created competitive environments in the HE sectors in this country? The answer is probably no. What the country has to do next is to hammer out a new assessment scheme which enables to build up competitive environments. One of the solutions may be discipline-based assessment like the one UK has been doing. Key elements are quality of research (R) and students’ learning outcomes (T).
Introduction

During the last two decades the playground of Higher Education has changed fundamentally. Globalization and Information and Communication Technology have given a further push to its internationalization. The required level of quality and effectiveness of higher education programmes are no longer determined at local or national level only, but are also referenced internationally today. This does not imply that it is felt like that by the majority of academic staff. One can observe a striking difference between opinions and viewpoints of governmental authorities, management of higher education institutions, its faculties and departments and academic leaders on the one hand and the average academic on the other. One can also see differences between world regions.

Nevertheless, whether one likes it or not, in particular for the more renowned institutions in every country competition in terms of attracting academic staff, young researchers / PHD-students and master and bachelor students has moved from the national to the international scene. Academics as well as students identify and select higher institutions which serve their interests best. Due to search engines, portals, web presentations, and new forms of communication such as Facebook and other social media, this process has become a global one. As a result the student body has changed in a large and growing number of institutions as has often - partly - the language of instruction. Academics are noticing that students have become more demanding regarding the content of educational programmes as well as the learning and teaching process. Not only because they have a wider and easier access to information about higher education institutions, what is taught there, how it is taught and by whom, but also because this information is easily exchanged via social media . This involves risks, for all involved, because images and reputations can easily be built or destroyed, rightly or wrongly. This is also the case for the quality of education which is offered. At present, there is more focus on this issue than ever before. Universities develop so-called quality cultures as a result of external pressure in particular. A key question in this respect is what and who decides what high quality programmes are and on which basis?

Besides what has been said above, and besides personal development, there is another dimension which has a growing impact on the content, implementation and modes of delivery of a higher education programme: its relevance for society. Relevance is understood here in terms of preparing for citizenship and for employability. In particular employability seems to be given more and more weight and is therefore competing with the actual interests and abilities of the student. This is understandable in a situation where an economic global crisis is impacting so many, but it might lead to wrong choices, possibly followed by a growing number of drop-outs. This can be a costly affair in more than mere financially. Although the chances for obtaining employment at a suitable level after graduation for one programme might be better than for another, this does not imply that less successful programmes in this respect
do not have an obligation regarding the transition to society of its graduates. Education is simply not intended to be ‘art for the sake of art’. Degree programmes are also not intended to mirror the academic profiles of the teaching staff itself in today’s dynamic world. This has implications for the design and delivery of programmes as well as the for the competences which are developed and the desired outcomes.

Searching for a new paradigm

In 2001 the Tuning Educational Structures in Europe project (hereafter: Tuning) was launched with these notions in mind. At the end of the 1990th a growing concern developed among the initiators and developers of the European Credit Transfer and Accumulation System (ECTS) that one credit system would not be a sufficient answer to societal challenges and was not the only panacea for (trans)national mobility of students and cooperation between higher education institutions in and outside Europe. It was concluded that the emphasis should be much more on the outcomes of the learning process, but also that more attention should be given - within degree programmes - to transferable or general skills. It was also thought necessary to stipulate the role of higher education institutions as major contributors to the welfare of society. This fitted in a trend in which higher education institutions were forced to show that they are accountable, responsible, and sustainable. This should not apply to the higher institution only but also to the individual degree programmes on offer. A number of conclusions were drawn from this analysis:

▶ Higher Education structures and programmes and qualifications should be reformed on a large scale to be able to respond to the needs of society;
▶ Academics should be given a key role in this process;
▶ The reform process would require the development of internationally shared reference points / standards at disciplinary / subject area level;
▶ A language for communication should be developed which would be understood by all major stakeholders, that is academics, students, graduates, (potential) employers of graduates as well as professional organisations;
▶ All stakeholders, including (potential) employers and professional organisations, but in particular graduates should be (indirectly) involved in the process of curriculum design and quality enhancement;
▶ The focus should be on diversification of degree programmes by profiling and stimulating flexibility;
▶ The reform should facilitate national and international mobility and the recognition of periods of study, including qualifications for obtaining access to the next level of programmes.

The concept of ‘competences’ was chosen as a means to communicate with stakeholders despite awareness of the different meanings throughout the world attributed to the term. The basic idea was and is that the role of education is primary to make the student / learner more competent as a result of a learning process. Competences should be understood - according to Tuning - as a representation of a dynamic combination of cognitive and metacognitive skills, knowledge and understanding / insight, interpersonal, intellectual and practical skills, and ethical values. In other words, high level competences required to operate with confidence and success in a leading capacity in society.

Besides competences Tuning also introduced as part of its methodology the concept of learning outcomes. Learning outcomes in Tuning state the level of competences to be developed in a course unit, module or degree programme. From the onset Tuning made a distinction between general or generic competences and subject specific (disciplinary) competences. This was done to raise awareness about - in particular - the generic competences. Although they might have been taught already as part of a degree programmes, it was noticed at the time that in most cases they were not made explicit in the course material. This may have changed somewhat over time but it still seems to be an important point of attention in many degree programmes.

Having said this, a more far reaching conclusion was drawn from the analysis made; the reform process would require a paradigm shift in the teaching, learning and assessment process. As a consequence of focusing on the outcomes of the learning process in terms of the competences to be developed, the switch should be made from input
or content based learning to outcomes or output based learning. To phrase this differently: to move from a more staff-centred approach, based on the concept of learning objectives to a concept of student oriented learning, based on the concept of learning outcomes. A learning outcome is understood here as a statement of what a learner is expected to know, understand and be able to demonstrate after completion of a process of learning. A learning objective outlines the material the teaching staff intends to cover or the questions related to the discipline that the class will address. This last approach means in practice that the focus is on the teaching process (not the learning process) and on knowledge transfer of the academic staff member to the students. Student centred learning is an approach or system that supports the design of learning programmes which focus on the learners’ achievements, accommodates different learners’ priorities and is taking into consideration student workload (i.e. workload that is feasible within the duration of the learning programme). It accommodates for learners’ greater involvement in the choice of content, mode, pace and place of learning.

From this perspective the Tuning initiative developed two main action lines: a methodology or approach to develop current, high quality and relevant degree programmes for all levels, bachelor, master and doctorate, and internationally established reference points or subject area based competences frameworks to support the (re-)design, implementation, delivery and quality enhancement of degree programmes. This paper focuses on the latter.

**TUNING competences frameworks**

Having outlined above that degree programmes are no longer developed and delivered in its own right (only), but should be referenced against at least national, but preferably internationally established reference points, Tuning took the initiative for developing international competences frameworks. More or less at the same time the Quality Assurance Agency in the UK initiated the development of national based benchmark papers. In both cases a bottom-up approach was applied by making a group of academics responsible for setting-up a subject area based framework, respectively benchmark. The approach applied was slightly different, although in both cases it was based on a process of discussion and reflection. With respect to Tuning, lists of key competences were formulated which were used as a basis for consultation among stakeholders. For all subject areas / disciplines involved in the Tuning project at the time a common list of 30 generic competences was drawn up and for each individual subject area a subject area list of key competences was produced (see below). A consultation of stakeholders took place in 2001-2002 and was repeated in 2008 with more subject areas (9 instead of 7) and with a slightly adjusted and improved list of generic competences. The stakeholders consulted were academics, graduates, employers and in 2008 also students.

The consultation was based - in both cases - on different variables, that is, first:

- the degree of importance: the relevance of the competence, in the opinion of the stakeholder (for work in their profession);
- the level of achievement: the achievement of this competence as a result of having taken this university degree.

To evaluate these two variables, the respondents had to use a scale: 1 = none; 2 = weak; 3 = moderate; 4 = strong. In addition they had to:

- rank the generic and subject specific competences; based on the categorisation of the five most important generic and the five most important subject specific ones The competence that was ranked highest in the survey was allocated five points, four for the second and so on, with one point for the last in the selection. If the competence was not chosen in the survey, it scored zero points.

The outcomes of the consultation were very informative and revealing and were used as input for producing the Tuning competences frameworks for the subject areas involved. What was learned in particular was the need in society for graduates with better developed generic competences. Although doing well regarding the acquisition of knowledge of and insight into the subject area, students did less well than thought necessary in developing abstract thinking, analyzing and synthesizing skills (seen by the respondents as the most important competence) as well as
applying knowledge in practice, problem solving, learning abilities, and written and oral communication skills. Also more attention should be given – in particular according to employers and graduates - to leadership and teamwork competences in the learning process. Furthermore, it was advised to offer more attention to creativity and the development of an entrepreneurial spirit. These kind of consultations of the same stakeholder groups were repeated in other regions of the world with adjusted lists (based on the situation, culture and opinions of the academics of that region that drew up the lists) but with more or less comparable outcomes.

The consultations and competence lists were only one of the basic elements for developing the competences frameworks. Others were and are a description of the academic field country by country and synthesized at regional level – at first instance European and later also other regions of the world. Furthermore the typical degrees offered were mapped, as well as the social and professional needs, including the (potential) employability field, and if possible future trends were identified. The last items showed us differences between countries and higher education institutions. It was a confirmation that degree programmes have and should have different profiles, in particular at master level, and that they are also partly regionally bound.

Constructing subject area based competences frameworks

Since 2001 Tuning has built up much experience regarding the construction of (inter)national competences frameworks. As part of the European Tuning projects a large number of documents called Reference Points for the Design and Delivery of Degree Programmes in [name of subject area] - covering a large set of disciplines by now - have been developed. These were validated by external peers before being finalized and published. In the European context the term Reference points was used instead of, for example the term ‘standards’ (which is in use in Russia and Australia to express the same idea). Although Tuning intends to be standard setting, it wants to avoid at any cost that these reference points are regarded as prescriptive. In the dynamic world of today, they simply cannot be set in stone and therefore revision is needed every six years – preferably on the basis of a new consultation round of stakeholders. The development of the reference points was based on a common format, independent of the subject area involved.

For each competences framework to be developed, a group of 12 to 15 international experts was established. Its members were selected from different countries and were representing their higher education institution in the field involved. At its start Tuning focused only on traditional research oriented mono-disciplines to be able to develop its approach and to avoid unnecessary complications. At a later stage multi-disciplinary and interdisciplinary degree programmes as well as more applied disciplines were covered as well.

The process started with the mapping process as described above and the consultation of stakeholders. As was already explained, to prepare for the consultation process a common list of generic competences was drawn up by the different groups together and each group developed its own list of so-called key subject area competence statements which should ‘frame the subject area’. This last list contained on average of 25 statements. That list was established on the basis of a collection of ideas and expectations regarding degree programmes in the subject area based on an open discussion. On the basis of the long list consensus was sought on the short list (key competences) to be used for the consultation process. Because the focus was on competences to be covered in the discipline as a whole, the first cycle / bachelor and the second cycle / master were not distinguished. To allow for the consultation process, each university draw up a list of relevant employers for its field, a list of graduates which graduated within the last 3 to 5 years and a list of academics to be consulted. In the first European consultation round (2001-2002) no (mature) students were consulted. The Tuning America Latina project decided in 2004 to involve also students. This has been standard procedure since. After finishing the consultation process, its results were analysed by each subject area group. This led to a redefining of the original subject specific competences lists. The next step in the process was to design academic and professional frameworks for each of the cycles, preferably formulated as descriptors and to
build consensus on the most relevant competences for each of them, combining both what is common for academic recognition and what is different (the specific features). It was also checked whether the subject area frameworks were consistent with another recent development at the time, the establishment of a Qualifications Framework for the European Higher Education Area. This framework was endorsed in Europe by the ministers of Education in 2005 as part of the Bologna Process. Below we will come back to the complementary relationship of these so-called overarching or meta-frameworks and the subject area based frameworks.

The following step in the process of establishing competences frameworks was to check whether these frameworks allowed for a realistic student workload. In other words: to ‘measure’ the required student workload to achieve the desired competences levels, to be expressed in terms of credits (besides learning outcomes). This step was followed by an open discussion to identify the most suitable modes and approaches for teaching, learning and assessment to develop the competences identified. This reflection took place against the proposed change of paradigm – student centred learning – as described above. The outcome was an overview of different possibilities and strategies for each of the subject areas from which the academic staff responsible for offering the degree programmes in each institution could make their own selections and combinations. The last step was to relate the competences to consistent mechanisms for quality control and enhancement.

This approach can be visualized in the following image:

As part of the Tuning America Latina 2 Project (2011-2013) ideas about the development of competences frameworks were further refined and deepened, by introducing the concept of meta-profiles. Both reference points and meta-profiles have the aim to identify and describe the core (elements) of a discipline / subject area. However, the meta-profile approach is slightly different. While in the original approach the focus is on identifying the core or key competences, both generic and subject specific, in the second approach the focus is on the clustering of generic and subject specific competences to derive to so-called meta-competences. The interlinked groups of meta-competences then serve as the basis for defining a meta-profile (competences framework) which captures the essence of the discipline in more general terms. When the meta-profile is decided, it can be used as a template for constructing individual degree programmes.

This new method which was and is being applied later in Tuning projects in other regions of the world such as Russia, Africa and Central Asia, offers us a more sophisticated way forward, because the existing template of collecting a long list of competences and then boiling it down to the more essential ones, lacked sufficient structure.

Two main approaches have been developed for the grouping of competences. The first way is to cluster the most related competences in a feasible number of groups, minimum 5 to 8 maximum. After having done so, a label for each group is decided which reflects best its content and purpose. However, it is also possible to work the other way around, defining labels for the group’s first and then using these as a basis for clustering the competences. Each group or meta-competence will contain a mix of generic and subject specific competences. This is fully in line with the Tuning philosophy, which requires that these are developed together. Although named generic and transferable skills or competences these are or should always be developed in conjunction with the main field(s) of study, and not in isolation. To illustrate this point: the competence abstract thinking, analyzing and synthesizing in history is based on a different theoretical and methodological framework than for example the one used in physics or mathematics. This also applies - in general - for oral and written skills, leadership, teamwork, entrepreneurial spirit etc. because each discipline has its own academic culture and paradigm.
Competences frameworks are not only a fundamental basis and reference for the design of new and improvement of existing curricula, but also for the assessment and comparison of learning in a national and global context. Different levels of competences frameworks can be distinguished. While Tuning has focused on the subject area level, others, such as governments and quality assurance organisations, have initiated overarching or meta-frameworks which intend to cover all levels of learning or part of them. According to Tuning these complementary frameworks are a requirement for the restructuring and/or enhancing of the quality of higher education sector and its qualifications.

Qualifications Frameworks are not a new phenomenon, but in recent years their content and structure have changed. In principle every country has its qualifications framework or system which – in the past or still - describes the different types of qualifications offered, its interconnections, as well as pathways to progress within the system. This traditional model has been replaced now in many countries by a model which focuses on the outcomes of a learning process, be it formal or informal. As its basis descriptors are used which describe the expected / achieved learning in terms of competences and (indicated in terms of level) in learning outcomes. This new type of qualifications framework has been developed at transnational level (Europe) as well as national level.

At European level two meta-frameworks have been developed which are the Qualifications Framework for the European Higher Education Area (QF for EHEA), initiated as part of the Bologna Process, and the European Qualifications Framework for Life Long Learning (EQF for LLL). Both can be defined as an overarching framework that makes transparent the relationship between European national higher education frameworks of qualifications. The first one is based on the so-called Dublin Descriptors. These were developed in the same period in which Tuning developed its first competences frameworks (2001-2005). In March 2002 at the official Bologna seminar Working on the European Dimension of Quality, which took place in Amsterdam, the main conclusion was that general descriptors for the different cycles and reference points at subject area level should go together. To quote from its published report: “There is a widely-shared consensus that the ‘Dublin Descriptors’ , defining key outcomes for Bachelors and Masters programmes in general are useful. These generic descriptors are complementary to the more specific outcomes of the Tuning project (…), which have been developed at the level of areas of knowledge ( ‘disciplines’ )” . The Dublin descriptors outline the essential components of any degree programme that leads to the completion of a Bologna cycle, that is bachelor, master and doctorate. They are based on the following dimensions:

- Acquiring knowledge and understanding
- Applying knowledge and understanding
- Making informed judgments and choices
- Communicating knowledge and understanding
- Capacities to continue learning
- Contributing to original research (doctorate only)

Besides the ‘Bologna’ QF- EHEA, the European Union established the EQF for LLL which contains eight levels, covering learning achievements at all educational levels. As far as higher education is concerned, the top 4 levels (that is 5, 6, 7 and 8) in this Framework are compatible with the three cycles, plus the short cycle (Associated Degree), included in the Qualifications Framework for the EHEA. Hence, their outcomes correspond to those foreseen in the Dublin Descriptors, even though they are expressed in a slightly different language. The EQF distinguishes three main categories to order its outcomes based descriptors: knowledge, skills and competences. Competences should be read here as wider competences, intending to describe responsibility and autonomy.

Within each subject area, discipline or professional sector, the QF for EHEA Dublin Descriptors and /or the EQF for LLL level descriptors can be applied and adapted according to the specific way that learning is acquired in that sector. Thus the Dublin Descriptors/EQF descriptors form general reference points at the European level in which any specific Degree Programme can be situated.
In particular during the last five years, the development of competences / learning outcomes based National Qualifications Frameworks (NQFs) has gathered considerable momentum. To establish such a framework, each country sets out its own qualifications framework according to its educational structures and traditions. A National Qualifications Framework can be described as a single description, at national level or level of an education system (for example, Scotland, England and Wales), which is internationally understood and through which all qualifications and other learning achievements may be described and related to each other in a coherent way. Good examples are the qualifications frameworks of Australia, South-Africa, Thailand, a growing number of European countries, as well as a model developed by the Lumina Foundation for the USA. A national framework enables students, employers and quality assurance and accreditation agencies to ‘compare and contrast’ the learning achievements of students and benchmark (i.e. position) them according to other national and regional frameworks.

The Australian Qualifications Framework (AQF) distinguishes – with regard to the higher levels – the following items: purpose and volume of learning (expressed in time) besides the categories knowledge, skills, applications of knowledge and skills. The Lumina USA Degree Qualifications Profile is – as the QF for the EHEA - based on dimensions: Broad, Integrative Knowledge, Specialised Knowledge, Intellectual Skills, Applied Learning and Civic Learning. In this last model more emphasis is given to the learning process as a classifying principle.

All these overarching competences frameworks have in common that they are based on broad descriptors which are general by definition because they include all types and orientations (applied, research driven etc.) of qualifications or other learning achievements.

Bridging meta-competences frameworks and Tuning subject area based competences frameworks

In 2008 the Tuning management team concluded that it would be useful to bridge the gap between the meta-level and the subject area level by developing an intermediate framework level by grouping academic programmes in terms of domains or sectors. A sector or domain is understood here as a combination of related fields of study which are based on more or less comparable learning profiles. Tuning distinguishes six sectors: Humanities, Creative and Performing Disciplines, Engineering, Natural Sciences, Health Care and Social Sciences. The basic aim for developing Sectoral Qualifications Frameworks (SQFs) was and is to produce a common set of statements about expected achievement levels for students in any and all of the disciplines represented, as well as seeking to define what these disciplines have in common. The framework intends to show what sets them apart from disciplines in other sectors and give them their distinctive character as a sectoral grouping. At the start of the project period there was much doubt among the disciplinary experts whether it would be possible to find sufficient common ground. During the project these hesitations were slowly replaced by enthusiasm about the opportunities a SQF would offer.

From 2008 to 2010 a first project was implemented: the development of a Tuning Sectoral Qualifications Framework (SQF) for the Social Sciences. The sector was represented by the following academic areas: Business Studies, European Studies, Education Sciences, Occupational Therapy and Social Work, Law, Psychology and International Relations. The project designed a SQF which covers the higher education sector levels 5 to 8 as well as the preceding levels 3 and 4. Its development proved to be a pioneering and innovative experience, but most of all a major step forward towards linking the different initiatives – QF for the EHEA, EQF for LLL, NQFs and Tuning Subject Area based Reference Points, - so far.

The Social Sciences SQF-project was followed by a project in which two SQFs were developed, one for the Humanities and another one for the Creative and Performing Disciplines. While the Social Sciences SQF project group limited itself to define competences statements covering the broad categories of knowledge, skills and (wider) competences, the other two SQF project groups went a step further. Initiated by the Creative and Performing Disciplines dimensions were identified to organize the competences statements. The catalyst for generating mutual
belief in the validity of a joint sectoral framework came from the realisation of the central role played in the Creative and Performing Disciplines by the dimension of Creativity (making, performing, designing, conceptualising) in all of the disciplines represented. Also for the Humanities such a central dimension could be identified in the concept of the Human Being (human condition, experience and expression). In the framework of the OECD TUNING-AHELO project which aimed to develop a Conceptual Framework of Expected/Desired Learning Outcomes in Engineering (2009) a set of dimensions for organizing the competences was decided as well. It is intended to develop SQFs based on dimensions for the other three other sectors in the near future. The following table offers an overview of the dimensions identified for the three sectors covered so far:

**Table 1: SQF Dimensions**

<table>
<thead>
<tr>
<th>Humanities Dimensions</th>
<th>Creative and Performing Disciplines Dimensions</th>
<th>Engineering Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Human Being</td>
<td>Making, Performing, Designing, Conceptualising</td>
<td>Basic and Engineering Sciences</td>
</tr>
<tr>
<td>Cultures and Societies</td>
<td>Re-thinking, Considering and interpreting the Human</td>
<td></td>
</tr>
<tr>
<td>Texts and Contexts</td>
<td>Experimenting, innovating &amp; Researching</td>
<td></td>
</tr>
<tr>
<td>Theories and Concepts</td>
<td>Theories, Histories and Cultures</td>
<td>Engineering Analysis</td>
</tr>
<tr>
<td>Interdisciplinarity</td>
<td>Technical, environmental and Contextual issues</td>
<td>Engineering Design</td>
</tr>
<tr>
<td>Communication</td>
<td>Communication, Collaboration &amp; Interdisciplinarity</td>
<td>Generic Skills</td>
</tr>
<tr>
<td>Initiative and Creativity</td>
<td>Initiative &amp; Enterprise</td>
<td>Engineering Practice</td>
</tr>
<tr>
<td>Professional Development</td>
<td>Professional Development</td>
<td>Professional Development</td>
</tr>
</tbody>
</table>

The development of Tuning Sectoral Frameworks based on dimensions seems to be a breakthrough. They not only bridge the two European meta-frameworks and national frameworks globally but also the meta-profiles / reference points at subject area level. The three sectoral frameworks or profiles developed so far offer the necessary precision which is required for degree programme design, delivery, quality assurance and enhancement and the recognition of degrees and periods of studies.

The following table offers descriptors of the first two dimensions of the Tuning SQFs for the Humanities and the Creative and Performing Disciplines. It illustrates the kind of information about what is actually (or should be) taught and learned in the framework of degree programmes which belong to the sectors involved.

This information should be used as a reference for the designing of individual degree programmes. The expression ‘reference’ is used here to express that each degree programme will and should have its own individual profile. The profile of a degree is not only defined by the competences framework but also by the mission of the institution, the role foreseen for its graduates in society, the particular strengths of the department offering the degree in terms of particular expertise(s), as well as the financial means available. According to the philosophy of student centred learning and therefore that of Tuning, degree profiles should be developed by the team of academics involved in offering the degree / qualification. This team, which includes preferably one or more mature students, formulates the learning outcomes followed by the outline of the degree programme and its individual modules. This structure should
allow for covering the development and achievement of all learning outcomes defined in such a way that in terms of student workload the programme is realistic and feasible. For the development of new programmes and enhancement of existing programmes Tuning has developed a ten steps approach, which is included in the publication *A Tuning Guide to Formulating Degree Programme Profiles*. Bilbao, Groningen, The Hague, 2010.
Tuning competences frameworks: tools for comparison and assessment of learning

The AHELO project has shown us that a Tuning competences framework is the foundation of an assessment framework for allowing comparisons regarding the level of achievement of learning in a particular subject area / discipline in transnational perspective. An assessment framework measures the level of achievement of individual learners but also of a system. This system can be the department within an institution, a local or regional system or a national system. In combination with meta-competences frameworks and sectoral frameworks, the Tuning competences frameworks are also important tools for recognition of periods of learning as well as of qualifications.

With the exception of the TUNING-AHELO conceptual frameworks, which were developed in a global setting, all other Tuning competences frameworks were defined regionally: Europe, Latin America, Africa, Russia. We can observe that in the different Tuning projects two main types of meta-profiles, or combinations of these have been developed. The first one focusses on the learning process; the second one derives from the subject area or an academic sector. Both are based on dimensions.

The dimensions identified for the ‘learning process approach’ are for example: Cognitive Competences, Interpersonal Competence, Civic Competences, Practical Competences, Methodological Competences and Professional Competences. The dimensions of the subject area approach are reflected in the table above.

Both models can be used to develop and to position individual degree programmes according to their own profile. These can be visualized as spiderwebs against the background of agreed competences statements of the Tuning subject area based competences frameworks. Depending on its mission and role and its academic strengths the different dimensions identified are more or lesser developed than the levels indicated in the meta-profile. This web visualizes individual profiles of a bachelor, a master and a PhD-programme:

As was stated earlier, Tuning has developed competences frameworks already for a large number of subject areas. All these frameworks cover the first and second cycle of higher education, that is the bachelor and master, and often also the doctorate. It is important that frameworks are developed which cover more than one level. It forces the designers / experts to decide on the level of achievement that can or should be developed for each of the competences levels involved. It will guarantee that a good balance is defined. Below, an example derived from the subject area of Architecture, is offered. This academic field decided to develop its competences framework on the basis of the SQF for the Creative and Performing Disciplines, not the one for Engineering. The reason for this
choice was that the experts of the European Tuning subject area group of architecture identified themselves much more with the dimension of creativity than the dimension of modelling (typical for Engineering). Descriptors were developed for the levels 6 to 8 of the EQF, covering bachelor, master and doctorate.

Table 3: Levels of mastering competences, the example of Architecture:

<table>
<thead>
<tr>
<th>EQF Categories:</th>
<th>Knowledge</th>
<th>Skills</th>
<th>(Wider) Competences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architectural Creation &amp; Architectural Creativity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level/ Dimension</strong></td>
<td><strong>Grades of Architecture are expected at level</strong>*</td>
<td></td>
<td><strong>To be able to draw upon the knowledge and skills gained within their architectural studies so as to act and respond critically and creatively in different situations</strong></td>
</tr>
<tr>
<td>6 / Conceptualising, Designing, Materialising Architecture</td>
<td>To have advanced knowledge of the processes, concepts and cultural values guiding architectural creation</td>
<td>To have the advanced skills to formulate critically, elaborate creatively and translate innovatively into spatial forms their own architectural concepts</td>
<td></td>
</tr>
<tr>
<td>7 / Conceptualising, Designing, Materialising Architecture</td>
<td>To have highly advanced knowledge of the processes, concepts and cultural values guiding architectural creation, some of which will be at the forefront of their field</td>
<td>To have developed to a high professional level their ability as architects to formulate critically, elaborate creatively and translate innovatively into spatial forms their own architectural concepts</td>
<td>To emerge as well-developed personalities, able to draw upon the knowledge and skills gained within their architectural studies so as to act and respond critically and creatively in situations that are complex, unpredictable and require new strategic approaches</td>
</tr>
<tr>
<td>8 / Conceptualising, Designing, Materialising Architecture</td>
<td>Knowing in depth all the relevant methods and techniques of inquiry related to a particular field of study of architecture</td>
<td>Integrating previous experience so as to demonstrate original creative insights in the domain of architecture</td>
<td>Comprehending the transferability of their research capabilities to other fields</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Functioning with complete creative autonomy</td>
<td>Displaying professional, creative and scholarly integrity</td>
</tr>
</tbody>
</table>

From the above it can be learned that different phrasing is used to indicate each level. The Tuning experience has learned – as is also shown in the table - that level is indicated by complexity and scope.

When developing a measurement instrument it is not only very important to have agreement on the level of mastery or achievement. This level will be expressed in the form of a learning outcome for that particular competence. That learning outcome, however, will still be rather general, and therefore it is necessary to break it down in (for example 3) indicators. These indicators in turn are supported by (for example 5) descriptors or assessment criteria, which cover the level of attainment. This subdivision is made for each generic and subject area / disciplinary competence to be assessed / measured. Also in this respect, Tuning has developed examples of good practice. With regard to the generic competences the publication of Aurelio Villa Sánchez & Manuel Poblete Ruiz, eds., *Competence-based learning. A proposal for the assessment of generic competences*. Bilbao, 2008 is referred to. It is is also advised to consult this publication when developing or enhancing an individual degree programme.
Conclusion

This paper started with the assumption that the higher education sector is challenged by dynamisms of society, such as globalism and information and communication technologies, which have raised (global) competition between higher education providers in terms of outcomes of achievement and level of appropriateness of what has been learned by its graduates to serve today’s society/ies.

An analysis of the situation has shown that higher education institutions should take their responsibility to respond to the needs of society. These needs are most of all well-educated graduates which have the knowledge, skills and competences to operate successfully in that society. To educate such graduates a change of paradigm of the educational model in use is required, which means in practice a move from a staff-centred approach to a student oriented approach. This last model implies that stakeholders groups are involved in the design, delivery and enhancement of the quality of degree programmes. It also implies that a language is used which is well understood by all parties. Such a language has been proposed and developed by Tuning, by using the concepts of competences and learning outcomes. Furthermore, Tuning has outlined a methodology for (re-)designing degree programmes, and it has developed competences frameworks both at sectoral and subject area level, against which individual degree programmes can be referenced.

It has been demonstrated in this paper that not only these Tuning competences frameworks at subject area level (meta-profiles) but also overarching competences frameworks, such as the Tuning SQFs and the European and National Qualifications Frameworks, are essential instruments for reforming the higher education sector and its degree programmes. In particular the Tuning meta-profiles allow for comparison of (the level of) learning. Therefore, they form the backbone of any instrument for measuring learning at sectoral / subject area level in a comparative setting, locally, nationally, regionally and globally. Level descriptors and indicators as developed by Tuning allow for fair measurement of performance and comparison of learning.

Why is it important to compare learning achievements of programmes and systems with another? It will first and foremost show us their relative strengths and weaknesses. This is important information in today’s world, which requires transparency and accountability. It will offer the precise data to enhance degree programmes in a sophisticated way, by allowing us to focus on its weaker elements. When the performance indicators are linked - as part of the assessment instruments - to civil and employment needs and requirements, it will also make it possible to tailor the programmes better to those. However, it has to be stressed that according to Tuning, higher education institutions and their academics should remain responsible for content and form of their higher education programmes. Nevertheless, serving society best is a role of academics for which they can and should take full responsibility as the Tuning experience has shown us. The new approach as explained in this paper, will allow academia to show what high quality programmes are. It also allows to prove that it is doing its job well.
Section 1: Presentations

1. Global challenges

- Globalization
- Information and Communication Technology
- Competitive Internationalized Higher Education market
  - Academic staff / Researchers
  - Bachelor, Master and PhD students
  - Language of instruction
- Opportunities on the labor market
  - Role of economic crises

2. Searching for a new paradigm

Finding responses to the challenges (1):

- Higher Education structures and programmes and qualifications should be reformed at a large scale to be able to respond to the needs of society;
- Academics should be given a key role in this process;
- The reform process should require the development of international shared reference points / standards at disciplinary / subject area level;
- A language for communication should be developed which would be understood by all major stakeholders, that is academics, students, graduates, (potential) employers of graduates as well as professional organisations;

3. Searching for a new paradigm

Finding responses to the challenges (2):

- All stakeholders, including (potential) employers and professional organisations, but in particular graduates should be (indirectly) involved in the process of curriculum design and enhancement;
- The focus should be on diversification of degree programmes by profiling and stimulating flexibility;
- The reform should facilitate national and international mobility and the recognition of periods of study, including qualifications for obtaining access to the next level of programmes

4. Searching for a new paradigm

Agreed common language: “Competences” and “Learning outcomes”

Tuning definition of competences

- Competences represent a dynamic combination of cognitive and metacognitive skills, knowledge and understanding / insight, interpersonal, intellectual and practical skills and ethical values.
- Fostering competences is the object of an educational programmes.
- Competences are formed in various course units and assessed at different stages.
  
[competences are obtained by the student]
What is a Learning Outcome according to Tuning?

**Level** of competence is expressed in terms of Learning outcomes (required or expected achievement):

- Statements of what a learner is expected to know, understand and be able to demonstrate after completion of learning.
- They can refer to a single course unit or module or else to a period of studies, for example, a first, a second or third cycle programme.
- Learning outcomes specify the requirements for award of credit.

[learning outcomes are formulated by academic staff]

---

**Student centred learning:**

An approach or system that supports the design of learning programmes which focus on learners' achievements, accommodate different learners' priorities and are consistent with reasonable students' workload (i.e. workload that is feasible within the duration of the learning programme). It's accommodates for learners' greater involvement in the choice of content, mode, pace and place of learning.

---

**TUNING competences frameworks**

Consultation based on different variables:

- the degree of importance: the relevance of the competence, in the opinion of the stakeholder (for work in their profession);
- the level of achievement: the achievement of this competence as a result of having taken this university degree.

To evaluate these two variables, the respondents had to use a scale: 1 = none; 2 = weak; 3 = moderate; 4 = strong.

And second:

- ranking of generic competences: based on the categorisation of the five most important ones according to academics, graduates, students and employers.

---

**EUROPE 2001** | **EUROPE 2008**

<table>
<thead>
<tr>
<th>Academics</th>
<th>Graduates</th>
<th>Employers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.46</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td>0.94</td>
<td>0.78</td>
<td>0.70</td>
</tr>
<tr>
<td>0.92</td>
<td>0.89</td>
<td>0.70</td>
</tr>
</tbody>
</table>
1. Academics, Top 6

- Basic general knowledge
- Cap. for analysis and synthesis
- Capacity to learn
- Creativity
- Cap. for appl. knowledge in practice
- Critical and self critical abilities

2. Employers, Top 6

- Ability for abstract thinking, analysis and synthesis
- Ability to apply knowledge in practical situations
- Knowledge and understanding of the subject area and understanding of the profession
- Ability to identify, pose and resolve problems
- Capacity to learn and keep up-to-date with learning
- Ability to generate new ideas (creativity)

3. Employers: Importance ratings

- Capacity to learn
- Cap. for appl. knowledge in practice
- Cap. for analysis and synthesis
- Problem solving
- Concern for quality
- Teamwork

4. Ability for abstract thinking, analysis and synthesis
- Ability to identify, pose and resolve problems
- Knowledge and understanding of the subject area and understanding of the profession
- Ability to work in a team
- Ability to plan and manage time

5. Overarching competences frameworks

- Complementary instruments:
  - Meta-Qualifications Frameworks (EQF / QF for EHEA)
  - National Qualifications Frameworks
  - TUNING Sectoral Qualifications Frameworks: Social Sciences, Humanities, Creative and Performing Arts
  - TUNING subject area meta-profiles or reference points for large range of disciplines
  - Diploma Supplement (as an instrument to publish content and outcomes of degree programmes)

- Under construction:
  - TUNING Sectoral Qualifications Frameworks for other domains/sectors
Section 1: Presentations

Overarching competences frameworks

European Qualifications Framework for Lifelong Learning: Level 7 (Master)

Knowledge
- Highly specialised knowledge, some of which is at the forefront of knowledge in a field of work or study, as the basis for original thinking and/or research
- Critical awareness of knowledge issues in a field and at the interface between different fields

Skills
- Specialised problem-solving skills required in research and/or innovation in order to develop new knowledge and procedures and to integrate knowledge from different fields

(Wider) Competences
- Manage and transform work or study contexts that are complex, unpredictable and require new strategic approaches
- Take responsibility for contributing to professional knowledge and practice and/or for reviewing the strategic performance of learners

6. Bridging frameworks

THE COMPETENCES FRAMEWORK PYRAMIDE

TUNING Sectoral Competences Frameworks

TUNING Subject specific Competence Frameworks

Bridging frameworks

Structuring Competences and Learning Outcomes by using Dimensions

Examples of general categories / dimensions:

<table>
<thead>
<tr>
<th>Qualification Framework for the DEE</th>
<th>Lumina USA Degree Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquiring knowledge and understanding</td>
<td>Broad, Integrative Knowledge</td>
</tr>
<tr>
<td>Applying knowledge and understanding</td>
<td>Specialised Knowledge</td>
</tr>
<tr>
<td>Making informed judgments and choices</td>
<td>Intellectual Skills</td>
</tr>
<tr>
<td>Communicating knowledge and understanding</td>
<td>Applied Learning</td>
</tr>
<tr>
<td>Capacities to continue learning</td>
<td>Civic Learning</td>
</tr>
<tr>
<td>Contributing to original research (doctorate only)</td>
<td></td>
</tr>
</tbody>
</table>

Examples of sectoral / subject area dimensions:

<table>
<thead>
<tr>
<th>Humanities</th>
<th>Creative and Performing Disciplines</th>
<th>Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Human Being</td>
<td>Making, Performing, Designing, Conceptualising</td>
<td></td>
</tr>
<tr>
<td>Cultures and Societies</td>
<td>Re-thinking, Considering and Interpreting the Human Basic and Engineering Sciences</td>
<td></td>
</tr>
<tr>
<td>Texts and Contexts</td>
<td>Experimenting, Innovating &amp; Researching</td>
<td></td>
</tr>
<tr>
<td>Theories and Concepts</td>
<td>Theory, Histories and Cultures Engineering Analysis</td>
<td></td>
</tr>
<tr>
<td>Interdisciplinarity</td>
<td>Technical, environmental and Contextual issues Engineering Design</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Communications, Collaboration &amp; Interdisciplinarity General Skills</td>
<td></td>
</tr>
<tr>
<td>Initiative and Creativity</td>
<td>Initiative &amp; Enterprise Engineering Practices</td>
<td></td>
</tr>
<tr>
<td>Professional Development</td>
<td>Professional Development Professional Development</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sector at level 6 EQF</th>
<th>Knowledge</th>
<th>Skills</th>
<th>Competences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced knowledge of a field of work or study, involving a critical understanding of theories and principles</td>
<td>Advanced critical, demonstrating mastery and innovation, required to solve complex and unpredictable problems in a specialised field of work or study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage complex technical or professional activities or projects, taking responsibility for decision-making in unpredictable work or study contexts</td>
<td>Manage complex technical or professional activities or projects, taking responsibility for decision-making in unpredictable work or study contexts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have advanced knowledge of the processes and concepts underlying creation and/or performance in their specific discipline</td>
<td>Have the advanced skills necessary to create, realise and express their own creative concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appreciate how the practice and creation generated within their discipline both stems from, and shapes, our humanity</td>
<td>Have the advanced skills necessary to create, realise and express their own creative concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional competencies needed to operate with an ethical awareness and to encourage the development and the well-being of other individuals and groups</td>
<td>Be able to draw upon their knowledge and skills gained within their studies to act and respond ethically in different situations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bridging frameworks
Section 1: Presentations

Tuning Subject Area Meta-Profile publications ...

Other subject area brochures: Business Administration, Educational Sciences, Gender Studies, History, Mathematics, Nursing, etc.


Also published: Tuning AHELO conceptual frameworks for Economics and Engineering (first cycle).

Bridging frameworks

Degree programme profiles:
- Own features
- In accordance with Overarching / Meta-framework
- Based on Sectoral Qualifications Framework (if available)
- Based on Meta-profile / Reference points subject area / discipline

Bridging frameworks

Profiles can be based on two types of dimensions: the Learning Process and the Subject Area. Examples:

IDENTITY
- Gathers the essence of what is - "should be" - the degree holder.

FUNCTIONS
- Detects the occupations and tasks which can be carried out by the graduate.

CONTEXTS
- Focuses on the environment in which the graduate is able to function successfully.

EDUCATION
- Defines the main expected learning outcomes in terms of competences – generic and specific.

Bridging frameworks

Consequences for degree programme design and enhancement:

Degree programme based on the Tuning methodology:
- Programme based on profile, sets of competences to be obtained, desired learning outcomes to be achieved, student workload based credits to be awarded
- Programme design is team work, based on consultation, discussion, cooperation
- Learning outcomes / competences to be developed are basis for credit allocation
- Teaching, learning and assessment approaches respect credit allocation: feasibility key factor

Bridging frameworks

6. Assessment of learning: measuring performance

Comparing levels of achievement of learning:
- Locally: in and between classes in same institution
- Nationally: between institutions / external reviews
- Regionally: Limited number of countries, e.g. EU, East-Asia, Latin America, North America
- Globally, e.g. AHELO

Outcome: Reliable way to compare effectiveness of learning strategies and approaches in relations to level of outcomes (taking into account mission and profile of one own’s degree programme)
Section 1: Presentations

Measuring performance

Levels of Achievement: CRITICAL THINKING

- FIRST LEVEL OF ACHIEVEMENT: Asking oneself questions about surrounding life and actively participating in discussions on it, analysing the judgements made and reflecting on the consequences of one’s own and others’ decisions
- SECOND LEVEL OF ACHIEVEMENT: Analysing the logic of own and others’ judgements, weighing their personal and social implications
- THIRD LEVEL OF ACHIEVEMENT: Arguing the pertinence of judgements made and analysing the consistency of own behaviour, based on underlying principles and values

An example of a subject specific competence in the field of Architecture: Conceptualising, Designing, Materialising

<table>
<thead>
<tr>
<th>LEVELS OF MASTERY</th>
<th>INDICATORS</th>
<th>DESCRIPTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OECD-AHELO
7. Conclusions

- Overarching competences frameworks and Tuning meta-profiles are essential instruments for reforming the higher education sector and its degree programmes.
- Tuning meta-profiles allow for comparison of (the level of) learning.
- Tuning meta-profiles form the backbone of any instrument for measuring learning at sectoral / subject area level in a (trans)national setting.
- Level descriptors and indicators allow for fair measuring of performance and comparison of learning.
- Degree programmes require a facelift to meet the challenges of today’s and tomorrow’s world.

Thank you for your attention!

http://www.unideusto.org/tuningeu/
The purpose of this brief paper is to describe the design and results of a Feasibility Study of the Assessing Higher Education Learning Outcomes (AHELO) initiative conducted by the OECD in 2009-2012, and to draw some conclusions about what was learned through the Feasibility Study from the perspective of its Technical Advisory Group (TAG). The TAG was established early in the Study to provide technical guidance to the AHELO Feasibility Study and, based on this experience, to draw conclusions on the conduct of a possible Main Study.¹

Background and Early Development of AHELO.

OECD’s Directorate for Education began planning for an international assessment of learning in higher education in the spring of 2007. An initial “experts meeting” was held in Washington, DC in April of that year and involved a group of invited assessment and policy representatives from a half dozen countries. While discussions at that meeting remained general, it was agreed that such an assessment would have value for participating countries and that OECD should design and implement an exploratory study in a few countries in order to identify the challenges involved. This was followed by larger experts meetings in Paris in early July and in Korea in late October.

By the end of the third meeting, the basic shape and timetable of the AHELO initiative had emerged. The assessment would involve several subject matter domains producing results at the individual institutional level for benchmarking purposes. Like the PISA assessments already offered by OECD, they would prominently feature task-like production measures. Two components of the assessment would examine outcomes in selected fields of study with international relevance. After some discussion, Economics and Engineering were chosen as the primary subjects of interest. Another component of the proposed study would be a direct examination of “generic skills” like critical thinking and problem solving that might apply to all fields. One reason advanced for the latter was that “valid and reliable” instruments were already available in this area such as the U.S. Collegiate Learning Assessment (CLA). The experts recognized that such generic skills are widely viewed as critical for individual success, but that any assessment of them should properly be set in the context of particular fields of study. It was also recognized that many fields like history, literature, and law, are so culturally conditioned that international benchmarking would make little sense. Finally, recognizing the vast differences in the contexts for higher education across countries including curricular structure, educational values, and predominant pedagogies, a final important component would be a “Contextual Dimension.” This would consist of background surveys completed by faculty and institutional administrators, as well as a student survey to be administered at the same time as each cognitive assessment.

¹This paper’s author served as the Chair of the TAG and its contents draw heavily on reports of the TAG issued throughout the Feasibility Study’s history.
Before establishing the detailed design of the Feasibility Study, additional decisions needed to be made about AHELO’s purpose and how it should be governed. With respect to purpose, OECD argued that the need to align higher education outcomes in key areas across boundaries in a time of growing graduate mobility was becoming imperative. This is a major objective of the Bologna Process in Europe, and is also reflected in initiatives like the “Tuning Project,” which is trying to coordinate academic standards across institutions in different countries in multiple subjects. Reasons for individual institutions to participate reflect the same motivations as institutions administering national standardized assessments like the CLA: such information can be useful for strategic planning and supporting external quality assurance reviews, as well as for benchmarking performance on locally developed measures of student achievement.

With respect to governance, AHELO is located within OECD in the Institute for Management in Higher Education (IMHE), a unit primarily concerned with providing services to institutions in common areas like strategic planning and assessment throughout the world. Policy within this unit is set by the Education Policy Committee (EdPC), which with the OCED Secretariat, assumed broad responsibility for the AHELO initiative. To provide more detailed guidance to the initiative, a Group of National Experts (GNE) was created consisting of representatives of participating countries and jurisdictions, as well as “observer” countries that expressed interest in the initiative but did not participate in the Feasibility Study. Members of the GNE are a mix of policy and ministry representatives and individuals with expertise in international testing.

The AHELO Feasibility Study was originally envisioned to be a two to three year effort involving a small group of countries. The primary questions that this effort was designed to answer were: a) could valid and reliable assessment instruments be developed in multiple languages across quite different country contexts; b) could such an effort be managed politically through OECD’s complex multinational governance structure; c) could the testing process be effectively implemented through the use of country and institutional coordinators, assessment administration manuals, and local training programs and; d) would the resulting data be able to be analyzed and prove useful to participating countries and institutions?

**The Design of the AHELO Feasibility Study.** Each country participating in the Feasibility Study was asked to select a set of six to ten institutions chosen to reflect the diversity of that country’s higher education system. The sampling approach for each institution then involved choosing 200 students who were nearing the end of their three or four year period of tertiary study. Because the principal purpose of the Feasibility Study was to try out an assessment approach and draw conclusions about implementation, the resulting data were never intended to support comparisons across institutions or countries. To achieve maximum efficiency, all assessments were “spiraled” so that different students completed different parts of the whole. All students participating in AHELO completed a ninety minute to two hour assessment instrument in one of the domains of interest, plus the short survey designed to gather information about student backgrounds and educational experiences. All of the assessments were administered electronically through secure computer links. Constructed response task (CRT) items were scored by trained reviewers in each country using specially-designed rubrics.

The core of the Generic Skills component of AHELO was a modified version of the CLA, the task-based assessment administered by the New York based Council on Aid to Education (CAE). Representatives of CAE attended each of the initial planning meetings for AHELO and CAE was chosen as one of the prime contractors because its CLA assessment was at that time the only generic skills assessment that used a constructed response format. Two CLA task prompts were selected for further development after an invitational meeting in New York attended by prospective countries—“Lake to River,” which deals with the pros and cons of proceeding with a dam project with uncertain safety consequences and “Catfish,” which deals with ascribing the likelihood that an observed biological anomaly might be due to a commercial polluter. In addition to completing one of these tasks, each student also answered a short battery of multiple-choice questions (MCQs) on generic competencies drawn from the Graduate Skills Assessment
The AHELO Economics assessment was developed by the Educational Testing Service (ETS) using a combination of existing multiple-choice test items drawn in Economics and various international inventories of economic content knowledge including results of the Tuning project and Subject Benchmarks created for the discipline by the Quality Assurance Agency of the United Kingdom. Construction of the assessment was guided by an assessment framework created by a group of subject matter experts. This group also reviewed the assessment itself after draft items were developed. Rather than concentrating on strict content knowledge, the focus of the assessment was on students’ ability to “think above content” and use the concepts and language of the discipline effectively. The assessment itself consisted of a set of constructed-response tasks plus a multiple-choice battery drawn from re-worked items on the Economics GRE.

The Engineering assessment was developed by an international partnership of assessment development organizations led by ACER, and was guided by existing national competency frameworks, a Japanese engineering licensure examination, and the results of the Tuning project in Europe. Because Engineering consists of various sub-fields, a decision had first to be made about which should be examined. The panel of subject-matter experts chose Civil Engineering, largely because it is the most conceptually straightforward. In addition, a number of more general skills in Engineering shared by all sub-fields such as analysis and design, as well as basic scientific concepts underlying all of them, were assessed. Construction of the assessment was also guided by an assessment framework, constructed by the subject matter experts. Like the Economics assessment, a focus of the Engineering assessment goes beyond mastery of content to examine students’ ability to “think like an engineer.” The design of the assessment itself was influenced heavily by the Engineering licensing examination in Japan. Like the assessment in Economics, the Engineering assessment consisted of a mix of multiple-choice items and constructed response tasks.

Although the original plan for the Feasibility Study called for only five or six countries, a total of seventeen countries eventually participated in twenty-five fields—nine in Generic Skills, seven in Economics, and nine in Engineering. All countries participated in contextual data collection, which involved the ten minute student survey administered in conjunction with the assessment, a survey of faculty, a survey of department chairs (in Economics and Engineering) and a survey of institutional administrators. Additional contextual data included a range of descriptive materials about curriculum and each country’s higher education system assembled by a study coordinator in each country. Testing began in the spring of 2012 and was concluded at the end of May, with scoring and analysis completed by the end of the summer of 2012.

Creation and Role of the TAG. The need for an advisory body responsible for reviewing and upholding technical standards for the AHELO Feasibility Study was recognized by the OECD Secretariat and interested countries from the outset of the AHELO initiative. This need was affirmed by the three expert group meetings held in 2007 in Washington, Paris, and Seoul.

The TAG was formally established in 2010 with eight members drawn from assessment and policy experts from throughout the world. The TAG reported to both the OECD Secretariat and the AHELO GNE. The Terms of Reference of the TAG were specified broadly, but essentially established a role that consisted of a) reviewing draft materials on all aspects of the Feasibility Study and suggesting mid-course corrections, b) providing recommendations on the eventual conduct of an AHELO Main Study and, c) providing a definitive recommendation on the feasibility of AHELO at the conclusion of the study. A fourth responsibility was added in Phase II of the Feasibility Study when the TAG was charged with serving as the expert group for the Generic Skills strand and the Context Dimension. Finally, the Terms of Reference established that the GNE could call on the TAG for advice on technical “or other matters”—a charge that allowed the TAG to consider policy and implementation questions with increasing frequency as the Feasibility Study progressed.

The TAG met eight times in the course of the Feasibility Study, three of which were face-to-face meetings and
the balance conducted via teleconference. Most meetings consisted of updates on progress guided by a review of documents and covered all facets of the study including the development of assessment frameworks, instrument development, sampling approaches, country coordination, assessment administration procedures, scoring arrangements for CRTs, analysis plans, and reporting arrangements. Recommendations for mid-course guidance of the Feasibility Study were developed by the TAG in the course of these reviews. After each meeting, the Chair of the TAG drafted a report, which was then forwarded to the GNE and the Secretariat. The Chair of the TAG also met with the GNE after each face-to-face meeting to report on issues and recommendations, and the Chair of the GNE also observed the final face-to-face meeting of the TAG in October of 2012.

**The TAG’s Overall Assessment of the Feasibility Study.** The AHELO Feasibility Study constituted an unprecedented multi-national data collection effort at the higher education level. Data on student learning outcomes were collected in three domain strands in seventeen different countries or systems, using assessment instruments comprising both production-focused CRTs and forced-choice MCQs. Data were collected on a wide range of contextual factors by means of surveys completed by students, faculty members, institutional coordinators and national project managers. Numerous implementation challenges including translation, contextualization, sampling, electronic test administration, CRT response scoring, data cleaning, statistical analysis, and reporting were, for the most part, 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 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 a Main Study should be implemented.

That said, some things went particularly well in the AHELO Feasibility Study and a few did not go so well. Most have implied lessons for any AHELO Main Study.

**What Went Well.** 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, presented 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 coordinators, and the robust administration procedures established all contributed to success here.

- **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, 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 Coordination.** Management and coordination of an enterprise as complex as the AHELO Feasibility
Study involved massive challenges of maintaining consistent procedures across five continents, seventeen unique cultural-political contexts, and numerous time zones. The administrative arrangements that were put in place to run the Study 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, some aspects of the Feasibility Study did not go so well. As a consequence, they constitute areas that must be particularly examined as the initiative moves forward.

- **Resources and Time.** 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-build instruments, full-scale field trials of administration and scoring arrangements, and more time for de-briefing and collective discussion of obtained results.

- **Uneven Student Cooperation Rates.** The validity of any assessment depends in part on obtaining a sufficient number of students chosen as part of the sample at each institution to participate. Countries and institutions achieved substantially different levels of cooperation in the AHELO Feasibility Study. In some countries (Colombia and Mexico, for example) almost all students completed the assessments. In others (Norway and The Netherlands, for example) so few did so that obtained results were not valid enough to use. This mixed track record means that unusual attention to obtaining needed levels of student cooperation will be needed for a Main Study.

- **CRT Difficulty and Contextualization.** 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 them to be rectified.

- **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. By the time this situation was addressed by re-structuring contractual arrangements so that CAE was a subcontractor of ACER under the Consortium, past history meant that it was difficult to establish a true culture of partnership.

**Some Particular Lessons from the Feasibility Study.** Experience with the AHELO Feasibility Study offers additional lessons that should be taken forward for the AHELO Main Study:

- **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 contextualize instruments and administration procedures were admirable and, for the most part, successful, a more collaborative approach might have yielded even 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 institutional participants were somewhat disappointed by the lack of attention to their needs for information resulting from the study. Institution-level reports with more detailed breakdowns across student populations would have been beneficial, as well as more fully documented institutional and country data files. A project-wide Quality Monitor should also be established, as well as a National Quality Monitor for each participating country/system. This is consistent with international standards in conducting such studies.
• More information should be made available about the costs and benefits to countries and institutions of participating in AHELO. Two primary questions will probably be raised by any country/system considering whether or not to join an AHELO Main Study: “what is it likely to cost us?” and “what are we likely to learn?” Because the AHELO Feasibility Study is only just concluded, little can be said about the second question at this point. But some information about costs is available. The direct monetary costs of developing, adapting, and administering the various instruments are known through OECD contracting records. Many costs incurred by institutions and systems for such activities as sampling, student recruitment, test administration, scoring and data reporting, and coordination/oversight are similarly known. But many are not documented because they constitute less tangible costs, for example the time devoted to AHELO by institutional and system personnel. As a consequence, a systematic effort to collect data about both direct and indirect costs should be included in any future Main Study.

• Tools such as “Readiness Criteria” should be developed and put in place to allow potential participants and the OECD to determine whether institutions and jurisdictions can actually undertake and benefit from AHELO. An explicit set of country and institutional readiness criteria should be established to govern institutional participation in any AHELO Main Study. These criteria should include the provision of a student population sampling frame, sufficient computing infrastructure and IT personnel to support computer-based testing, commitment to participation in training, and effective internal management. It should also include a formal commitment to carry out Study protocols and to abide by the AHELO Technical Standards.

• Further work is needed about how to effectively assess Generic Skills in multiple disciplinary and national/cultural contexts. A major design choice for the AHELO Main Study is whether or not to include a dedicated Generic Skills strand. The existence of these competencies independent of discipline or field of study is a contested issue in the field of higher education assessment. Some generic competencies transfer relatively well across domains, other generic competencies are developed, applied, and assessed much more appropriately within the contexts of particular domains. Results of the Feasibility Study on Generic Skills CRTs suggest that these tasks might perform better if they were better contextualized. The two Generic Skills CRTs used in the Feasibility Study were contextualized to a “real world” problem-solving situation. However obtained results suggest that the manner in which these tasks were culturally situated and perceived varied substantially across countries and systems. How appropriate contextualization of Generic Skills should be accomplished in any future AHELO Main Study is still a matter for consideration. One option is to continue down the path of including “discipline-specific generic” components in each disciplinary assessment. This was done in Engineering in the Feasibility Study and, to some extent in Economics. If further development along these lines is pursued, these “discipline-specific generic” competencies should be more appropriately aligned with one another to ensure that they address some parallel content. If a decision is made to continue with a separate Generic Skills strand, the performance tasks might be situated in the context of broad disciplinary groupings like the sciences, social sciences, humanities and fine arts.

• To provide meaningful information for improving teaching and learning, a mix of item types is required in international assessments at the higher education level. Another design choice about instrumentation is whether production based CRTs should be included in an AHELO Main Study at all. Decades of research have shown that CRTs will never perform as well in terms of reliability as a battery based solely on MCQs. Results of the Feasibility Study confirm this conclusion for all three domains. The question for an AHELO Main Study is whether the use of CRTs adds enough validity to be worth this inevitable price in lost reliability. On this question, results of the Feasibility Study in Engineering suggest that some of the most important information that could drive improvements in teaching and learning was obtained through the CRTs. The major drawback of including CRTs is substantially increased costs. If the main purpose of AHELO is held to be instructional improvement, the inclusion of CRTs will undoubtedly increase the usefulness of results. On the other hand, if the
main purpose is to provide the most reliable international benchmarks of institutional performance with respect to student learning outcomes, the greater reliability and lower cost of adopting an approach based solely on MCQs may be preferred.

- **An acceptable response rate based on an accurate probability sample is required to assure comparability of results across institutions.** In the Feasibility Study, probability sampling or a census of students was used by almost three-quarters of participating institutions. For the remaining institutions, it is not apparent why such a sample was not used. For the Main Study, participating institutions should be required to compile a list (or lists) of eligible students (or groups of students) and to use either probability sampling or a census. That said, there should be some flexibility regarding the choice of probability sampling method. For example, cluster sampling of class groups may be reasonable when the number of eligible students is large. It may also be reasonable for AHELO to impose a fixed minimum response rate threshold, at the level of the country or the institution, below which data will be excluded from the data analysis. Finally, measures to increase response rates should be actively researched before any new AHELO data collection.

- **AHELO should be better located and integrated with the international scholarly community examining student learning outcomes and the policies and practices that support better learning.** The past decade has seen a sharp increase in policy and scholarly interest in improved 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 U.S., the rise of qualifications frameworks in many nations, and the emergence of multinational 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 AHELO Feasibility Study experienced serious resource shortfalls which, in the course of implementation, negatively affected many of its components. This occurred incrementally and its effects were complicated by the fact that the project included more countries than a “feasibility study” should probably have included. A similar under-resourced condition cannot be allowed for a Main Study. The OECD and participating countries will need to ensure adequate resources in moving forward. If this cannot be guaranteed, implementation will have to wait until it can.

**Moving Forward?** The OECD has concluded that the results of the Feasibility Study were sufficiently positive that a Main Study should be conducted and has already distributed a paper describing its main features and inviting country participation. This verdict by no means assures that AHELO will become more broadly operational, however. The real question that will govern moving forward is not whether or not it is possible to conduct these assessments. It is instead whether or not the effort is cost effective. The cost side is by this point readily apparent. The AHELO Feasibility Study cost more than nine million Euros to implement, most of which was borne by participating jurisdictions and institutions. Benefits to participants, on the other hand, have been mixed—largely depending upon the amount of effort jurisdictions and institutions invested in undertaking local data analyses and disseminating the results. Consistent with OECD’s hopes for a Main Study, participating institutions may find internationally benchmarked assessment results helpful in their strategic planning efforts. Participating countries, meanwhile, will be able to rehearse how various kinds of results can be used to evaluate higher education performance. Whether either of these actors, as well as a host of non-participating institutions and nations, will come to believe that full implementation is worth the investment remains to be seen.
The AHELO Feasibility Study: Study Results and the Conclusions of the Technical Advisory Group (TAG)

Peter T. Ewell
National Center for Higher Education Management Systems (NCHEMS)

NIER Conference
Tokyo, Japan
December 10, 2013

Background and Early Development of AHELO

- Parallel Efforts in PISA and PIAAC
- Domains: Economics, Engineering, and Generic Skills—As Well as a Contextual Dimension
- Decision to Launch a “Feasibility Study”

Design of the AHELO Feasibility Study

- Seventeen Countries (Jurisdictions)
- Six to Ten Institutions per Country
- Two Hundred Students per Institution who are Nearing the End of their Academic Programs
- Ninety Minute to Two Hour Testing Period
- Mix of Multiple Choice and Constructed Response
- Test Items “Spiraled” Across Test Takers

The AHELO Generic Skills Assessment

- Administered in Nine Countries
- Two Constructed Response Prompts Based on the Collegiate Learning Assessment (CLA)
- Multiple-Choice Items Drawn from the Australian Graduate Skills Assessment
- Contextualized and Translated Under Direction by Each Country
- Scored Locally in Each Country

Examples of Generic Skills

- Cognitive:
  - Communications (Written and Oral)
  - Critical Thinking and Problem Solving
- Non-Cognitive
  - Empathy and Ethics
  - Working with Others
  - Taking Responsibility

Graduate Attributes: Top Choices for Employers

Percent Reporting “Very” and “Somewhat” Important:
- Ethical Judgment and Integrity – 96%
- Comfortable with Diverse Backgrounds – 98%
- Demonstrated Capacity for Professional Development – 94%
- Interest in Giving Back to Communities Our Company Serves – 71%
Teaching Generic Skills: Common Approaches

- A “General Education” Block of Courses
- Embed These Skills in Disciplinary Teaching
- Particular Pedagogies:
  - Collaboration and Group Work
  - Internships and Fieldwork
  - Capstone Courses and Projects
  - Role of the “Co-Curriculum”

The AHELO Economics Assessment

- Administered in Seven Countries
- Constructed Response Prompts Based on Tuning and UK Subject Benchmarks in Economics
- Multiple-Choice Items Based on the ETS Graduate Record Examination in Economics
- Focused on Student Ability to “Think Above Content”
- Scored Locally in Each Country

The AHELO Engineering Assessment

- Administered in Nine Countries
- Confined to Civil Engineering
- Constructed Response Prompts Based on Tuning
- Multiple-Choice Items Based on a Japanese Engineering Licensure Examination
- Focused on Student Ability to “Think Like an Engineer”
- Scored Locally in Each Country

The AHELO Contextual Dimension

- Designed to Gather Background Information in Terms of Which to Explain (and Adjust) Cross-Country Variation
- Elements of the Contextual Dimension
  - Student (Test Taker) Survey
  - Faculty Survey
  - Institutional Background Survey
  - Country Background Information

Some Findings

- All Assessments Performed at or Above Expected Levels of Reliability
- Constructed Response Items Less Reliable than Multiple-Choice Items [as Expected]
- Constructed Response Items Clearly Tapped an Independent Dimension of Student Knowledge
- Some Evidence of Convergent Validity for All Assessments

The Technical Advisory Group (TAG)

- Charged with Reviewing the Technical Adequacy of All Aspects of the AHELO Feasibility Study:
  - Assessment Frameworks and Instruments
  - Sampling and Test Administration Procedures
  - Analysis and Reporting
- Charged with Making Final Recommendations on Feasibility and the Future Conduct of the AHELO Main Study
The TAG’s View: What Went Well

- Assessment Administration
- Technical Aspects of the Data Analysis
- Instrument Design for Purpose-Built Instruments
- Overall Coordination

The TAG’s View: What Did Not Go So Well

- Resources and Time
- Uneven Student Cooperation Rates
- Contextualization and Difficulty of Constructed Response Tasks
- Contractual Arrangements

Lessons for the AHELO Main Study

- More Opportunities for Stakeholder Participation in Assessment Design and Analysis
- Full-Scale Try-Out of Instruments and Administration Procedures
- More Up-Front Information on Costs and Benefits
- Development of Country/Institution “Readiness Criteria” to Assist Decisions to Participate
- More Work on the Assessment of Generic Skills

Lessons for the AHELO Main Study (Continued)

- Continue Mix of Item Types (Multiple-Choice and Constructed Response)
- Use Census or Probability Sample for All Institutions
- Better Integrate AHELO into International Discussions of Quality in the Scholarly Community
- All Will Require More Time and Adequate Resources

Moving Forward?

- OECD Has Concluded that AHELO is Feasible and is Planning to Move Forward
- Country and Institutional Decisions About Participating will Depend on Individual Decisions About Costs and Benefits
- Cost Side Known, Benefits Not Yet Assured
- We Will Know within a Year Whether and How AHELO Will Move Forward
This is the last presentation before the break. My name is Kishimoto. I’m from the Tokyo Institute of Technology. I will talk about AHELO and its prospects in Japan. As has been discussed, our country has participated in the engineering field, and we have conducted various activities in regard to that. First, I will give you an overview of what we have been doing in the engineering arena. Second, I will give you an idea of the kinds of questions that were created for the examination. And third, I will talk about the significance and prospects of our participation.

First, here is the outline of the initiative. As Dr. Ewell mentioned, its purpose was to measure university education learning outcomes worldwide through a common test. The final report is being translated, and is expected to be published in March 2014; the original version can be downloaded from the site that you see on the slide. Regarding what is being measured, from the perspective of Tuning AHELO, the framework was created and broken down into the questions on the test. A similar methodology was applied to both economics and engineering. This is the structure of Japan, this is AHELO and the national center, and validation was done. It was conducted under the framework of the university reform initiative of MEXT, and my university, Tokyo Institute of Technology, served as its representative. The AHELO consortium was commissioned by the OECD to carry out this project, and we utilized the test – We asked the students to solve the problems on the test and analyzed the results. The AHELO consortium created the actual test. Therefore, we did not merely participate in AHELO, but have worked on the creation of the test and were thus able to get an overview of the initiative. The AHELO initiative had two phases: creating the test and conducting it in a small scale. The areas of improvement and scoring criteria for the test were selected. In the second phase, its scientific feasibility and practicality were evaluated, a final meeting was held in March 2013, and the report was produced.

We now go to the engineering arena. What was the framework for the creation of the tests? On May 4 and 5, 2009, experts met in Brussels to discuss the topic. As you can see, the agenda included nine items. The overall discussion concerned the engineering domains. What kind of specialized academic domains are there? What are the major degrees and occupations? The report was created under the Tuning framework. Agenda item number 5, the definition of expected and desired learning outcomes, was very important. What, then, is a learning outcome? As Prof. Wagenaar has said, this is the definition of the learning outcome; specifically, how do we select it? This was the challenge that was given to us. We decided to select the learning outcomes as follows. We decided to use ABET (Accreditation Board for Engineering and Technology) engineering criteria 2000. These are the criteria used for accreditation in engineering areas, so we decided to utilize the framework there. This is the equivalent of the Washington accord criteria; in Japan, it’s the Japanese criteria – JABEE: Japan Accreditation Board for Engineering Education. In Europe, EUR-ACE was utilized. This is also an accreditation organization and they have learning outcomes. We referred to the outcomes of the first cycle of bachelor degrees.
Because this was a feasibility study, everything was considered equal, and with the framework, we made some considerations. The Tuning-AHELO conceptual framework was thus derived; this chart shows the conceptual framework of the EUR-ACE used in Europe on the left-hand side. As you can see, the standards are categorized into knowledge and understanding, engineering analysis, engineering design, investigations, engineering practice, and transferable skills. As for ABET, if the criterion deals with the ability to apply the knowledge of mathematics, science, and engineering, it was categorized as a) to correspond to the categories of EUR-ACE. We decided to select the common parts, and a Tuning-AHELO framework was set up, as you can see on the right-hand side. After identifying the common areas, we decided to pick several of them in creating the scheme. Our selection included the basic engineering sciences, engineering practice, engineering design, and engineering analysis. These are engineering skills, and they were included in the Tuning-AHELO framework of learning outcomes.

The consortium’s work was to create a test from the framework. This is how we did it. The generic skills in engineering are required across the board, and above them are the engineering fundamentals. In civil engineering, we put in the specific skills needed, and above that, engineering analysis, engineering design, and engineering practice were put in place.

For this area, the test we created was composed of multiple-choice questions to examine the level of basic knowledge and skills. On the other hand, for the areas related to activity, constructed response questions were created, as mentioned in the earlier presentation. What I mean is that the students would select 25 out of 30 items and solve one of the three items for the constructed response tasks. I would like to give you an idea of the specific questions given. We developed the questions based on the conceptual framework, and the consortium prepared the draft. As stated in the previous presentation, the multiple-choice questions were proposed by Japan. We had with us the provisions of the first professional engineering examination given by the Institution of Professional Engineers of Japan, and we translated it. Also, within the Japan Society of Civil Engineers, there is a civil engineering licensing examination to test the ability of those who graduated from the civil engineering department. We translated that as well and made a proposal to the consortium. We extracted the questions which were appropriate for AHELO. On the other hand, the constructed response tasks were proposed by Australia, and the appropriate questions were selected.

At the meetings of the international experts, the proposal from the consortium was submitted. They chose from the proposed questions and reviewed the selection to raise them to the internationally accepted level. The Japan-specific questions drawn from the professional engineering examination and civil engineering licensing examination were excluded from the common test, as these would not be appropriate for such. The areas that are not taught so much in other countries were also excluded. For example, some countries said they did not need countermeasures for earthquakes because they did not have any earthquakes. If the meetings had been scheduled after the Great East Japan earthquake, we could have made a strong proposal. At that time, however, countermeasures for earthquakes were considered irrelevant and, therefore, excluded. Also, there were countries that did not have railways, so the railway questions were considered inappropriate for them. It was very difficult to select the appropriate questions for the test. To illustrate specific questions, the multiple-choice questions look like this. This is called a truss; when there is a framework and a certain load is put on it, what kind of power would have to be applied? The students would pick the answer from choices A, B, C or D. The professional engineering examination has five choices, but we decided to cut the number of choices to four, even though it would alter the degree of difficulty.

We have a lot of experience in the creation of these types of questions. Through these questions, we are able to determine whether the basics have been acquired by the students. This is an example of the CRT (constructed response task) questions. On the top part is the address of the website; if you download that, you’ll be able to read the original text. There’s Hoover Dam, and this question is related to it. We posted a photograph of the proposed site of the dam. A sketch or plans of the proposed reservoir were also provided. This is the set of questions. The questions were not limited to four; there were more.

The first question is “Explain why this is a good dam site for hydroelectric power generation,” and it asks you to
discuss at least two aspects. There are various answers to this, and the person who will score the test will check whether the answer is appropriate, half-scored, or unacceptable. The second question is “Explain the two main design features that contribute to the structural strength and stability of the Hoover dam.” There are various other questions; for example, “In the midst of constructing this dam, you found a defect. How would you respond to that?” These types of questions help a future professional engineer determine whether he or she has learned enough to be able to respond to such problems appropriately. The details of each question revealed the learning outcomes that we wanted to assess. For the first example, it was the features that were suitable for a dam that would be used to generate hydroelectric power. To come up with the appropriate answer, they will need to have sufficient professional knowledge and the necessary competencies. The aspects range from A to F; it’s not just one. Therefore, if they write one aspect, that’s one point; two aspects, two points; and so on. The learning outcome assessed by the second question was the ability to determine which features contributed to the structural strength and stability of Hoover Dam. Two of several possible answers are “arch shape” and “robustness of the material in the canyon walls.” I forgot to mention this earlier in the first question: here it is written that the dam location will have minimal social impact. Well, you might answer that there are no residents nearby, but such a response was not initially included in the model answer. Many of the Japanese students wrote similar answers to this, so we incorporated that as a correct answer, too. We needed to discuss all the areas—the kinds of answers or what the students will think of when answering these questions. Next, we asked the students what they thought about the multiple-choice questions. They said most of the questions were about what they had learned at the university, and that it was not difficult to solve them. They also said the questions were one-shot; they appeared to be wide and shallow, and could use more depth. But does this reflect the actual score outcome. I think the impressions of the students and the actual results were not the same. Regarding the constructed response test, in the first phase, we had the students write their answers on paper, and they wrote a lot of sentences. In the second phase, we used the computer, and the answers were very short and abrupt. So, the content of the response was different when they used a pencil or a computer, and I think that’s another issue that we need to tackle. But first of all, the Japanese students said that they found it good that the questions in the first phase of the constructed response test concerned a practical issue, as they often deal with theoretical questions at the university. They also said that they enjoyed the ethical questions, too. Other comments said they found it interesting that it was a well-thought-out question that “made us think of the cause and then respond to it. I think it will be useful in practice in the future.”

Other comments were that perspectives such as those adapted to CRT-type questions were important. One student said, “I feel we should have more group discussions and case studies in our classes, meaning that we haven’t had the opportunity to learn them before, so we would like to learn these approaches to problem solving better.” The result of the feasibility study is one thing, but that the students showed this type of response is an important element when we think about the future education of the students. I would like to briefly summarize the significance of participation. First of all, we were able to participate in the discussion regarding the framework. And in our country, we have an accreditation institution for licensing engineers. Thus, by participating, we could express our views to make the framework consistent with the criteria used for licensing engineers and actual conditions in the educational arena in Japan. We will also be able to express the thoughts behind the creation of such questions in the international arena. In the consortium of colleges or within the expert meeting, our opinions were heard with respect. The reason for this is that Japan’s engineering is respected globally and is supported by the graduates who have become engineers. So, regarding engineering education, we found out that there is a strong interest in our country; therefore, we ought to contribute more in these areas. Furthermore, we were able to work not just with the engineer experts, but with education experts as well. Thus, we realized that it is important to not limit ourselves to working only with engineers; we should also work with experts in other fields, such as education. We participated in the engineering domain of the feasibility study, and we feel that it is necessary to explain and pass on the findings, achievements, and insights to other subject areas. Regarding the multiple-choice questions, we found out that the questions Japan proposed were
at the internationally accepted level. We tried a question that was made somewhere else (not in Japan) and it was not used. So, the questions that our colleagues have been making for long years have been proven good for assessing elsewhere what we wanted to assess here in Japan. Regarding the constructed response test, there were difficulties in developing questions and in scoring. The questions will be translated into Japanese. Since the language frameworks of English and Japanese are different, are we going to translate it as it is? In Japan, we don’t express the subject that much, but if we don’t express the subject within the sentence, let’s say for an ethical question, what happens if there’s a defect in the dam? If no subject is mentioned, the answers will be derived from somebody else’s opinion, and not from the engineers.

That is one of the experiences that we learned from. Regarding the constructed response task, the ratio of students who got the right answer has a correlation with the ratio of students who scored high in the multiple-choice questions. But I think we can conclude that combining the two types of questions is quite effective. I would also like to touch on AHELO’s potential impacts. We were able to reconfirm Japanese engineering education from a global perspective. We need to maximize AHELO’s education improvement effect, so that by disclosing all this information, such as the questions and scoring criteria, we will be able to maximize its global effect. And let me reiterate that in the future, I think these efforts need to be continued in various places. However, as pointed out previously, we also need to devise a way to make this sustainable from the perspective of manpower and costs. In other words, let’s move this forward gradually without rushing into getting results. I have run out of time so I will have to skip my summary. But I have put in bullet points of what I have been speaking about on my slide. That is all. Thank you for your kind attention.
**Section 1: Presentations**

### AHELO in Japan and Its Prospects

Kikuo Kishimoto
Dean, Graduate School of Engineering, School of Engineering at the Tokyo Institute of Technology

AHELO Engineering Expert Group

1. Task Overview (Engineering)
2. Questions for Examination
3. Significance of Participation
4. Future Prospects
5. Summary

### OECD-AHELO Assessment of Higher Education Learning Outcomes

- **Purpose:** To measure university education learning outcomes through a common test around the world.
- The feasibility study implemented between 2008 and 2012 intended to determine whether it was possible to globally assess learning outcomes.
  - Implemented for subject areas (Engineering and Economics) and generic (= background survey)
  - Japan participated in the field of Engineering (12 universities, 504 students, 196 university teaching staff)
  - Final Report (Japanese edition is scheduled for publication in March, 2014)
- The employed competency framework was Tuning-AHELO (Engineering and Economics)

### OECD-AHELO Feasibility Study

The Experimental Study to Assess Feasibility - 3 Phases

- **Phase 1:** January 2010 – June, 2011
  - Initial proof of concept (Qualitative evaluation)
  - Development of assessment tool (measuring instrument + test) to draw reliable conclusions regarding learning outcomes while considering the diversity and specificity of each country?

- **Phase 2:** December, 2011
  - Scientific feasibility and proof of practicality (Reliability assessment – quantitative evaluation)
    - Is the assessment tool valid and reliable?
    - To what extent possible test university students for their participation and implement assessment appropriately?
    - Large-scale field implementation and scoring

- **First meeting:** March, 2012
  - Whether to move AHELO forward with full-scale implementation is examined based on the results of (1) and (2)

- **3 areas (Generic Skills, Economics, Engineering + background)**

[http://www.oecd.org/document/22/0,3746,en_2649_39263238_40624662_1_1_1_1,00.html](http://www.oecd.org/document/22/0,3746,en_2649_39263238_40624662_1_1_1_1,00.html)

### Management Structure - Japan

- **Secretariat**
- **Decision-making Group of National Experts**

- **AHELO National Center (NIEER)**

- **International Council**

- **Central Council**

- **Panel of Experts**


- **Representative:** Tokyo Institute of Technology

- **Central Council for Education University Sub-committee Meeting AHELO – WSG**

- **Februrary 2008 – March 2008, August 2012 – March 2013**

### Tuning in AHELO Engineering

Main agenda matters at Experts’ Meeting in Brussels (May 4 – 5, 2009)

1. Academic/specialized domains of engineering
2. Major degrees in engineering
3. Major occupations
4. Categorization and ordering of learning outcomes statements
5. Definition of expected/desired learning outcomes
6. Expected/desired learning outcomes in each subject area
7. Definition of level indicators
8. General statements for engineers
9. Approaches to learning, teaching, and assessment
Section 1: Presentations

What are Learning Outcomes?
"Learning outcomes are statements of what a learner is expected to know, understand, and/or be able to demonstrate at the completion of a process of learning."

AHELO - Selecting learning outcomes in engineering
Selection will be done referring to ABET Engineering Criteria 2000 (member countries of Washington Accord use the same criteria while Japan uses JABEE Criteria) and EUR-ACE Learning Outcomes for First Cycle Bachelor Degrees.

=> “A Tuning-AHELO Conceptual Framework of Expected Desired Learning Outcomes in Engineering” OECD Education Working Paper No.60 (http://dx.doi.org/10.1787/5kgqthcn8mwn-en)

List of Tuning AHELO Key Concepts in Engineering

Development of Questions based on the Conceptual Framework

- The Consortium prepared the draft
  - Japan proposed multiple-choice questions (MCQs)
    - First-Step Professional Engineer Examination by the Institution of Professional Engineers, Japan
    - Civil engineering licensing examination of Japan Society of Civil Engineers
  - Australia proposed constructed-response tasks (CRTs)

- International experts’ meeting:
  - Makes a choice from proposed questions
  - Reviews selected questions to improve them to the internationally accepted level

A Tuning-AHELO Conceptual Framework of Expected Desired Learning Outcomes in Engineering

<table>
<thead>
<tr>
<th>Desired Learning Outcomes</th>
<th>AHELO</th>
<th>EUR-ACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge and Understanding</td>
<td>(a) Multiple-choice questions, analyze and interpret data, apply knowledge to new situations and other engineering problems</td>
<td>(a) Multiple-choice questions, analyze and interpret data, apply knowledge to new situations and other engineering problems</td>
</tr>
<tr>
<td>Engineering Analysis</td>
<td>(b) Ability to identify, analyze and solve engineering problems</td>
<td>(b) Ability to identify, analyze and solve engineering problems</td>
</tr>
<tr>
<td>Engineering Design</td>
<td>(c) Apply basic principles of design and innovation to real-world problems</td>
<td>(c) Apply basic principles of design and innovation to real-world problems</td>
</tr>
<tr>
<td>Investigative Skills</td>
<td>(d) Ability to work in a team and contribute constructively</td>
<td>(d) Ability to work in a team and contribute constructively</td>
</tr>
<tr>
<td>Engineering Practice</td>
<td>(e) Ability to lead, manage, and work in teams</td>
<td>(e) Ability to lead, manage, and work in teams</td>
</tr>
<tr>
<td>Transferable Skills</td>
<td>(f) Ability to learn independently</td>
<td>(f) Ability to learn independently</td>
</tr>
</tbody>
</table>

2. Questions for Examination

An Example of MCQ

A load P is applied to a Warren truss as shown below. If the self-weight of the members is ignored, which of the following statements is correct?

- Compressive force exists in both the upper-chord member (p-q) and the lower-chord member (r-s).
- Tensile force exists in both the upper-chord member (p-q) and the lower-chord member (r-s).
- Compressive force exists in the upper-chord member (p-q), while tensile force is applied to the lower-chord member (r-s).
- Tensile force exists in the upper-chord member (p-q), while compressive force is applied to the lower-chord member (r-s).

Focused competencies: engineering systematic understanding of the key aspects and concepts of major branch of engineering

* MCQs were reviewed to meet the internationally accepted level based on civil engineering licensing examination of Japan Society of Civil Engineers and First-Step Professional Engineer Examination by the Institution of Professional Engineers, Japan.
Section 1: Presentations

Scope of Learning Outcomes and Definition of Standards
An Example of CRT


The Hoover Dam is a 221-meter high concrete arch-gravity dam in the Black Canyon of the Colorado River in the United States of America. It was built to provide irrigation water, to control floods, and to provide water for a hydroelectric power station at the base of the dam.

1. Explain why this is a good dam site for hydroelectric power generation. You should discuss at least two aspects.
2. Explain the two main design features that contribute to the structural strength and stability of the Hoover dam.
3. The maximum electrical power generated by the turbines at the Hoover Dam is 1.745 × 10^9 W. What is the approximate amount of water that flows through the turbines at this power output? If the power station operates at 96% efficiency? (Note: \[ V \times A \times 10^3 \times \eta = P \])
4. Imagine that a new dam is being planned today in a different location. Briefly explain two environmental effects of the dam (which could also be upstream or downstream) that an engineer would need to consider in an environmental impact statement.

Scope of Learning Outcomes and Definition of Standards
An Example of 2nd Question

- Learning outcomes to be assessed
  - Understanding of features that contribute to the structural strength and stability of the Hoover dam.
- Background competencies:
  - Engineering Design (an ability to understand and apply methods to designs systems, components, and processes to satisfy requirements within the realistic constraints)
- Answering aspects
  - a) Arch
  - b) Material in the canyon walls must be robust
  - c) The heavy weight of concrete
  - d) Tapered shape of concrete wall/low center of gravity
  - e) Spillways and/or tunnels are equipped

Excerpt student comments (from the first phase)

[Multiple-choice question (MCQ)]
- Most of the questions were about what we learned at universities, and it was not difficult to solve them.
- All questions were one-shot, thus appeared to be wide and shallow. Questions can be more in-depth.

[Constructed-response task (CRT)]
- I found it good that the questions was on a practical issue, as we deal with more theoretical, abstract questions more often at university. I enjoyed the ethical question, too.
- It was a well-thought-out question that made us think of the cause and then responses to it. I think it will be useful in practice in the future.
- If perspectives such as those adopted in the CRT-type questions are important, I feel we should have more group discussions and case studies in our classes. We haven’t had them before, so we didn’t have an opportunity to learn approaches to problem-solving.

Achievements from Participation in AHELO-FS

- Through the participation in establishing the framework for learning outcomes, we could express our views to make the framework consistent with criteria used for licensing engineers and actual conditions in the educational arena in Japan.
- Having participated in the question developing process (selecting subjects, making items, and developing/examining scoring rubrics), we could actively send information to the world from the perspective of Japan’s engineering education.
  - In the world, there is a strong interest in and a high expectation from Japan’s engineering education supporting Japanese science technology. Accordingly, it was recognized that Japan was expected to make international contribution in assessing learning outcomes in the field of engineering to respond to such an interest and expectation.
- Another thing that was realized through the participation was the importance of cooperative efforts among experts in the fields of engineering and education to improve educational systems within both national and international frameworks.
  - It is necessary to closely examine possible influences of such efforts in engineering on other subject areas, and to explain and pass on findings and achievements to other subject areas.
Insights Obtained from Experiences of Developing, Implementing, and Scoring the Examination

- Multiple-choice question (MCQ) (Engineering fundamentals) – It was found that questions Japan proposed were at the internationally accepted level.
- Professional Engineer Examination by the Institute of Professional Engineers, civil engineering licensing examination of Japan Society of Civil Engineers.
- No significant variation was identified among participating countries in terms of average score and distribution (first phase).
- Constructed-response task (CRT) (Engineering processes) – Difficulties in developing questions and in scoring were shown.
  - Difficulties that can be got over by accumulating experiences.
  - Appropriateness of difficulty.
  - Amount of information that can be dealt with within the given time of an examination.
  - Translation skills (word order, unfamiliarity with passive voice, revision of subject, and place/local).
  - Difficulties of uniformly measuring the ability to think.
  - Difficulty in presenting the key point of a question: clearer the key point is made, more limited the scope of answer gets, and that inevitably depletes thinking.
  - Efforts to standardize scoring criteria actually review the acceptable range of a correct answer.

Future Prospects

- More experts will have opportunities to be involved in developing questions for examination and scoring rubrics through continued efforts.
- At the same time, it is necessary to make good use of the international assessment of learning outcomes to improve education by disclosing information on obtained findings and achievements.
- Expert groups are required to explore more effective international methods for assessing learning outcomes with a long view.
- It is necessary to identify such survey design and information disclosure that are useful to improve university education, and to take measures for their realization.
- It is necessary to devise a sustainable means from the aspects of human resources and cost to identify an international assessment of learning outcomes that has significance for participating universities.

Insights Obtained from Experiences of Developing, Implementing, and Scoring the Examination

- Conceptual framework “Engineering process is supported by learning engineering fundamentals”
  - CRT and MCQ are highly correlated with each other (First phase).
  - Is it really necessary to include CRT-type questions in an examination? – Yes
- Meaning of CRT
  - “Measuring competencies of engineering process using CRT” is significant by itself in terms of experimental study. At the same time, presenting a clear message, “competencies are important in the engineering process,” has an essential implication as an initiative. Such international efforts are expected to be influential.
  - CRT can encourage university teaching staff to discuss in a cooperative and interdisciplinary manner about education programs necessary for students to gain such competencies, thus urging university teaching staff to embark on an effort for educational improvement.
    - Structuring education programs for the acquisition of competencies.
    - Devise education methods (e.g., PA, linkage among subject area, and internship).

5. Summary

- It was recognized in a tangible way that an internationally common view was emerging regarding contents of learning outcomes in engineering education.
- It was realized that cooperative efforts among experts in the fields of engineering and education were important to improve educational systems within both national and international frameworks.
- In the world, there is a strong interest in and a high expectation from Japan’s engineering education supporting Japanese science technology. Accordingly, it was recognized that Japan is expected to make international contribution in assessing learning outcomes in the field of engineering to respond to such an interest and expectation.
- A long view is required when discussing assessment methods for learning outcomes, especially regarding how and what are measured and compared with. For this matter, international expert groups need to provide further consideration. Japan’s active participation in such efforts is highly significant.
Thank you for your attention.
Section 2: Panel Discussion

Report
Satoko Fukahori
Senior Researcher, Department for Higher Education Research, NIER
Informing Universities for Educational Improvement: The AHELO Feasibility Study Experience in Japan, Canada, and Australia

Panel Discussion
Facilitator
Motohisa Kaneko*
Professor, Research Center for University Studies, University of Tsukuba

Panelists
Robert Wagenaar
Director of Undergraduate and Graduate Studies, University of Groningen
Peter Ewell
Vice President, National Center for Higher Education Management Systems
Kikuo Kishimoto*
Dean, Graduate School of Engineering, School of Engineering, Tokyo Institute of Technology
Mary Catharine Lennon
Senior Research Analyst, Higher Education Quality Council of Ontario
Daniel Edwards
Principal Research Fellow, Australian Council for Educational Research
Satoko Fukahori*
Senior Researcher, Department for Higher Education Research, NIER

N.B.
• The * mark indicates that the original language of the speech was Japanese and that the transcript is a tentative translation based on the simultaneous interpretation provided during the symposium.
• The transcripts include changes made after the symposium for purpose of publication.
• The affiliations and professional titles of the speakers are as of December 10, 2013.
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).

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), 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.

---

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

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).

### 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

---

²A similar conceptual competence framework based on the Tuning approach in the Economics strand (OECD, 2011b).
taken to establish the international comparability of engineering degree programs through mutual recognition of accreditation and cross borderer 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. 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 borderer 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>
<thead>
<tr>
<th>Competence cluster</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic and Engineering Sciences</td>
<td>Ability to apply knowledge in mathematics, science, and engineering.</td>
</tr>
<tr>
<td>Engineering Analysis</td>
<td>Ability to design and conduct experiments, and analyze and interpret data. Ability to identify, organize, and solve engineering tasks.</td>
</tr>
<tr>
<td>Engineering Design</td>
<td>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.</td>
</tr>
<tr>
<td>Engineering Practice</td>
<td>Understanding of professional ethical responsibility. Knowledge of issues in modern society. Ability to utilize techniques and the latest engineering tools necessary for engineering practice.</td>
</tr>
<tr>
<td>Engineering Generic Skills</td>
<td>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.</td>
</tr>
</tbody>
</table>

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.
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.

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](image-url)
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.

### 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

![Figure 5. Validity of the instruments: evaluation of program and test scores](image)
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.

Figure 6. Allocation of time and test scores
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.

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.

As shown in Figure 7, cross-tabulations of course types and test scores show that students that take more “lectures”
Section 2: Panel Discussion

(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 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 fulfill 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.*
- Center for Research on University Management and Policy (CRUMP), The University of Tokyo (2008), “National Survey of University Students”.

**Websites**

Section 2: Panel Discussion

Informing Universities for Educational Improvement
The AHELO Feasibility Study Experience in Japan, Canada, and Australia
Satoko Fukahori
National Institute for Educational Policy Research

OECD-AHELO State of Progress

- OECD-AHELO: An international test that aims to measure what students in higher education know and can do upon graduation.
- OECD-AHELO Feasibility Study
  - Is the assessment scientifically possible?
    - How? • Defining measurable learning outcomes that demonstrate mastery of a given competence, and developing instruments that measure attainment of those learning outcomes.
    - Scoring? • Defining through scoring rubrics the scope and level of learning outcomes to be attained, and sharing an understanding of the framework.
  - Is the assessment practically possible?
    - Getting universities and students to participate.
    - Can we verify the reliability and validity of the instruments?
      - The conclusion of the Technical Advisory Group.

If there is a Future AHELO
- For whom and for what?

- Who will benefit from a future AHELO and how?
  - How will the results be analyzed, and how will they be reported?

※ In order for a future AHELO to be successful, we will need to share a clear vision on “for whom and for what”

Satoko Fukahori
2013.12.10 NIER Symposium

The AHELO-FS Research Design

- A convenience sample of universities
  - The university sample does not represent the country’s HE system.
    - The “university population” is not self-evident – what is a university?
    - Randomly sampling a university from the university population, and gaining their participation at a high rate will not be easy.
  - AHELO-FS was not designed to compare countries.
- A random sample of students
  - The student sample represents the university’s program’s student population.
  - Students responded to different sets of items based on the item response theory model.
  - Information such as the difficulty level and student distribution of each item are necessary in order to standardize and make comparable student test scores. Some such information is not yet available, it is not at this point possible to conduct a robust analysis with students as the unit of analysis.
- The unit of analysis is the “university.”
  - The purpose of AHELO is to provide institutions with information that will help them improve their education.

Satoko Fukahori
2013.12.10 NIER Symposium

What are the kinds of information that universities seek from AHELO?

- Consultation:
  - NIER Project Study on the Analysis of the AHELO-FS Results(2013).
- 1. International benchmarking
  - How are their students performing in relation to their international peers?
  - Not ranking.
- 2. Diagnosis of student competencies.
  - What are the strengths and weaknesses of their students’ competencies?
  - How can programs and student support systems be improved?

Satoko Fukahori
2013.12.10 NIER Symposium
Test Scores
Do they measure “ability to think autonomously?”

- SWLE (Scaled Weighted Likelihood Estimate)
  - Combined total score of Multiple Choice Questions and Constructed Response Tasks.
  - Estimates of student ability calculated based on the item Response Theory model.
  - Average 500, Standard Deviation 100.

- Student Comments to the Test.
  - Multiple Choice Questions
    - Most of the contents were covered in class, so everyone should be able to solve them.
    - The questions were objective, superficial, and unconnected. It would be better to delve more deeply into the subject matter.
  - Constructed Response Tasks.
    - Interesting questions, asking what caused the problem and how to solve it. Useful in practice.
    - Sometimes, we deal mostly with theoretical and abstract problems. The rest was interesting because it dealt with real world problems. I liked the ethical questions.
    - If constructed response tasks are important, we should have more classes with group work and case study. We have not had such experience, nor chosen to learn approaches to solving these types of questions.

Hypothesis 1: The lower the proportion of “lectures” and higher the proportion of “seminars” and “group-work,” the stronger the capacity for “autonomous study.”

- The proportion of “lectures”
  - 81% of the students in the Japanese sample responded that “61-75%” or “75%+” of their classes were of “lecture” type. The test scores for these groups were relatively high.
  - A similar pattern was found in the Country A sample.
  - In the Country B sample, students responding “41-60%” and “61-75%” scored relatively high on the test.
  - For the three country-samples, around 60% “lecture” type seems to be the preferable proportion.

The Relationship Between Teaching and Learning Environments and Student Test Scores
- an Exploratory Analysis


Recommendation 1: Induce qualitative change in Bachelor level education through encouraging autonomous study.
  - Increasing time spent on active learning such as discussion and debate.
  - Encouraging learning outside the classroom, such as internship.
  - Hypothesis 1: The lower the proportion of “lectures” and higher the proportion of “seminars” and “group-work,” the stronger the capacity for “autonomous study.”

Recommendation 2: Increase quality learning time.
  - Securing sufficient total time for autonomous study—preparing for class, taking classes, and later developments
  - Hypothesis 2: The longer total time spent on study, the stronger the capacity for “autonomous study.”

Sharing National Data
An MOU between Australia, Canada, and Japan

- For the purposes of educational research each Party desires to disclose and share its respective national Data with each other Party (“the Purpose”).
- The Parties agree that no Party shall disclose to any third Party, publish or otherwise disseminate any other Party’s national Data disclosed to it under this Memorandum of Understanding or findings or research based on such Data without the prior written permission of the other relevant Party or Parties.
- Any major public announcement in connection with this Memorandum of Understanding must be agreed by the Parties before it is made, except if required by law or a regulatory body, in which case the party required to make an announcement must, to the extent practicable, first consult with and take into account the reasonable requirements of each other party.

For the Japanese sample, students spend relatively little time on active learning activities. However, those taking more “seminar” or “group-work” type classes do not necessarily score better. The preferable proportion seems to be about 25%
Section 2: Panel Discussion

Aim “11-15 hours” “preparing for class.” Reduce time spent on “paid work unrelated to study.”

References

- Tuning Association. Tuning Educational Structures in Europe. (http://www.unieduva.org/tuningeu/)

Hypothesis 2: The longer total time spent on study, the stronger the capacity for “autonomous study.”

Number of Hours Spent on Activities
First Semester of Third Year (per 7-day week) (average)

Informing Universities for Educational Improvement based on Higher Education Learning Outcomes

- The information on education that universities seek
  - Information that will allow universities to better understand the state of learning outcomes of their students in relation to international standards.
  - International benchmarks
  - Competence frameworks
  - Information that will provide implications on ways to improve education.
  - The relationship between teaching and learning environment and student learning outcomes.
- When does learning outcomes assessment most effectively inform universities?
  - The competences being measured are aligned with the competences being pursued through university education.
  - The importance of developing competence based degree programs.

Thank you for your attention!

fukahori@nier.go.jp
tuning-aheло@nier.go.jp
Motohisa Kaneko: Thank you for giving me the microphone, but the burden is overwhelmingly big, because a huge number of questions have been received. It’s going to be quite difficult to have all of these questions responded to. I believe that there are three basic issues. I would like to list those three major points. And I want to try to summarize all of the questions. I would like to apologize if I miss some of the questions in my summary at the outset, but please understand that there are time constraints. So, there are three fundamental issues. First of all, in order to do the test, in one way or another, the competency is conceptually summarized, and this is very relevant to Tuning meta-profiling or meta-competence or meta-skills. It’s necessary to, in some way, conceptualize capability or competency. In the case of the United States, “generic skills” is the phrase that is frequently used. But the concepts themselves are based on certain logic. What is that logic? Tuning versus critical thinking is also frequently discussed in the United States. Is it possible to integrate these two concepts? Are they similar or the same? How are these concepts developed and how are they emerging? What would be the basis or logic if they are to be used empirically or demonstratively, especially when we try to use them as demonstrative tools? That is the first area. The second area concerns this: when we conduct such a large-scale test, what are the practical or logistical problems? You said that you were able to conduct the tests, but what were the practical and logistical results, especially in terms of cost? Dr. Ewell was mentioning this. How much expenditure or cost was incurred and what were the benefits? Who should be bearing the cost? How do we convince the general public that these tests are beneficial? How do we identify the effectiveness of these benefits or tests? Those are the questions related to the second issue.

The third issue is something that you are all interested in. What will happen in the future? The first sub-point is whether this can be developed under the umbrella of OECD on an international basis. Secondly, is it possible to apply this beyond the current scope to other areas? When I say “other areas,” I mean other fields or subjects. For example, in the case of engineering, the area was limited to civil engineering, but even within the universe of engineering, there are various other subject fields. What about liberal arts? Is this applicable to liberal arts? Or further, inside Japan, this may sound closer to PISA in secondary education, but can we also expand this and include other tertiary education institutes within the scope?

Those are the three issues that have been raised, and the many questions submitted can be classified under these three major groups. Unfortunately, we have less than an hour for discussion. I wish I could ask each of you to answer all these questions, but I’d appreciate it if Prof. Wagenaar and Dr. Ewell could focus on the first question, which is the framework of capability or competency. Dr. Edwards will be asked to comment on the Generic skills assessment, which is based on CLA and ACE that were used to choose the problems or questions asked. If you compare them, can you point out the differences or anything that
comes to your mind by way of comparison? And with regard to the second issue, the practical issue of costs and benefits, who will bear them? I would like to ask Ms. Lennon to share with us the Canadian case and Dr. Fukahori, to share with us the Japanese case. How was the arrangement done in Japan and how will it be done in the future pathway? Of course, we will have to ask Dr. Ewell to comment. Prof. Kishimoto, we would appreciate it if you could expound on whether it’s possible to do this further in Japan and whether you think that it’s possible to do this in other subareas of engineering other than civil engineering. You can touch on other points first, then the three, but for the three particular points that I raised, I will be asking you those questions.

Is there any problem with the questions? Okay, sure. So, who would like to start? Prof. Wagenaar? Please take the floor. From your side, we go one by one, following the seating order.

Robert Wagenaar : OK, this is a difficult question and I think I have to be open with you. From the very start I had my doubts about CLA, to be frank with you. And the reason is -- which was already explained by Peter -- that it is an instrument which has been developed in one context, the American context. And I suppose in one context it has proven its usefulness but the moment you move it to another context, it might lead to problems. And the CLA involved two issues here. Of course there are differences between cultures throughout the world. That’s one issue. Secondly and I focused on this already during my presentation, it makes a difference whether it’s one discipline or another discipline. So the people I work with, the many academics, thousands of academics throughout the world now, they seem to be convinced that if you develop generic competency, and they should be developed -- that’s without any question, because it’s important for society -- it should be done in the context of a domain of knowledge. If you develop them in the domain of knowledge, you should also test them, you should measure them in that domain of knowledge. That seems to be fair. But the fact that we are operating in different cultures, and you get the example of, for example, the competency conceptualisation. We have to find out whether there are differences, we still have to find out whether there are differences between different regions. I think for good reasons, Tuning organized its frameworks region by region. And at the moment, we are busy comparing what has actually been the outcomes, whether we have a common understanding in terms of the cultural and also the educational settings. And that will show us, whether we can actually compare those concepts. I think there’s a good chance that it will be possible because we also learnt from engineering and economics that if the moment you, in particular engineering in this case, if you link them to a field of study, there’re a lot of common grounds.

So I think there is prospect, and you have asked me about the future. I’m not sure or maybe I’m sure that AHELO was to a certain extent a risky affair. Because we have checked, we have tried to find common ground between 17 jurisdictions and some of those jurisdictions came in rather late and also that might have been a problem. I’m not sure whether we should not test this first region by region and then compare the regions and I brought this also up when we were discussing this in the TAG, and that step-by-step approach might probably be more promising than what we see at the moment.

There’s also another model: we start in one region and we open that region up to other regions, in a sort of circular system. That might take time, but also might be more rewarding. A few things from the perspective of Tuning about the cost and benefits. Because we have been working very hard in trying to frame what the disciplines stand for, we learned a lot from each other when making these frameworks, working all over the world. We also learned that some systems are more effective in developing a certain set of competences than others, but to be sure, we have to compare in one way or another, to learn more about it. But most of all it’s the responsibility, or should be the responsibility of the academic world. The issue, I wouldn’t say the problem, but at least a challenge with regard to AHELO was that
it was a top down approach. It was initiated by the Ministries and the universities were invited to participate at random. But also, with regard to the students, as Peter showed us, we weren’t sure that actually students showed up, in such a way that you get reliable outcomes. Maybe we should turn it around, maybe we should give a more prominent role to the universities, a more leading role in the next step. There will be a next step. I’m absolutely convinced because I think there’s not much alternative if we are working in a competitive world, also we should be sure that what we are doing can be compared and therefore, be measured. But we have to find the right framework and I’m not convinced that the framework we have developed so far seems to be the right one.

Peter Ewell: We had no framework for generic skills. We began with an existing instrument, and there is nothing more “top-down” than beginning with something that should have been developed incrementally. I think we learned that lesson and I would not recommend doing this again if we were to go forward.

I think what you’re seeing in many respects with Tuning and AHELO is the top-down and the bottom-up growing out of the discussions among teaching staff, is that they are meeting in the middle. I think that what we’ll find is a situation—you’ve called it sector competencies or skills—where we can span these things a little bit more. We had many discussions in the TAG about what the architecture of an assessment might be that could do that and we did suggest two significant, modifications to what was done. If there is to be a free-standing generic skills instrument, it should be heavily discipline “flavoured.” Ironically, the examination that the CLA was derived from was not about everyday situations. It was anchored in the disciplines and I think we need to honour this approach in moving forward.

The other thing that we recommended is that there has to be, regardless of discipline, a common core of items that is administered to everybody, so that you can see the way in which different disciplines perform. Getting back to the question of the conceptual basis of any AHELO instrument, it is hopeless to try to assess a domain which is not well specified. That’s why you have to do the Tuning work first in order to get some kind of agreement on what you’re really talking about. So I think there will be disciplines, and this is skipping to your last question, Professor Kaneko, which just cannot be done in AHELO. They are simply too culture specific and context specific. I would feel fairly confident about assessing all parts of engineering in AHELO and I think that there are many practice disciplines that could be done quite well. But Law? I don’t think so. History? I don’t think so. There are a number of disciplines that would be very, very difficult.

Kikuo Kishimoto: All three questions are very difficult to answer. So I would like to provide the answers to the extent that I can. One is whether the scope will expand from engineering. In terms of the learning outcome conceptual framework, what would be the basis for that? In the case of engineering education, as the students and engineers had high mobility, the conceptual framework had to be sorted out in the accreditation of the framework in that perspective. Well, of course, there’s the question of whether what we have is perfect or not. Nevertheless, the opinions of the people of the universities, the employers, and the engineering students have been reflected in the discussion of what is the important learning outcome.

In that sense, the group in Europe and the Washington Accord Group have almost the same results, and what is required of engineers has been fixed to some extent. However, the capability required of engineers has changed with the times, and the framework must be deepened. We need to continue the discussion
from that perspective.

At AHELO, questions will be created and tested. First, we’ll have to find out whether the learning outcomes can really be checked. We haven’t had enough time to reflect on that, as there are still other things that we need to work on in terms of application capabilities. How we can measure that is one issue. As Prof. Wagenaar mentioned, the capability to work within a team is considered very important. But if the testing is measured on an individual basis, can we really measure the teamwork capability? With regard to communication capabilities, scientists and engineers communicate in various ways, including charts and graphs. It would be ideal to have a test in place with which we would be able to measure all of those capabilities. We may not go there 100 percent, but by continuing at this, I think that we would need better means of measurement.

The students are learning not only in one country, but in various countries. As they go from one university to another, the kind of education being conducted in each university must be considered. In terms of whether this would expand, Japan intends to provide globally competitive education programs in engineering. We thus need to check where we are and what kind of performance can be expected from the students enrolled in our programs. I think that this should be done with other universities, within the cooperative framework.

We have been working with various universities under the AHELO initiative, and I feel that other university professors may want to do this kind of cooperative framework. But who would be paying for that? And who would be providing resources to conduct this kind of initiative? These are very big challenges. In addition, our collaboration with other universities and international consortiums has shown us that Asian countries want to undertake this kind of initiative. So, how do we proceed with that? Well, in the AHELO initiative, OECD has taken the lead. When the name OECD is attached to an initiative, it’s easier to find people who want to participate. Also, if we are able to create university frameworks, maybe on a regional basis, it would be better. That is the comment that I would like to make. Thank you very much.

MC Lennon: Thank you. I have been asked to speak a little bit about some of the practical and logistical results, if the cost benefits actually justify some of the work that we’ve done. I think that as a feasibility study it’s naturally a slightly more expensive endeavour than a full scale activity would be. But have the costs actually justified the work? I know in Canada, because we don’t have a Federal government that’s responsible for Education, we run provincially, a number of the provinces did not choose to participate, because they felt it was a very costly endeavour, because many of the provinces don’t actually have that many institutions to compare. Some of them have one institution per province.

So that upfront cost and the uncertainty of what was actually going to be provided back to the provinces, was certainly a consideration that we dealt with. It was Ontario’s choice to participate and entirely take on the costs associated with the project, because we were not only investigating the learning outcomes assessment through the OECD’s AHELO project, but we were involved in a number of other initiatives as well. So we recognised that there’s a value in investigating and investing in learning outcomes by participating in Tuning projects, by participating in implementation activities for institutions to actually embed learning outcomes in their projects.

So ultimately we of the province felt that there was actually good reason to invest in this activity, and the impression or the desire of the province was that we were hoping to get some international comparative information to really see how we benchmark ourselves across the world. And as it turned we were going to be able to have some comparative information, and the OECD chose not to present that to us in a formal way, but it was something that we as the Australians, the Japanese and the Canadians, felt was a really important part of this study, and so collaboratively we decided to seek out that comparative information.
Our Institutions were very interested in participating and they did so free of charge, we gave them a very nominal fee to engage in the activities, and they did so with gusto, they were really quite happy to participate in it, and felt that they were going to be given significant information back to help them inform their teaching and learning and to help their students actually and engage in the international world that we live in, particularly in the engineering programmes in Ontario. They’re a very international student population, they’re a very international faculty, and we know that they’re very mobile as well. So they were very, very interested in participating with the international comparisons.

I think overall, the province, the faculty and the institutions were quite satisfied with what they received based on the costs that they’d put into it. And though the institutions were very much looking forward to having a bit of a better idea about the competencies, the certain clusters of competencies and how they compared to each other and compared to other jurisdictions. We realised that as a feasibility study, there are a number of issues that still need to be worked out. And we realise that there’s costs associated with that. So they, though disappointed with certain aspects, they didn’t lose heart in the project, and having read the report and received both the institutional reports from the OECD, and the jurisdictional report that’s recently been presented to them, they’re pleased with the results, they’re interested in working with them and, but mostly they’re very much looking forward to our collaborative report that we’re going to be working with.

So the benefits I think are there for the faculty and certainly for the institutions. And I think they’re continuing to want to engage in these sorts of projects. The question I suppose, is whether or not the province itself has as much to gain out of it, other than supporting the institutions and their activities, without the opportunity to have international comparisons, it’s a question as to whether or not the province benefits as much as perhaps it might, if the comparisons were available.

Daniel Edwards: Thank you. I’ll speak to the three points, but starting with the conceptualising the competencies and starting with the generic skills, I think Professor Ewell and Wagenaar have already sort of covered a bit of that. But I think I’d like to juxtapose the way that the framework for engineering was developed and then the instrument versus the way that the generic skills were developed that were pretty much the opposite way, so the generic skills, as we’ve said, the test existed, but a framework was retro-fitted on top of it, whereas the way that, in civil engineering is where I did a lot of, had more of the involvement, as Professor Kishimoto has said, it involved employers, it involved students and it involved academics most importantly and educators in identifying these competencies through the Tuning project and then I guess ret-tuning it to fit AHELO. And then thinking about once we had these competencies, what sort of questions would be best to put on top of the framework and actually give us some outcomes or give us a measurable, questions that we could find in the outcomes. That issue didn’t happen with generic skills, because of the way that things were designed, there’s no necessary blame involved, but it also points to the difficulties with generic skills as a concept. And I think if you have a look at the civil engineering framework, you will see that there are special noted engineering generic skills and it’s one of the areas of the framework and I think and I agree with Professor Wagenaar in particular, who’s said a number of times today, how important it is to think about the discipline based generic skills. There are engineering generic skills that just simply aren’t generic skills, in sociology for example. But then there are other generic skills which span all disciplines, but perhaps are used slightly differently by people in different disciplines.

So I think it’s a very difficult and contested area, even in terms of something relatively more simple like civil engineering, we produced the framework, if you got another ten academics together, they would produce a different framework. There would be overlap of course there’s overlap, but there’s no right or wrong
way of doing this. We’re doing some other work in medicine at the moment and I’ll talk about that at the end but we have a similar issue. You can never really nail it as we might say, there’s always someone who’s got a different opinion and a valid opinion to change things.

I was involved in AHELO as part of the consortium but also as the National Project Manager for Australia, so in terms of some of the logistical issues, I’ll just mention it briefly. I think one of the key take-homes for Australia was logistically this can be done. We had the systems, the information technology but also the student record data, and the interest from academics to do this, so logistically it can be done but I guess the big question is what is the cost? For Australia, the costs were relatively large and the government did contribute a lot to the OECD but also to running the programme nationally. But it’s not just money that’s involved, it’s commitment of time, and resources within institutions which we found in Australia to be the most important part in encouraging student participation, and encouraging more than just people going and sitting and taking the test and spitting out data. What we found was the institutions that put in more commitment got more out of it, in terms of discussion about what this is all about. And ultimately what this is about is about making sure that the young people who go through our universities come out at the other end with skills and competencies that will equip them in the work they would like to do in the future. And sometimes that gets lost I think.

So what we found in Australia while I think there are limited, there are significant limitations in the data, we had low responses from students and we were trying to play down the emphasis on the data, what we did get was the beginning of conversations about learning outcomes, about what it is that students should be learning, the framework development and the Tuning process has been very good in terms of thinking about the competencies. And then the AHELO process has been very good in terms of thinking about well how do you assess those, how do you measure whether the students who are coming out actually have these competencies. And that’s been very good and I mean the idea of thinking like an engineer, was sort of a catchphrase that we used all the way through. And it helped spark the discussion not only between academics, deans and their faculty but also between students. In the more engaged universities, there were very good discussions between students about, oh hang on, what we do in class is quite different from what this test is asking us to do. And Professor Kishimoto mentioned this before that especially the constructed responses and they were saying, we want more of this project-based, project management problem based approaches in terms of assessment, to be able to understand what we might face out there in the real world. And prove perhaps that we can think through these issues and provide you know solutions that others think are relevant. So that was a very important part for Australia I think, in terms of what we gained from it, it’s the conversations, and it’s the starting to think about that.

Just briefly, on the future, I think the others have mentioned, or question, there’s been some questions about the use of the OECD as a country-based model for disseminating this kind of information and I think there are good questions about how this must be done in the future whether the OECD is the right conduit for producing, for managing such a system.

At the moment, I guess they’re the ones who are putting their hands up trying to find the money and you are taking the initiative. But that’s not to say that that initiative will be followed through and we’ll wait and see what happens there.

One thing’s for sure and I know Robert has said the same thing is this, almost certainly, is going to happen. We have, and I’m sure Japan is interested, Australia is fixated with university rankings, but all the metrics that you see in the university rankings are research based metrics. OECD and AHELO doesn’t want to be a ranking and I personally don’t think any of it should be but the point is that the way the comparisons and benchmarks are being made at the moment are all based on research, research metrics, bibliometrics stuff. The teaching and learning side of things, which is one of the essential core roles of the university, is not benchmarked and universities who are doing a good job, cannot necessarily prove that using a measure. So I think it will go ahead.
In terms of other areas we’re doing some work with Canada, with Japan, talking with Germany about mechanical engineering, a more collaborative exercise that we’re thinking about. It’s much more bottom up than top-down. And also in Australia we’ve been developing quite a successful collaboration in medicine, where we’re looking at, we’ve built a framework in assessing medical students’ outcomes. It’s Australian based at the moment but there are similar initiatives going on elsewhere which could match up. So I think medicine is another one of those ones which is, its context dependent on a lot of things, but there are some areas that are common.

Satoko Fukahori: Thank you very much. I would like to provide an answer to the second question. In regard to cost, MEXT has contributed money to OECD, and also to the five year research project with the engineering experts to examine optimal approaches to participating in the AHELO initiative. NIER has covered significant costs associated with translation, adaptation, and verification of instruments. Honorariums were provided to students for their participation. So, the monetary cost was not small. Apart from the monetary cost, I must emphasize that the five-year commitment by engineering and education experts have been enormous, and I would like to extend my appreciation to all involved for their generous contributions.

So, what were the benefits derived from those costs? I have been involved in the initiative for five years as a coordinator, and as Prof. Kishimoto mentioned in his presentation, we did indeed learn a lot from the experience. We were able to confirm that the internationally developed competence framework was acceptable in Japan and that the Japanese situation was relatively well aligned with those competences. We were able to experience the process of defining concrete learning outcomes from abstract competences, and substantiating them to assessment items. Through scoring, we experienced building consensus on the scope and level of expected learning outcomes for bachelor students. This experience had rich implications in itself. I’m hoping that the people who have been involved in this initiative feel the same way, and that this experience will have some irreversible impacts on them. And I am hoping that the professors will share what they have learned from this experience with their colleagues. What I regret is that we have not been able to provide the resources necessary for professors to persuade their colleagues. Because of confidentiality issues, only a part of the test items and scoring rubrics developed has been publicized. If AHELO were to have an impact on the quality of education, we need to share the test items and scoring rubrics with faculty, who need the information to be able to share competence frameworks and agree on the scope and levels of bachelor learning outcomes. We need to publish these materials to be widely utilized for faculty development. I would have liked to provide participating universities with outlooks for extensive feedback on benchmarked competence profiles, so that professors could discuss more seriously with their colleagues about the benefits of participating in these kinds of initiatives. That is something we need to work on, but if we are able to do that, I think initiatives like AHELO would be of great benefit to universities. For any university education reform to have an impact, it is the professors that need to be convinced and take action. The role of MEXT, or the Ministry of Education, is to provide a platform for that to happen through issuing initiatives and competitive grants. The role of public research centers like NIER would be to provide coordination and expertise. In other words, we need an action scheme, leadership, and coordination to enable faculty, universities, networks of faculties or universities, and academic societies to take primary roles in their reform efforts.

Motohisa Kaneko: Are there any people who would like to add to the previous comments? Please.

Peter Ewell: I think that a theme throughout all of our responses is that data of this kind is best used to start
a conversation about improvement. That proposition was not intentionally put on the table as part of the AHELO feasibility study. It’s happening now quite nicely, though almost by accident. We should continue those conversations and it’s important to recognise that the data necessary to start a good conversation do not need to be as technically sound as the data needed to make an estimate of a country performance. The cost of producing the kind of data that could be used, as in PISA, to make a country estimate is very, very high. The level of precision needed to start a good conversation in a discipline is a good deal less. Starting with regional consortia is a very promising way to proceed, I believe. The difficulty is that the OECD has never run a process like that. The OECD’s experience is with large scale testing that produces a point estimate for a country. I therefore think that their institutional culture was to try to do the same thing with AHELO in producing a formative assessment. And it didn’t really work for that purpose.

**Motohisa Kaneko:** Dr. Fukahori and Dr. Ewell both considered whether or not this in itself was a success. Putting that aside, I think we can rightly say that this is the starting point of various discussions. But this framework itself went through an international organization, and there’s a certain inconvenience in the usage of that which poses a constraint. At the moment, we’re faced with difficulty even in using the existing data. Bearing in mind that point, what is the least extent to which we can use what we already have and do you plan to do anything further with what you already have in terms of data? Dr. Edwards mentioned that there are initiatives underway in Australia, and Dr. Fukahori also mentioned the initiatives in Japan. But do you at least have any plans to use the collected data to do something in the future? Does anyone like to volunteer?

**Satoko Fukahori:** Thank you. If I may, first of all, Ms. Lennon and Dr. Edwards have been invited to the closed meeting of the universities participating in AHELO, which will be held tomorrow. In that meeting, we plan to update ourselves on what’s ongoing in each of the countries. Dr. Hamanaka, who is serving as the emcee, will do the data analysis; we are trying to provide as much feedback as we can to the participating universities. By the end of March, NIER’s report on the analysis of AHELO will be written and made public. The extent to which we can include data in that paper will be subjected to further study, but we are trying to come up with a document that will be useful to universities when they think about strategies for educational improvement.

**Motohisa Kaneko:** Dr. Edwards, you touched upon the plan slightly, but contextual analysis has hardly been done. Any comments on that?

**Daniel Edwards:** Yeah, thank you professor Kaneko. We are working with Japan and Canada on a combined data set and we’ve started to do some analysis looking at the contextual data. The plan, one of our plans, I think we have multiple plans, but one paper that we’ve been working on at ACER has been around the contextual data, and looking at certain aspects. I mean Satoko did put up some interesting data already about the amount of time spent in work, paid work unrelated to study etc. We’ve been doing some more of that, and some more looking at what issues make a difference and which issues don’t make a difference.

One of the key things and it’s not surprising from the result work we’ve done and the work that Professor Ewell has done, is that when we do a regression analysis and control for a whole lot of variables, the students who say they are quite good in class so they get high marks, are doing better on, perform better on the AHELO assessment. So we know there’s some concurrent validity in the AHELO assessment. We also know that, and some of the stuff that I’ll talk about tomorrow with the institutions, certainly from the Australian data we can see that up to a certain point, students who are working in paid work that is related to their fields, so they’re working as an engineer already, up to a certain number of hours, if they work too much it doesn’t work, but up to a certain number of hours, we can see a correlation with that work and their outcome on the AHELO assessment.
So there’s that, the other key finding, I think that I’ve put in my paper is relating to the extent to which students, what they plan to do after graduation. The students who are planning to become an engineer and work in that field did better on AHELO than the students who were not planning on working in that field once they left the university. So this is a start but there is more to be done.

Motohisa Kaneko: Dr. Ewell?

Peter Ewell: Daniel and I compared the US results on generic skills and the Australian results on Engineering and, although they are different domains, we found the same general patterns of results. So there is definitely something there. And we both found that the level of self-reported academic ability and the level of self-reported effort can be used as statistical controls because we found that that both were strongly related to outcomes. But at the same time, other variables such as what the student was going to do after graduation and how relevant AHELO questions were to this also showed significant relationships with outcomes after controlling for academic ability and self-reported effort. At the same time, some student-reported curricular experiences were present in our data that even if you regressed out self-reported ability and self-reported effort, also remained significant. As a result, I think that we have some solid foundations to further explore teaching and learning processes as a result of our AHELO results.

Motohisa Kaneko: Dr. Kishimoto?

Kikuo Kishimoto: This time around, we not only looked at how much they (the students) were able to score, but we were also able to get the contextual data. What we wanted to do was to look at the correlation between the contextual data and the actual test results; we believe that kind of correlation is very important. When the universities change the way that they educate people, we want to see how the results will be reflected in the change. In order to see that, it would not be enough to do it just one time. Maybe we should wait for several years and take this data again. We need to have this as a continuous process, so we need to continue taking data.

Motohisa Kaneko: I would like to finish this soon. In Canada, the US, or in the EU, do you see the experience of AHELO linking to a different movement? Or what kind of impact do you think it will have on any activities or movements that are in place or being conducted right now?

Robert Wagenaar: I can give it a try. Actually not much is happening, at least not globally. But I think there’s great awareness among universities that this is important. So the fact that it’s accepted as a notion that it’s important, also in Europe, although Europe -- as you have noticed has not been very much represented in AHELO, although the OECD is an European organisation from its origin which is amazing. I talked with the engineer organisations for example recently, and they have a strong interest. But I’m not so sure, and I said this before, that the format we are using at the moment is the one which actually will be sustainable. I personally think and others have referred to this so this seems to be common ground that a sustainable basis seems to be a bottom up approach, and not a top down approach. Universities should see the advantages, academics should see the advantages from the very start. They should be involved from the very start in the exercise, as also Peter has mentioned.

The moment that happens, so people see the advantage because they can learn on the spot more or less, from the dialogue, with colleagues, whether it’s in a region or worldwide, that we have to see how it develops further then there will be interest. Also students will have an interest, if they can get a personalised let’s say outcome with regard to the test they are doing, of course they want to take it, because they can throw it away when it’s
not good. If it’s good, they can show it to a future employer.

So there are means of strategies to draw attention to this approach. But the fact that we want to compare, because we are comparing as was stated already with regard to research, we also want to compare teaching and learning. It forces us more or less -- at least morally -- to come up with a follow-up. So we have to look for the right framework, and that framework might also be a sort of consortium of universities that actually set this up, which are subsidised by organisations like the OECD -- that’s probably asked too much because they don’t have money -- governments, the European Commission, organisations like Lumina, or other organisations throughout the world, that have an interest. Because these organisations exist, they exist in Germany, they exist in the United States, they exist in other countries. So that might be a better approach than we have had so far. That doesn’t mean that what the OECD has done so far as an initiative has not been of tremendous value, because we would never have found out where we stand today if they wouldn’t have done this. So we should be grateful, I think, to all that have been involved, but also to the OCED as an organisation for the work done.

Peter Ewell: I very much agree with that. We owe the OECD a good deal of gratitude for getting this moving. But whether the OECD is the correct vehicle for sustaining a formative assessment like this, I think is a very open question. Again, methodological purity is not necessarily the goal here. The goal here is to get data that are good enough to support a conversation about improvement. So I think regional consortia of countries is a very good idea. Look at the countries that are not represented here-- countries like Columbia, countries like Egypt, countries like Mexico-- that put a lot of effort into AHELO. These countries got close to 100 percent student cooperation rates, are analysing their data now and, there is a Latin American initiative between Mexico and Columbia under way to share their data. None of this was originally envisioned in the OECD’s plan, but it happened. And it happened because the need is there. So my prediction for the future is that we will begin to see small sets of countries getting together to create AHELO-like initiatives of their own, hopefully with support from third party foundations. And all of these should be working from the same kinds of assessment frameworks in the disciplines that they choose to examine, not identical frameworks, but the same kinds of frameworks.

MC Lennon: Thank you. I think that AHELO is part of a massive movement towards establishing, implementing, and measuring learning outcomes. Not only around the world, but in various levels, both in the government, the system level, in the institutions and also in disciplines as well as programmes. So I think that with the interest of those various levels, the approaches to assessing learning outcomes has the opportunity to be done at those different levels. In the engineering programmes for Canada anyway, we have recently developed graduate attributes, graduate learning outcomes, and the institutions are being mandated to not only implement them but also have them demonstrably assessed by 2016. So they’re working very, very hard on trying to figure out these assessment tools amongst themselves, they were very interested in participating in AHELO for that purpose and also looking forward to participating in the conversation about developing the mechanic engineering assessment as well. So we’re seeing pockets of innovation come together, but I think that having the opportunity to collaborate internationally, certainly helps forward our conversations and our conceptual thinking about how to do it. So I think that the OCED offers a very vocal and a very prominent platform to have these conversations. But I don’t think that it’s the only one and I think that there’s the opportunity to support smaller innovative activities as well.

Motohisa Kaneko: The others, do you have anything else to add?

Daniel Edwards: Just a, I’ll try to be, a quick point, but comparing, talking about this bottom up versus top down approach. In the Australian experience, so we’ve run AHELO with civil engineering and I’ve been involved
in medical programmes so I’ve had pretty much the same role in two different programmes. One where I guess government started it through the OECD and we were contracted to do it. And we then said to universities who wants to take part and from there we’ve had relatively good participation from the institutions but students haven’t been in the institutions, have had little bit to do with developing the assessment, but not much. And to be honest, if we tried to run something like this in Australia, we would probably struggle to get this many people. So the commitment isn’t there as much.

On the other hand, the medical programme we’ve done has actually started from institutions contacting our organisation saying, "Hey we want to do something like this. We know that the Australian Council for Education Research has expertise that we don’t have. We need someone to facilitate, how about we band together, get a grant and build something”. We started the medical assessment collaboration with two medical schools. We ran a forum about two months later and we had a room full of people like this. And a year and half later we have sixteen medical schools and we didn’t have to try, everyone wanted to be part of it, because the way that we’ve been building the assessment is through workshops with the institutions, the academics themselves are coming, and the clinicians themselves are coming and doing this work. We’re developing items that are relevant to their curricula, but also that can be commonly used across schools. And also the institutions feel that they are improving their assessment more generally, because of this involvement. So this is a small example but I think an important one and a successful example of how that bottom up approach can work and now I guess our future aim is to then expand regionally, with this idea but starting small for us on that side of things has been good. And the one last thing that I want to say is about that initiative. In the medical schools we are giving the students a pretty detailed student report, every single student who’s offered an opportunity is sitting that test. In engineering, we had to, well we couldn’t but we would have dragged them kicking and screaming to get them to sit the test. It was very difficult in most institutions, so again there’s completely different outcomes on things that are trying to kind of achieve the same thing, but done differently.

Motohisa Kaneko: Thank you. It’s about time. So let me wrap things up with my personal impression. At the beginning, Prof. Kishimoto said that Japanese educational reform will have to focus on competition. In principle, what can be done at the universities, and what difference do the individual faculty members make in terms of the learning outcomes of the students? They have become specific agenda items for the Japanese education system already. In one way or another, measuring what the students are doing is important, not too ambiguously, too abstractly, but in a more concrete manner. In that sense, AHELO is an ambitious project because it tries to establish an international framework that is measurable. But if we actually embark upon such an effort as exposed this afternoon, many problems may arise, such as the conceptualization of the test and various other issues. But this is not the end, and when I say international, even if not all OECD members can participate, a few countries can volunteer. It would also be meaningful if a few institutions can participate on the domestic level. I think these are some of the lessons learned through our trials and errors. In that sense, I think this conversation was extremely useful as Japan enters into a more concrete phase of measurement and comparison. Again, I apologize if I missed some of the questions that were submitted. However, all the responses were greatly interesting and provocative. So, once again, let us thank the panelists with a round of applause.
Closing Remarks

Tsuyoshi Sugino*
Deputy Director-General, NIER

N.B.
• The * mark indicates that the original language of the speech was Japanese and that the transcript is a tentative translation based on the simultaneous interpretation provided during the symposium.
• The transcripts include changes made after the symposium for purpose of publication.
• The affiliations and professional titles of the speakers are as of December 10, 2013.
Ladies and gentlemen, thank you very much for your participation. Thank you very much for attending this symposium despite your busy schedules. We come from different universities, private sector organizations, and others. Members of the embassies in Japan have also participated. And we witnessed very interesting presentations. The flow of higher education reform in the last 30 years was shown by Mr. Kimura at the beginning of the symposium. We were able to position this as the base for other symposiums. Prof. Robert Wagenaar presented the latest information regarding specific methods of Tuning, which builds degree programs based on competence frameworks of specialty areas, and the relationship of cross-discipline efforts and learning outcomes studies. Dr. Peter Ewell shared the overall picture of AHELO, explained its results and issues, and described the future of the main study in a way that was very easy to understand. Prof. Kikuo Kishimoto presented the specific efforts made in Japan in the engineering domain of AHELO. It made us acutely aware of the fact that the mere participation in international learning outcomes studies is not enough. There is a need to participate at the development stage of the questions to gain sufficient outcomes. Dr. Satoko Fukahori, senior researcher of NIER, reported the analysis of the study results of the three countries – Canada, Australia and Japan – on which she has worked together with Ms. Mary Catharine Lennon and Dr. Daniel Edwards.

This time it was a feasibility study. However, the participatory work done in these three countries is extremely meaningful. The panel discussions after the keynote lectures and presentations were quite frank, and very valuable exchanges were conducted. I would like to once again show my appreciation to our facilitator, Prof. Motohisas Kaneko, who made the discussion a very fruitful one by effectively organizing the diverse points of discussion. We can say that Tuning-AHELO provided an important insight into Japan’s higher education policies and the practices of university education. As Prof. Kimura said, higher education is the weak point of Japan. However, this symposium opened the opportunity to undertake university education improvement reforms with a deeper insight. I would like to thank all the participants for attending this symposium despite your busy schedules. Thank you for your kind attention.
Appendix

Reference Paper
The AHELO experience - implementation, outcomes and learning from an Australian perspective
Daniel Edwards
Principal Research Fellow, Australian Council for Educational Research

AHELO: The experience of Ontario's institutions
Mary Catharine Lennon
Senior Research Analyst, Higher Education Quality Council of Ontario

Presenters' Biographies
Introduction

This paper provides background, implementation processes and outcomes relating to Australia’s participation in
the Organisation for Economic Cooperation and Development’s (OECD) Assessment of Higher Education Learning
Outcomes (AHELO) Feasibility Study. The Federal Department of Innovation, Industry, Science, Research and Tertiary
Education (DIISRTE) funded the Australian participation in this project via direct contribution to the OECD, and
through funding of a National Centre to implement the study in Australian universities.

The AHELO Feasibility Study was a major OECD project. Its objective was to determine a robust approach to
measuring learning outcomes in ways that are valid across cultures and languages, and across the diversity of
institutional settings and missions. Further detail including aims and final reports relating to the project can be found

The AHELO Feasibility Study involved the testing of students in three strands – Generic skills, Economics and Civil
Engineering. For the Feasibility Study, Australia participated in the Civil Engineering strand and it is participation in
this strand that is the focus of this chapter. In addition to these assessments, contextual dimension questionnaires
were deployed to students, faculty and institutional staff at all participating institutions.

The implementation of the Feasibility was split into two phases. Phase One was undertaken in 2011 and involved
engaging institutions, coordinating focus groups in participating universities; undertaking an academic review of
draft assessment instruments; marking of student responses; and the collation and communication of feedback
from institutions and students. Phase Two took place in 2012 and involved re-recruiting universities, sampling of
students and staff, technical assistance in preparation for AHELO assessments, administration of the assessments and
questionnaires, scoring of test scripts and reporting of results.

This paper offers insight into the processes, activities, outcomes and dissemination of results from the AHELO
Feasibility Study in Australia with particular emphasis on the student-related data. It begins by providing a brief
background about the concept of AHELO and the practicalities relating to the project. It then examines the process
of implementing the study in Australia, beginning with Phase One and the collection of feedback relating to the draft
instruments. The implementation of the AHELO assessment (Phase Two) is then detailed, alongside some insight
into participation numbers and some notable but broad outcomes reported. The final section of the paper discusses
overall outcomes of the Feasibility Study from an Australian perspective and considers some possibilities for
alternative approaches to dissemination and reporting.

AHELO Background

This section provides a background to the AHELO project. It was jointly prepared with members of the AHELO
Consortium and provides insight into the aims of the study. It is included here to provide an overall perspective for
framing the work detailed in this paper. The overall aims and ambitions of the Feasibility Study are important and are
sometimes lost in the detailed discussions of process in this work, so this section offers the chance for reflecting on
the ‘bigger picture’ of what was trying to be achieved.

Is it possible to undertake an international assessment of final-year students’ capacity to use, apply and act on the
knowledge and reasoning they have gained from their degrees? Is it possible to assess these outcomes in an efficient
and internationally comparable way? Can policymakers, institutional leaders, faculty and students be convinced that the assessment of higher education learning outcomes is an essential checkpoint in the educational process? These questions lie at the core of the OECD’s AHELO Feasibility Study.

AHELO involves the development and validation of assessments in three core areas – Generic Skills, Economics and Civil Engineering, as well as the development of contextual instruments to aid with the interpretation of assessment data. The assessments are targeted at students in the final year of bachelor degrees and aim to assess their capacity to apply their skills and knowledge to real-world problems. AHELO is an ambitious project in that it aims to develop new methodologies and technical standards for the assessment of higher education learning outcomes. It is taking place on a global scale, with 17 countries participating in the development and validation of assessments, and engagement from experts, institutions, governments, and key higher education bodies from around the world. The project is run by a consortium of international organisations, led by the Australian Council for Educational Research.

AHELO responds to a critical information gap. Efforts to improve the quality of teaching and to enhance students’ learning outcomes are stymied by the absence of reliable information which enables comparative judgments to be made about the capabilities of students in different institutions and in different countries, or about the quality of teaching. In the absence of such data on core higher education activities of learning and teaching, the standing of a higher education institution is based largely on reputation and research performance. AHELO’s objective is to create a richer source of information through designing and testing measures which give due weight to teaching practices and learning outcomes.

The AHELO assessments were developed through the collaboration of experts from around the world and subsequently translated and adapted for use in all participating countries. Validation of the assessments with students in participating countries was completed in 2011. The assessments that were developed through for AHELO were implemented across 17 countries in 2012.

The study has had an impact on the way in which discussions about higher education and assessment are carried out, suggesting that AHELO has the potential to reshape the higher education landscape in important ways. First, discussions among some stakeholders appear to have moved on from whether learning outcomes should be measured to how they can be measured. At the same time, the engagement of national systems and institutions (both within and outside the OECD) has increased significantly throughout the life of the Feasibility Study. Taken together, these indicate the existence of a desire for data on the quality of teaching and learning, both to inform improvements and also to demonstrate quality.

**Facilitating AHELO in Australia**

Australia’s participation in AHELO was funded and supported by the Federal Government. The Government facilitated the organisation of a National Centre which included a National Project Manager (NPM), some basic support and stipend for travel. The NPM in Australia was essentially responsible for implementing all aspects of the AHELO Feasibility Study, under the direction of the AHELO Consortium and the OECD Secretariat for the study. Many of the key activities of the NPM and National Centre are detailed in the sections of the paper dealing with activities in Phase One and Two. One aspect of the work not covered as much in these latter sections is the element of communication required through the duration of the project. This aspect is discussed briefly here.

The NPM maintained close and frequent communication with the Federal Government through the Department of Innovation, Industry, Science, Research and Tertiary Education (DIISRTE) throughout the Feasibility Study, providing monthly progress reports during the busy times in the project and having regular teleconferences. The progress reports and discussions helped keep the Department aware of the day-to-day activities of the NPM as well as providing insight into the extent to which the project was being embraced by universities, university leaders, academics and students.
The National Centre also played a critical role in developing relationships and communication with participating institutions. Contact and collaboration with institutions was maintained during the study through teleconferences, a symposium, site visits and email. The key institutional relationship maintained by the NPM was with each Institution Coordinator, who was in charge of activities within their university. Building personal connections with these people was an important and highly valuable facet of the project.

Another key role of the NPM that is not explored as much in the other sections of this paper is that this role provided a channel through which Australian national interests are represented in the implementation of the AHELO Feasibility Study. The Australian Government were interested in ensuring that there was important Australian input into the development and adaptation of relevant test instruments and in the processes and procedures used in data collection.

Implementation of AHELO

This section of the paper provides an overview of the AHELO Feasibility Study as a whole from the perspective of the in-country implementation of the study. Implementation within countries and institutions participating in the Feasibility Study was carried out in two phases. Phase One involved qualitative testing of items through surveys and focus groups and providing feedback to the OECD and AHELO Consortium relating to the instruments constructed for the study. Phase Two involved the implementation of the AHELO assessments in universities and with students under test conditions as well as the completion of staff and institution-level surveys.

During Phase One of the Feasibility Study, qualitative testing of the draft assessments was undertaken in a number of countries. During focus groups and cognitive labs, students provided feedback on the assessments, while a number of senior academics in the area of civil engineering were asked to provide comment and review of the draft AHELO assessments. In summary, the key facets of Phase One were:

- **University engagement and recruitment** – securing the interest and participation of universities.
- **Support for participating universities** – providing detailed information to universities about participation, including, a Manual, discussions and training for Institutional Coordinator and ongoing contact and support.
- **Implementation of focus groups** – carried out at each university and facilitated by the Institutional Coordinator. Involved students taking a sample of the draft assessment and providing feedback through a survey and focus group.
- **Academic Review of AHELO instruments** – by experts in civil engineering education.
- **Collation of test results and feedback from students and academics** – organised by NPM and included marking of test papers, collation and coding of surveys and construction of a national database.
- **Providing feedback on test instruments** – involving analysis of responses and collation of key points for feedback to the AHELO Consortium.

Phase One activities in Australia were undertaken in 2011 and involved ten universities (further detail is contained in the ‘outcomes’ section of the paper).

Phase Two involved the full implementation of the AHELO assessment in universities to cohorts of students online under test conditions. It also involved the collection of survey data from academic staff and Institution Coordinators.

In summary, Phase Two involved the following activities for Australian participation:

- **Institutional recruitment** – re-securing the interest and participation of universities.
- **Development of manuals for institutions** – to assist in implementation of the online testing and in recruiting of students.
- **Institutional Coordinator training and support** – to ensure all participating institutions were familiar with the AHELO test system and AHELO test administration protocol.
- **Institutional sampling** – collecting population data on eligible students and staff members in participating...
institutions.

- **Setting up test computers** – providing advice and information on establishing secure online profiles for testing centres.
- **Test administration** – involving the implementation of the AHELO tests to students in participating universities and monitoring the administrations to confirm test sessions were run according to protocol.
- **Survey administration** – sending the staff and institutional survey to identified population and following up to secure maximum response numbers.
- **Scoring of assessment responses** – involving recruitment of a Lead Scorer, training in scoring and facilitating scoring among a group of experts.
- **Supporting data processing and international analyses** – ensuring adherence to AHELO protocol and maximising international comparability of data.
- **Attend international training and dissemination meetings** – maximising the value of lessons drawn in the Australian context as well as benefitting from insights in other countries.

Phase Two was conducted in Australia during 2012, with testing being undertaken between March and May in eight universities (details in ‘outcomes’ section).

**Recruiting Institutions**

Based on requests from the OECD and AHELO Consortium, Australia aimed to have ten universities participate in the Feasibility Study. As such, early in 2011, the Vice-Chancellor and Dean of Engineering at ten Australian universities were written to and invited to participate. All ten invited institutions agreed to be part of at least Phase One of the study.

Selection of universities involved liaison between the NPM, the Department (DIISRTE) the Australian Council of Engineering Deans (ACED). Universities were selected on the following basis (in order of selection priority):

1. They had actively shown prior interest in AHELO (through informal discussion with the Government).
2. They had previously provided contact details for an Institutional Coordinator to DEEWR.
3. They had an active Civil Engineering program.
4. The ‘sample’ included universities from a range of geographic locations.
5. The ‘sample’ included universities from a range of different ‘groupings’ of institution (i.e. research intensive, ‘new’ universities, regional universities etc.)
6. The ‘sample’ included universities of various sizes.

As each university confirmed participation in the Feasibility Study, they were contacted by the NPM and provided with an overview of the project. Institutional Coordinators were identified and directly contacted by the NPM to provide details on involvement in the project.

Across the two years of the project, students from 11 different Australian universities were involved. These institutions are listed below. Institutions with one asterisk (*) participated only in Phase One, these institutions generally cited resourcing issues as the key impediment to further participation. The one with two asterisks (**) participated only in Phase Two following an invitation from the NPM in early 2012. All other institutions listed were involved in both phases of the Feasibility Study.

- Charles Darwin University
- Curtin University of Technology*
- James Cook University
- Swinburne University of Technology
- The University of Adelaide
- The University of Melbourne
Outcomes

In this section the broad outcomes from students in Australian institutions participating in AHELO are discussed. The discussion is relatively broad and focuses more on the participation levels than it does on the actual ‘results’ from the AHELO assessments.

Qualitative testing – Phase One

Phase One was carried out in ten participating institutions in Australia in April and in August 2011. In total 78 students took part in focus groups in which they were asked to take some sections of the draft assessment and then complete a survey and provide feedback in a focus group setting. The draft assessments included in this session were:

- a Constructed Response task - where students are presented with a scenario or problem and asked a number of questions in relation to this. In most cases these required ‘open answer’ responses (i.e. they were not multiple choice); and
- a set of Multiple Choice questions.

All sessions were run within participating universities by the Institution Coordinator or a designated member of faculty. Sessions lasted approximately one hour and were undertaken by final year Bachelor of Civil Engineering students. Table 1 displays the demographic characteristics of the students who participated in the focus groups for Phase One.

After taking part in the test, students were asked to focus on the following topics during the moderated discussion:

- The task challenged me to think (challenge)
- The materials stimulated my interest in the task (interest)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>(n)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>76.9</td>
</tr>
<tr>
<td>Female</td>
<td>18</td>
<td>23.1</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>5.1</td>
</tr>
<tr>
<td>21</td>
<td>18</td>
<td>23.1</td>
</tr>
<tr>
<td>22</td>
<td>26</td>
<td>33.3</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>15.4</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>11.5</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>3.8</td>
</tr>
<tr>
<td>26-30</td>
<td>3</td>
<td>3.8</td>
</tr>
<tr>
<td>31 and above</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Citizenship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>54</td>
<td>69.2</td>
</tr>
<tr>
<td>Other country</td>
<td>23</td>
<td>29.5</td>
</tr>
<tr>
<td>Language background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>52</td>
<td>66.7</td>
</tr>
<tr>
<td>Other language</td>
<td>25</td>
<td>32.1</td>
</tr>
<tr>
<td>Total students</td>
<td>78</td>
<td>100</td>
</tr>
</tbody>
</table>
- The task made me apply knowledge and skills in real-world ways (apply)
- The task assessed an appropriate range of knowledge and skills (range)
- The task was relevant to my program of study (program)
- The task was relevant to future professional practice (future)

The responses of students were scored and input into a national dataset for Phase One that was provided to the AHELO Consortium. At the national level, the National Centre undertook some analyses of the student responses.

The main outcomes derived from Phase One in Australia was that students rated the test mostly positively on the criteria of challenge, interest, application, range, relevance to program and future professional practice. The Constructed Response tasks were rated slightly higher on these criteria compared to the Multiple Choice questions.

![Figure 1: Feedback on Constructed Response Tasks](image1)

![Figure 2: Feedback on Multiple Choice Tasks](image2)
An overview of the Australian student responses to survey questions about the AHELO assessment instruments is provided in Figure 1 and Figure 2 below. They show that students were highly challenged by the Constructed Response tasks and were also generally positive about the other specific areas explored through the survey – with relevance to program and relevance to their future work being of particular interest (Figure 1). Levels of challenge and relevance to the degree program for these students was also relatively good for the Multiple Choice items, while the detail in Figure 2 suggests interests, application and the future relevance of these items were given lower scores by participating students.

In addition to rating the test through a questionnaire, the student focus groups also entailed a moderated survey format. In brief, the major responses of these moderated discussions were:

- When asked about the Constructed Response tasks, in general students found these made sense and the instructions were adequate. Overwhelmingly students found the task interesting, because of the diagrams, real-world problems, and because the test was different to how we students are tested usually.
- Student opinion about the Multiple Choice tasks was generally that these were more like the tasks they were familiar with at university. Students found the tasks to be more recall oriented and while relevant to their degree seemed less relevant in the long term for their future career.

A review of the draft assessment instruments was also undertaken by three academics in Civil Engineering from participating Australian universities. The outcomes of the academic review can be summarised as follows:

- In general, the feedback was similar to that of students.
- There was some suggestion that the content was focused on the early stages of degree, rather than final years.
- There was more specific item-based feedback relating to technical explanations, language used etc., suggested for change in order to make some items clearer.

Based on the results from the student focus groups and academic review both in Australia and in other participating countries, a number of changes were made by the AHELO Consortium to the instruments used in the Civil Engineering strand. A major outcome at this point in the study was that more time was allocated for the test than originally planned. One Constructed Response task was completely removed from the test since it was too difficult. Other Constructed Response tasks were amended.

Quantitative testing – Phase Two

Phase Two of AHELO was conducted in 8 Australian universities in 2012. Students participating in the Feasibility Study undertook the AHELO Civil Engineering assessment, which involved a 90 minute test including one Constructed Response task and a module of Multiple Choice Questions. Students also completed a context questionnaire at the end of the test session which collected data relating to demographics and other background characteristics of students.

Sampling and student recruitment

A major task in preparing for test administration was the collection of sampling frames from each participating university in accordance with the AHELO protocol. To ensure international comparability of the AHELO data and to facilitate processing of the data collected, all participating institutions were required to provide lists of all eligible students and staff members, in a pre-defined table format. Internationally, this format enabled the validation of target populations and where necessary the selection of a sample. In Australia, Civil Engineering faculty sizes were sufficiently small to allow for a census rather than a sample. Eligibility for being included in the test population was defined as students enrolled in civil engineering programmes in their final year before Bachelor completion.

The provision of sampling frames to the NPM was a significant effort on the part of the institutional coordinators, since it required up-to-date information on the enrolment status of students and a range of demographic
characteristics. Institutions were invited to add any background indicators to their sampling frames they deemed enriching to the data, and some provided information such as the programme students were enrolled in, country of birth of the student, or student aptitude measures such as grade point average. Other institutions were more constrained by regulations of data sharing, and were only able to provide essentials such as gender.

Test administration took place during April and May 2012 in all institutions. For the administration of the student assessment and questionnaire, most institutions organised a number of sessions, while other smaller institutions organised a single session. In some institutions, the institutional coordinator supervised all test administration sessions, whereas others hired test administration assistants financially supported through the NPM budget. During test sessions, the NPM was available for support regarding the AHELO test system and any other queries regarding the AHELO protocol.

A major challenge for the Australian implementation of AHELO was motivating students to participate in the study. This was a problem also highlighted by some other countries during the international NPM meeting in October 2011. Given the variety of institutions participating in Australia, the NPM took a flexible approach to incentivising participation, allowing institution coordinators to propose localised incentive ideas that were agreed to and funded by the NPM. Table 2 details the kinds of incentives offered by institutions to participating students. It shows a range of different options and approaches offered. In some institutions, full student population numbers were small enough to offer a voucher to each participant, in others it was necessary to have a draw of prizes for participants. Besides monetary incentives, students were motivated to participate by explaining to them the relevance of the test to their future work experience, and the value of their participation to the success of the study. One university chose not to offer any monetary or ‘prize’ incentives for students and instead focus solely on the experience of the assessment as being an incentive for participants.

All institutions were provided with participation certificates to provide to students who sat the test. These certificates were developed by the AHELO Consortium and the OECD and included the OECD logo.

### Table 2: Student participation incentives

<table>
<thead>
<tr>
<th>Institution</th>
<th>Incentive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uni A</td>
<td>$AU100 voucher for each participant</td>
</tr>
<tr>
<td>Uni B</td>
<td>ipad draw. Lunch and discussion of Outcomes Assessments for participants</td>
</tr>
<tr>
<td>Uni C</td>
<td>$AU50 voucher for participating students</td>
</tr>
<tr>
<td>Uni D</td>
<td>ipad draw. Voucher draw in each room (8 different sessions) chance in each room to win $AU200 or a $AU100 voucher.</td>
</tr>
<tr>
<td>Uni E</td>
<td>2 ipads drawn, 5 $AU50 vouchers drawn, a lunch</td>
</tr>
<tr>
<td>Uni F</td>
<td>$AU1000 prize draw, and offering student society $AU1000 for a high overall participation rate (75% - not achieved).</td>
</tr>
<tr>
<td>Uni G</td>
<td>none</td>
</tr>
<tr>
<td>Uni H</td>
<td>Made test a 'requirement' for a core subject.</td>
</tr>
</tbody>
</table>

Table 3 provides the participation numbers and participation rates (based on the numbers of eligible participants identified through the selection of a sampling frame) for both students and staff for each university. In total 187 students from Australian institutions participated in the AHELO assessment. Across the institutions involved, one in five final year civil engineering students participated in the assessment. However, this average does not adequately represent the notable variation in participation rates of students between institutions. Five of the eight institutions had participation rates below 25 per cent, while one institution (Uni H) had almost full participation from its cohort.

Numerous issues affected the participation rates of students in Australia. These included students having limited ‘on-campus’ time in their course due to final year internships, a large proportion of students in part-time work and therefore little or no free time, and short notice of the testing window.
However, as the table shows, the Uni H was able to secure a large participation rate, proving that it is possible within Australian higher education to achieve wider engagement of the student cohort in this kind of exercise. Given the success of this institution in securing this level of engagement from students, further information of their approach is detailed in the box below.

### Participant characteristics

The characteristics of the student participants in AHELO are displayed in Table 4. Notable student demographics include the dominance of male students in engineering, and the high proportion of students with a main language other than English. The average age of students participating was 23.2. Most students were studying full-time and on campus. A basic comparison of the sample frame with the secured participation group suggests that by gender the sample is representative, although it appears that the participant age may be slightly older on average than the full population.
Assessment outcomes

Discussion relating to the data outcomes from AHELO participation by Australia is limited in this paper due to the fact that this data is relatively sensitive and the sample numbers and representativeness of students for most of the universities involved in Australia was below the level that would be required for the collection of reliable data. Instead, some relatively broad level findings are shown here using the full Australian data set. Two issues are discussed – relationship between student grades and outcome on AHELO, and the link between work ‘practice’ and outcomes in AHELO. For broad reference, in examining the scores below, the international mean for AHELO is 500 with a standard deviation of 100.

Figure 3 shows the relationship between student outcomes on the AHELO assessment and their grades during university study – self-reported by students. It clearly shows that the Australian students who identified themselves as being among the top in their class achieved a much higher average score in AHELO than other students. This goes some way to suggesting there is some concurrent validity present in the AHELO instruments.

An important issue in Australian Higher Education is the link between study and work and the extent to which study is preparing young people for their future careers. The outcomes from Phase One displayed earlier suggested that at least in terms of the content of the tests, students saw substantial relevance with future work in the Constructed Response tasks. Interestingly, the AHELO data for Australia shows that those students who are engaged in paid work relating to their degree in their final year of university, and those who are planning to work in the industry related to their study when they graduate are much more likely to have higher scores than students not currently and not planning on working in their field in the future.

Figure 4 shows that students who worked at least six hours – but less than 31 hours – per week in a paid job that was related to their study on average had a higher outcome on the AHELO assessment than those not working in a paid job in a related field or those working more than 30 hours in such employment. This is an interesting finding in

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Category</th>
<th>Participant population (% of total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>19.9</td>
</tr>
<tr>
<td>Age</td>
<td>20 and below</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>20.5</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>25-30</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Above 30</td>
<td>4</td>
</tr>
<tr>
<td>Enrolment type</td>
<td>Mainly part time</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Mainly full time</td>
<td>95.9</td>
</tr>
<tr>
<td>Enrolment mode</td>
<td>Entirely on-campus</td>
<td>91.2</td>
</tr>
<tr>
<td></td>
<td>Mix of external/distance and on-campus</td>
<td>8.8</td>
</tr>
<tr>
<td>Country of birth</td>
<td>Australia</td>
<td>53.8</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>14.6</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>Malaysia</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>22.8</td>
</tr>
</tbody>
</table>
the context of considering the extent to skills gathered through working in the field might help influence the extent to which students were able to successfully answer questions in the AHELO assessment. It also suggests that for those students with a very heavy work burden during their final year of study (paid work of more than 30 hours), this paid work may be having an adverse impact on their ability to perform at university.

The other work-study related issue that shows some interesting data among the Australian participants in AHELO relates to the post-graduation ambitions of students. The data in Figure 5 indicate strongly that those students who had plans to pursue a career in Civil Engineering straight after completing their degree were clearly more likely to have a higher score on the AHELO assessment. It is likely that these are more motivated students at this time in their degree and perhaps more focussed than those students with no plan to pursue work in the field. Interestingly, those students intending on pursuing further study in the area had a higher mean score than those not planning on working in the field, but lower than those intending on entering the Civil Engineering workforce.
Conclusion and ideas for the future

The Australian participation in the AHELO feasibility study has resulted in a number of valuable lessons. Firstly, it has shown that the Australian sector is equipped to participate in this type of international study. There was interest from stakeholders in participating, and administrative systems in institutions allowed relatively straightforward production of the required student and staff data. The process of a National Centre overseeing a network of institutional coordinators functioned well, and coordinators report finding this type of cooperation valuable.

Secondly, participation in the process of developing and trialling an internationally applicable engineering test has proven to be insightful both to participating institutions and students. AHELO has proven to be a truly ground-breaking test in the sense that students reported not having been exposed previously to the type of applied, integrated problems mirroring their future work. The cooperation between engineering experts from multiple countries both in meetings and via online communication has provided an excellent opportunity to strengthen international bonds. In addition the cooperation between participating institutions opened up opportunities for future cooperation within engineering faculties.

In practical terms, AHELO has shown that students in Australia are not easily motivated to participate in a voluntary test or questionnaire. Future iterations of AHELO or similar studies would do well to integrate as much as possible these types of assessment into the curriculum, using implementation models such as that used by Uni H. Though participation rates in the feasibility study for Australia were disappointing, the process of implementation has built substantial knowledge on the processes and systems needed for engagement among students and institutions in future studies of this kind.

AHELO has provided some small insights into the current state of engineering bachelor education at selected universities in Australia. While this data does not allow any strong conclusions about the skill level of Australian students as it compares to that of students in other countries (and therefore reduces the outcomes analysis of this paper), it allows a glimpse into what could be possible in future iterations of the study. Importantly, involvement in the AHELO Feasibility Study has provided Australia with valuable lessons and models for future implementation of
such assessments in the future.

For Australia, there are some worthwhile considerations for future international participation that have become apparent through the feasibility study. One is that in the testing window used in this phase, Australian students were technically one semester behind those from institutions in other countries (except Japan). The test window fell in Semester 1 of the final year of the Australian students, but in the final semester of study for those in most other participating countries. For accurate international comparisons, future iterations of the study should be implemented in comparable times during academic years across all countries. A second is that a number of the institutions involved in the Australian participation have substantially large final year internships or research projects in the final year, meaning that students spend significant time in this year off-campus. As such, being able to find a time in which a large cohort is able to participate in a secure assessment is difficult for institutions. Longer term planning for the running of such assessment could help in minimising the impact of these key events in the final year. However, the importance of the flexibility of internships and research projects may make this a challenge to achieve. A final issue recommended for consideration in the future is the production of student-level reports for individuals who participate. It is recognised that this was beyond the scope of the feasibility study, but Australia believes building such capabilities in the future would help to stimulate engagement of students in these types of studies.

Since the conclusion of the Feasibility Study, the Australian NPM has worked with colleagues on identifying useful ways of developing reports for institutions within the constraints of the study (in particular the fact that the assessment items are still secure and confidential). Some innovative work has been undertaken in this regard and an example institution level and national level summary report are presented at the end of this paper. These reports show outcomes at the individual item level, order items by level of difficulty, and provide a reference to the type of capability being tested by each item. The example here is a useful one in consideration of developing such reporting in the future. It is the opinion of the author of this paper that greater momentum following the implementation of the Feasibility Study would have been gained by having clearer and more detailed reporting at the student and institution level.

Acknowledgements

Thanks to colleagues and members of the ACER Consortium for advice and content for this paper and assistance throughout the Feasibility Study Eva van der Brugge, Sarah Richardson, Hamish Coates, Jacob Pearce, Yan Bibby and Xiaoxun Sun.
# EXAMPLE OF FORMAT FOR REVISED INSTITUTIONAL REPORT – PATTERNS OF STUDENT PERFORMANCE

## Civil Engineering Learning Outcomes

<table>
<thead>
<tr>
<th>Item Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engineering Generic Skills</strong></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>use diverse methods to communicate effectively with the engineering community and with society at large</td>
</tr>
<tr>
<td>B</td>
<td>demonstrate awareness of the wider multidisciplinary context of engineering</td>
</tr>
<tr>
<td><strong>Basic and Engineering Sciences</strong></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>demonstrate knowledge and understanding of the scientific and mathematical principles underlying civil engineering</td>
</tr>
<tr>
<td>D</td>
<td>demonstrate a systematic understanding of the key aspects and concepts of their branch of engineering</td>
</tr>
<tr>
<td>E</td>
<td>demonstrate comprehensive knowledge of materials and construction</td>
</tr>
<tr>
<td>F</td>
<td>demonstrate comprehensive knowledge of structural engineering</td>
</tr>
<tr>
<td>G</td>
<td>demonstrate comprehensive knowledge of geotechnical engineering</td>
</tr>
<tr>
<td>H</td>
<td>demonstrate comprehensive knowledge of hydraulic engineering</td>
</tr>
<tr>
<td>I</td>
<td>demonstrate comprehensive knowledge of urban and rural planning</td>
</tr>
<tr>
<td><strong>Engineering Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>apply knowledge and understanding to identify, formulate and solve engineering problems using established methods</td>
</tr>
<tr>
<td>K</td>
<td>apply knowledge and understanding to analyse engineering products, processes and methods</td>
</tr>
<tr>
<td>L</td>
<td>select and apply relevant analytic and modelling methods</td>
</tr>
<tr>
<td>M</td>
<td>conduct searches of literature, and to use databases and other sources of information</td>
</tr>
<tr>
<td>N</td>
<td>design and conduct appropriate experiments, interpret the data and draw conclusions</td>
</tr>
<tr>
<td>O</td>
<td>demonstrate workshop and laboratory skills</td>
</tr>
<tr>
<td><strong>Engineering Design</strong></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>apply knowledge and understanding to develop designs to meet defined and specified requirements</td>
</tr>
<tr>
<td>Q</td>
<td>demonstrate an understanding of design methodologies, and an ability to use them</td>
</tr>
<tr>
<td><strong>Engineering Practice</strong></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>select and use appropriate materials, equipment and tools</td>
</tr>
<tr>
<td>S</td>
<td>combine theory and practice to solve engineering problems</td>
</tr>
<tr>
<td>T</td>
<td>demonstrate understanding of applicable techniques and methods, and their limitations</td>
</tr>
<tr>
<td>U</td>
<td>demonstrate understanding of the non-technical implications of engineering practice</td>
</tr>
<tr>
<td>V</td>
<td>demonstrate understanding of the health, safety and legal issues and responsibilities of engineering practice</td>
</tr>
<tr>
<td>W</td>
<td>demonstrate knowledge of project management and business practices</td>
</tr>
<tr>
<td>Task 1</td>
<td>Task 2</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Item A</td>
<td>Item B</td>
</tr>
<tr>
<td>Score 1</td>
<td>Score 2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Item Descriptions**

- Item A: General knowledge question
- Item B: Specific knowledge question
- Item C: Application-based question
- Item D: Problem-solving question
- Item E: Critical thinking question

**Australian Sample**

- This Institution
- Australian Sample

**Appendix**

- CML Engaging Learning Outcomes
- Example Australian University
- AHELO Feasibility Study 2012

---

**Test Administration**

- Partial Credit Item Score
- Total Score

---

**Table 1**

- Table showing raw scores and item descriptions for each task.

---

**Table 2**

- Table showing item-level data for each task.

---

**Table 3**

- Table showing aggregated item-level data across tasks.

---

**Appendix**

- Additional information and references for the study.

---

**Codebook**

- Details on the coding scheme and data collection process.
### Appendix

<table>
<thead>
<tr>
<th>SD1</th>
<th>SD2</th>
<th>SD3</th>
<th>SD4</th>
<th>SD5</th>
<th>SD6</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image-url" alt="Table Image" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Each student was administered 5 out of 6 items sets.*

---

Multiple Choice Items

CIVILEngEering Learning Outcomes

Example Australian University

AHELo Feasibility Study 2012
## EXAMPLE OF FORMAT FOR REVISED NATIONAL SUMMARY REPORT

### AHELO Feasibility Study 2012

#### Civil Engineering Institutions

**Civil Engineering Learning Outcomes**

**Constructed Response Learning Tasks**

### Task 1: Dams

<table>
<thead>
<tr>
<th>Item description code</th>
<th>Item</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGCRTM17</td>
<td>B, U</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>ENGCRTM16</td>
<td>C, V, W</td>
<td>0.7</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM14</td>
<td>V</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM11</td>
<td>D, J</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM15</td>
<td>K, L</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM12</td>
<td>Q</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### Task 2: Storm Barriers

<table>
<thead>
<tr>
<th>Item description code</th>
<th>Item</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGCRTM28</td>
<td>M, W</td>
<td>1.2</td>
<td>2.1</td>
<td>3.0</td>
</tr>
<tr>
<td>ENGCRTM23</td>
<td>P, R</td>
<td>0.6</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>ENGCRTM24</td>
<td>S</td>
<td>0.6</td>
<td>1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>ENGCRTM25</td>
<td>O, Q</td>
<td>0.4</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM27</td>
<td>M, T</td>
<td>0.4</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM21</td>
<td>K</td>
<td>0.4</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM26</td>
<td>D, T, V</td>
<td>0.3</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM22</td>
<td>J, Q</td>
<td>0.2</td>
<td>0.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Task 3: Concrete and Bridges

<table>
<thead>
<tr>
<th>Item description code</th>
<th>Item</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGCRTM36</td>
<td>A, B, U</td>
<td>1.3</td>
<td>2.2</td>
<td>2.5</td>
</tr>
<tr>
<td>ENGCRTM34</td>
<td>T</td>
<td>0.7</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>ENGCRTM33</td>
<td>K</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM32</td>
<td>D, T</td>
<td>0.6</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>ENGCRTM35</td>
<td>L, O</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>ENGCRTM31</td>
<td>N</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>
AHELO Feasibility Study 2012
Australian Institutions
Civil Engineering Learning Outcomes
Multiple Choice Items

<table>
<thead>
<tr>
<th>Set</th>
<th>Item</th>
<th>Item description code</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>AHELO Sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Institution 100</td>
</tr>
<tr>
<td>Set 1</td>
<td>ENGMCQ3</td>
<td>G</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ1</td>
<td>C</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ2</td>
<td>E</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ5</td>
<td>H</td>
<td>20</td>
</tr>
<tr>
<td>Set 2</td>
<td>ENGMCQ7</td>
<td>C</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ8</td>
<td>F</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ9</td>
<td>F</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ10</td>
<td>G</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ6</td>
<td>C</td>
<td>47</td>
</tr>
<tr>
<td>Set 3</td>
<td>ENGMCQ13</td>
<td>C</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ14</td>
<td>E</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ11</td>
<td>H</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ12</td>
<td>H</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ15</td>
<td>E</td>
<td>22</td>
</tr>
<tr>
<td>Set 4</td>
<td>ENGMCQ16</td>
<td>F</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ17</td>
<td>F</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ18</td>
<td>F</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ19</td>
<td>H</td>
<td>18</td>
</tr>
<tr>
<td>Set 5</td>
<td>ENGMCQ22</td>
<td>F</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ23</td>
<td>G</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ25</td>
<td>I</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ21</td>
<td>C</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ24</td>
<td>H</td>
<td>24</td>
</tr>
<tr>
<td>Set 6</td>
<td>ENGMCQ29</td>
<td>F</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ27</td>
<td>F</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ30</td>
<td>H</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>ENGMCQ28</td>
<td>I</td>
<td>6</td>
</tr>
</tbody>
</table>
AHELO: The experience of Ontario’s institutions
Mary Catharine Lennon
Senior Research Analyst, Higher Education Quality Council of Ontario

Introduction

In 2011 Ontario joined the Organisation for Economic Co-operation and Development’s (OECD) Assessment of Higher Education Learning Outcomes (AHELO) feasibility study. The Higher Education Quality Council of Ontario (HEQCO) led the project on behalf of the Ministry of Training, Colleges and Universities (MTCU) and in cooperation with the Council of Ministers of Education, Canada (CMEC).

Initiated in 2006, AHELO is a feasibility study to determine if standard generic and discipline-specific tests can be used in different countries to measure what university students know and are able to do. Intending to contribute to the international conversation on establishing better indications of learning quality, the study aimed to develop common learning outcomes and assess student performance at the end of a bachelor’s degree (first cycle) in a variety of educational cultures, languages and institutions through standard tests. The feasibility study developed three assessments: one for generic skills, and two for discipline specific skills in economics and civil engineering.

Seventeen countries are represented in this global project, and Canada is one of nine jurisdictions participating in the engineering strand. Nine out of ten Ontario universities with civil engineering programs participated in the study. Nearly 450 Ontario students participated, representing approximately 61% of all civil engineering graduating students.

The following report reviews the experience of Ontario’s participation in the feasibility study, focussing primarily on the implementation and administration activities, and the value to institutions. While the institutions did not gain specific insight on their programming, AHELO generated considerable interest in international assessments and comparative understanding, and provided significant experience in the administration of large-scale assessments.

Background and Rationale for the AHELO Feasibility Study

The late 20th century saw significant changes to higher education worldwide. The massification of higher education produced a diverse profile of institutions, programs and students unlike the small elite systems of previous times (OECD, 2012a). This expansion was aided by a number of elements. Higher education is no longer contained by either bricks and mortar or national/jurisdictional boundaries. The proliferation of technology-enhanced learning allows programs to operate whenever and wherever the student chooses, and international student and faculty mobility have opened institutions to the world. Furthermore, the student demand for education has supported the creation of new and alternative providers, such as private institutions or those with specialised programming.

The now complex and global “market” for postsecondary education (PSE) demands new forms of governance, accountability and signalling mechanisms. In line with traditional forms of accountability and governance, measures of inputs (funding, library holdings, etc.) and outputs (retention rate, graduation rate, publications) have been the yardstick of performance both within systems and internationally. International rankings based on these performance indicators play a significant role in weighing the research capacities of institutions.

What these traditional measures fail to capture is the quality of education: the teaching and learning that is the heart of most institutions. Recognition of this gap led to the inclusion of another set of indicators that can be called “proxies”. These “proxies” of quality include student evaluations, surveys of student satisfaction and engagement,

---

1Information contained in this section is largely based on OECD documentation. For further information, please see Tremblay, Lalancette and Roseveare (2012) and OECD (2013a)
labour market outcomes, etc. These indicators suggest that if the student is satisfied and has fared well beyond PSE, then the institution must have provided high-quality education.

More recently, and considered by some to be a paradigm shift (OECD, 2012a: 35), there has been a focus on learning outcomes as a means to understand, demonstrate and assess educational quality. Defined learning outcomes – clear statements of what a learner knows and can do – along with appropriate assessment measures, provide a transparent means to measure student learning. It is thus possible to gain an objective picture of the quality of teaching and learning, and ultimately the quality of education provided.

Clear indications of teaching and learning quality are beneficial in a number of ways. They support a better understanding of educational value to students, employers, and the public at large. They also enhance institutional and programmatic improvement in coordination, curriculum development and teaching practices. Furthermore, they can provide measures of what has been considered intangible until now in the world of quality assurance and accountability – educational quality. They also provide a transparency of programming, which allows for greater international and comparative understanding for institutions and programs. Hence, learning outcomes demystify education to the benefit of the program, institution, and wider public and international community, but ultimately they put students at the centre of it all. This indicates a significant shift towards student-centred learning.

The purpose of the feasibility study was to provide ‘proof of concept’ (OECD, 2009a: 15) to determine whether it is possible to measure what undergraduates know and can do at the international level, to provide relevant information to higher education institutions (HEI’s), governments, and other stakeholders including students and employers (OECD 2009b: 2).

The primary questions that emerged from the work included:

- Is it possible to have international agreement of expected learning outcomes?
- Is it possible to implement the same test across cultures and languages?
- Are the assessments valid and reliable?

Three primary areas of work made up the feasibility study. Three strands were developed to assess student performance. One sought to assess generic skills, such as capacities in critical thinking, analytical reasoning, problem solving, etc. The other two assessments concentrated on discipline-specific skills in economics and engineering. Rather than assessing content knowledge, both discipline assessments focused on the application of knowledge (i.e., can a student “think like an engineer”). An additional strand, “value-add,” was intended to explore the contribution of higher education to student learning. It is expected that there will be gains in skills and competencies from when a student enters PSE and when they graduate (see OECD, 2013b).

In addition to these two areas of work, a Context Dimension was developed to survey students, faculty, institutions and jurisdictions to better understand the educational environment of each jurisdiction and identify factors that might explain differences in observed performance of students.

Concurrent to the development of the assessment frameworks, the OECD invited member countries and other interested countries to join the feasibility study, share in the development of the assessments and pilot the test to a sample of their institutions. Seventeen jurisdictions agreed to take part in the study, representing six continents and 12 languages. Ontario participated in the Civil Engineering Strand.

---

2 For more information, see Tremblay et al. (2012); p 112-114.
3 For more information, see OECD 2011a; Tremblay et al. (2012), p 116-118.
4 For more information, see OECD 2011b; Tremblay et al. (2012), p 121-123; OECD, 2012a; OECD 2012b
5 For more information, see Tremblay et al. (2012) p 128-132; OECD, 2012c
6 For a list of member countries, see http://www.oecd.org/general/listofoecdmembercountries-ratificationoftheconventionontheoecd.htm
AHELO in Ontario

In July 2011, the Ministry of Training, Colleges and University in Ontario decided to join the study independent of other Canadian provinces. The Higher Education Quality Council of Ontario (HEQCO), an arms-length agency of the Ontario government, was asked to lead the project.

As part of a broad research agenda to understand and demonstrate the value of postsecondary education, Ontario, through HEQCO, has engaged in a number of research projects to explore learning outcomes. HEQCO has been engaged in a multitude of activities in the area, including establishing learning outcomes across sectors of disciplines (the Tuning project; see Lennon, Frank, Humphreys, Lenton, Madsen, Omri & Turner, forthcoming A); measuring generic learning outcomes through piloting the Collegiate Learning Assessment (see Lennon, forthcoming B); as well as supporting the incorporation and demonstration of learning outcomes through a variety of institutional activities. Thus, participating in the international study supported the research base by providing comparative information on systems and programming.

Acting as the National Project Centre, HEQCO was responsible for liaising with the OECD and AHELO Consortium in order to administer and implement the assessment. This entailed acting as national experts to the OECD AHELO conversation, providing a National Project Manager (NPM), implementing and administering the study with participating institutions, and providing analysis.

There were a number of activities involved in setting up and administering the study in Ontario (see Figure 1). The feasibility study, operationalized by the Consortium, provided guidance and protocols for implementation in order to ensure that comparable activities took place in all jurisdictions. The following sections review the primary areas of field work and implementation, documenting both OECD protocols and Ontario’s actions.

Figure 1: Basic Timeline for Ontario’s Implementation Activities

Institution Recruitment and Participation

The OECD recommended that each jurisdiction have a convenience sample of 10 institutions to represent a range of

---

7The Council of Ministers of Education Canada represents higher education at the OECD table and, with agreement from all of the provinces, engages in various OECD projects. Despite general interest, CMEC was unable to gain consensus from the provinces to join the feasibility study.

8See www.heqco.ca for more information
types of institutions providing bachelor-level (or first-cycle) civil engineering degrees. It was desirable to have public and private universities/polytechnics/colleges, institutions of different sizes and with different mandates (teaching or research), and those in rural and urban settings, etc. The goal was to ensure that all types of institutions and of students were adequately represented in order to have a valid and reliable understanding of each of the systems.

In Ontario, 10 institutions provide bachelor-level civil engineering programs. Hence, each institution was invited to participate via a letter to the vice-presidents academic and the deans of engineering. To support participation in the project, each institution was offered a nominal sum to cover the costs of administration. Responses from the institutions were immediate and extremely favourable: nine out of 10 institutions agreed to participate, noting their interest in taking part in this international assessment as a way of understanding their own students and program, as well as those internationally.

Ontario Engineering programs were already familiar with notions of learning outcomes and strategies to assess them. As part of the Washington Accord agreement, and for accreditation purposes, engineering programs across Canada are moving to outcomes-based programming, assessment and accreditation for 2016 (EGADS, 2013). Hence there was considerable interest in exploring this large-scale assessment as a tool. Approximately 90% of all of Ontario’s civil engineering students were represented in these programs.

The participating institutions were all public institutions offering a broad range of arts and science programs up to the doctoral level. Located primarily in urban areas, they ranged in size from 14,595 to 75,941 full-time equivalent (FTE) students. Participating institutions included:

- Carleton University
- University of Ottawa
- Ryerson University
- University of Waterloo
- University of Windsor
- McMaster University
- Queen’s University
- University of Toronto
- Western University

The civil engineering programs were housed in faculties of engineering and were occasionally partnered with environmental engineering. The programs had between 17 and 40 faculty members. The full-time equivalent (FTE) student populations in the programs ranged from 231 to 573. The number of undergraduate degrees awarded annually ranged from 20 to 101 in the 2010 academic year.

**Adapting, Vetting and Validating the Documents**

A Lead Scorer was identified in each country to provide discipline/content expertise in the development of the assessments. The individual, a recognised expert in his or her field, was required to review all test documentation, adapting the assessments and translations to ensure appropriateness for their jurisdiction, and vet the assessment documents through pilot tests of students and faculty. The lead scorers from each country worked together to determine appropriate grading of test items and were responsible for training an in-country scoring team.

Ontario’s Lead Scorer, in collaboration with the National Project Manager, organised various activities to vet the test in the Ontario context, including piloting the test to students and seeking feedback on the test from faculty members and senior graduate students. The feedback commented on the suitability of the content and difficulty level of the questions and their appropriateness for the Ontario context (including issues of language nuances and technical jargon).

This information was presented to the OECD and incorporated into the assessment. Ontario recommended few significant changes. Indeed there were very few items of contention within the entire international scoring team, and jurisdictional lead scorers came to agreement rather quickly.

---

9Five third-year students wrote the test and provided feedback. Third-year students were selected so that the population of fourth-year students would not be compromised.
Institutional Activities

Each institution participating in the study identified an institutional coordinator. This individual was responsible for administering the study within their institution and liaising with the HEQCO as the National Centre. Playing a vital role in the AHELO study, institutional coordinators were responsible for operationalizing AHELO within their institutions.

Standardising field activities was critical to ensure that any resulting information would be reliable. Regardless of how well-designed the assessment might have been, administration and implementation issues could have impacted the validity and reliability of the data. Thus, an important part of the feasibility study involved securing common field activities.

Operationalizing AHELO Within Institutions

Each institution was in close contact with its National Centre to ensure consistency of implementation across a jurisdiction, and also to receive any support required. In Ontario, Institutional Coordinators (ICs), many of whom were the chairs of civil engineering, were identified within their institutions. The nine individuals attended a one-day training session at HEQCO to provide them with context and background on AHELO, and a roadmap for activities. Following the face-to-face meeting, the ICs and the NMP had weekly teleconferences to provide support and advice on activities.

There were a number of tasks involved in operationalizing the field work, and each institutional coordinator developed a team to assist them in the tasks. Recommended team members included an IT specialist and test administrators. The majority of Ontario institutions followed this recommendation, some with larger teams than others.

One of the ICs’ first tasks was to apply for institutional ethics approval to implement the test. Unlike other jurisdictions, Ontario institutions require permission to perform research on human subjects, including students. Based on the short timelines and the need to implement the test quickly, Ontario institutions were required to modify the parameters of the research in order to obtain ethics approval.

For example, whereas AHELO had requested that the ICs have access to students’ administrative information (to ensure that the sample was representative of the general population and to better understand the assessment results based on GPA, high school average, etc.), Ontario institutions were not given permission to link the AHELO test results to student files\(^1\). This resulted in less reliable information on the student sample writing the test and reduced institutions’ ability to examine the data. For example, had institutions had access to administrative files, they would have been able to examine individual AHELO test scores with other indicators of student ability, such as GPA. Institutional information of this nature would have supported a better understanding of the validity of the AHELO scores in Ontario\(^1\).

Implementing Field Work

Student and faculty recruitment

Student recruitment was a critical aspect of the AHELO feasibility study. Each institution was requested to identify a sample 200 students to write the test and aim for a 75% recruitment rate. For those institutions with more than 200 students in the designated area (engineering or economics, or the entire institution in the case of Generic Skills), ICs were required to provide a sampling frame indicating which students were identified as potential participants.\(^12\)

This purposeful sampling frame was intended to ensure that a representative sample of the population wrote the test.

---

\(^1\)Only ICs were to have access to the student administrative data. The national centre, consortium and OECD would not.

\(^2\)Despite the challenges of gaining ethics approval for this feasibility study, any subsequent assessment of this nature would likely be accepted without significant modification given additional time to develop ethics applications and to communicate the goals of the study.

\(^12\)This sampling framework was conducted under the guidance of the Consortium.
In cases where institutions had less than 200 potential test writers, the goal was to have all students participate as a census. All Ontario civil engineering programs had less than 200 final-year students, so all eligible students were identified as potential candidates for the assessment at each institution.

Jurisdictions independently determined their own student recruitment strategies. While some made student participation mandatory, others made it voluntary. Those with voluntary participation did not always incentivize students, though the majority did (OECD, 2013a, p. 169). Student participation in Ontario was entirely voluntary. ICs recruited students in any way they felt was appropriate, as long as it was within the guidelines of their ethical protocol. A range of promotional activities and incentives were offered to students to entice participation, such as posters and information sessions, offers of gift certificates, prize draws and donations to civil engineering class societies. Recognising that it was imperative to have sufficient numbers to conduct any type of analysis, student recruitment was by far the most time-consuming and anxiety-provoking activity for ICs. As student participation was voluntary, all Ontario institutions ended up with a non-random, voluntary sample.

Faculty participation in the short context survey was conducted in a similar way. If a faculty (or institution in the case of generic skills) had more than 40 members, a sampling frame was conducted to ensure that a representative sample was recruited. If there were fewer than 40 faculty members, as was the case in all participating Ontario institutions, a census of all students was attempted. Unable to mandate or entice all faculty members to participate in the survey, Ontario institutions ended up with a voluntary sample of volunteers.

**Test administration**

The AHELO test system was operational between February and June 2012, and institutions could test their students at any time during this window. Each institution was required to run test of the computer platform to ensure there were no technical difficulties. The times and dates of test sessions were reported in advance to the NPM (in case of technical or emergency issues), who reported this information in turn to the consortium to ensure that the online system was prepared to handle the influx of activity. Institutional coordinators were responsible for organizing the AHELO test session and ensuring that the test administrator (invigilator) was suitably trained (based on AHELO training guidelines).

In Ontario, the timing of the test window was challenging. The academic calendars schedule a one-week break in mid-February followed by mid-term exams and the close of the academic year in April, preceded by final projects and exams. As a result, a very short period of time remained for test administration. Most institutions ran their tests in early to mid-March. Institutions offered the test at a variety of times and days, often taking into consideration student class schedules.

Despite testing the system, one Ontario institution ran into technical difficulties and students were unable to submit their tests. While the Consortium rectified the technical issue, all but 8% of student responses were lost at that institution.

**Scoring**

While the multiple choice questions on student assessments were scored automatically using a computerized system, each jurisdiction was responsible for the manual scoring of the constructed responses of its own students. The lead scorers attended two training sessions with fellow lead scorers to finalize the assessment questions and determine appropriate scoring matrices. This activity ensured the consistency of marking around the world. The lead scorer, in collaboration with the NPM, was responsible for recruiting a small team of scorers and training them in both the test system and the scoring matrix.

---

13 Participating programs had between 34 and 137 final year-students
14 Student were excluded if they were out of the country or had disabilities requiring special arrangements.
In Ontario, a team of six engineers made up the scoring team. The majority of scorers were ICs interested in reviewing the assessments and the work of Ontario students. Scoring took place over two days in June 2012 at the National Centre. As part of a memorandum of understanding (MOU) with Australia, Ontario scored some Australian answers and vice versa. This was conducted as an experiment in inter-rater reliability between scoring jurisdictions and as a point of interest for the scoring teams (who would not otherwise have seen other jurisdictions’ student responses).

Successes and Challenges of Administration

A primary objective of the AHELO feasibility study was to understand if it was administratively possible to implement a standard assessment, online, in a common way, to students around the world. Indeed it was proven to be possible: experts and faculty members agreed on the common learning outcomes and the assessment questions, and project management and execution of tasks followed a common protocol around the world. There were minor administrative issues in various countries—some institutions dropped out, others had very low response rates or technical issues—but overall the administration of the international assessment was proven successful. The first volume of the AHELO report lays out a number of international lessons learned from the administration phase of the study (OECD, 2012a, Ch. 6).

There were considerable differences in administrative successes and challenges between participating jurisdictions, as well differences between the three strands. For example, one country recognised that the institutions did not have enough computers for the number of students completing the assessments, and thus required a travelling van to deliver computers to the institutions. Incredibly, Egypt was scheduled to implement the test during the Arab Spring, which created significant— though not insurmountable— challenges. The Generic Skills strand faced challenges in student recruitment.

These implementation issues are critical to understanding the practicality of administering the test—a primary goal of the study. Hence, a great deal of work has been done in collecting and international experiences (see OECD, 2013a; Ch. 8 for a description of each nation’s experience).

Within Ontario, challenges and successes were both generic to all large-scale testing and specific to the AHELO context. Generally, student recruitment for low-stakes testing is extremely challenging. It is time consuming in both the advertising (posters, emails, class visits, etc.) and in the organizing of test sessions. Furthermore it can become extremely expensive when students are provided with material incentives, financial or otherwise. Despite the challenges of student recruitment, the institutions were extremely creative in their strategies, and despite being the most time-consuming activity, it was quite rewarding for the ICs to see their students participate.

As one would expect in a feasibility study, there were a few hiccups in Ontario. The need to obtain ethics approval quickly required institutions to modify the assessment framework to reduce the institutional capacity to link the AHELO results to individual student level data. This reduced the potential value of the results for the institutions; at the time of writing, no institution has yet analysed their own institutional data but have expressed interest in examining it when they receive the comparative jurisdictional report.

In follow up interviews and surveys student and faculty participants indicated appreciation in participating in an international exercise, thereby validating their decision to participate in the study. This suggests that rather than being fearful of benchmarks or comparisons, there was true interest, from all levels, in understanding their programmatic characteristics, strengths and weaknesses compared to those of others in jurisdictions around the world.

---

15More on the MOU sections below.
Discussion and Conclusions

Analysis of Ontario’s Results

Nine jurisdictions and more than 70 institutions participated in the engineering strand of the feasibility study. In Ontario, out of the ten institutions that currently offer a civil engineering program, nine of them participated in the engineering strand of AHELO. Context information was collected from institutions. Across the nine participating institutions in Ontario, there were 155 faculty members and 443 final-year students who participated in the engineering strand of AHELO. Ontario had a response rate of 72% for faculty and 61% for students. While the demographic characteristics, employment status and qualifications of the faculty who participated in AHELO were similar across the institutions in Ontario, there was a noticeable difference in the amount of time faculty reported to spend on teaching and research. Focusing on the student population within the civil engineering program that participated in AHELO, there were also differences in involvement characteristics, such as how students spend their time preparing for class, attending class, or working in paid work related to or unrelated to field of study. Please see Lennon, forthcoming C, for more information on the institutional, faculty, and student characteristics of the participating institutions in Ontario and how they compared to Australia and all nine participating jurisdictions.

What does AHELO contribute to our understanding?

Recall that this feasibility study was not intended to provide any comparative ranking data, but rather to explore the potential for this type of work in both field work and in trialling the assessment tools. Thus, the potential for digging into the data was either a) not agreed to in the framework, or b) impossible due to data limitations discovered later.

Acknowledging that the data gathered must be interpreted with extreme caution, it is impossible to comment on the results and value they provide to stakeholders (i.e. jurisdictions, institutions, faculty members and students), as they are unable to presently use the information.

Thus, the feasibility study revealed that the tools need to be refined in order to say what contribution the information can make. What also became apparent, is that any reworking of the frameworks or assessments further begs the question of who is this information valuable to, in order to tailor it appropriately.

Let us now consider the value of the feasibility results in two ways:

1. What did the various stakeholder groups hope to gain from participating in AHELO?
2. How could the assessments be tailored to suit the needs of the various groups?

Jurisdictions

At the outset of AHELO it was made clear that the purpose was not for international rankings of countries or institutions either internationally or within a jurisdiction. However, the potential for comparing and contrasting across and within countries/jurisdictions was a significant draw. For governments, the ability to know how their institutions and programs are organised and examine the impact on student learning presents valuable information. This has the potential to lead to better comparative policy understanding, and improve the way systems interact with their institutions.

For a variety of reasons, this was not possible in the AHELO feasibility study, but the potential for system-level learning exists. For example, if the information was presented in a way that did not permit for rankings on simple data points, but instead provided comparable information to peer countries or an international average, the potential for international comparisons is possible.

\[16\]

For institution 9, out of the 36 students that participated in the assessment, 33 of the responses were not recorded as a result of a technological difficulty with the online system. The response rate for Ontario was determined using the number of students that participated in the assessment rather than the number of actual recorded responses.
Institutional and program level data

The primary goal of the AHELO feasibility study was to provide information to institutions on how their students performed in various capacities compared to others. Similarly, information collected in the context surveys was intended to provide insight into the characteristics of the educational environments in which students work.

The assessments used in AHELO were not sensitive enough to provide institution level information on student capacities by specific competency areas\(^\text{17}\). Thus, it was not possible to compare institutions on anything other than the aggregate score of their students. This was disappointing to institutions participating in the feasibility study, however, with a refined tool it would be possible to develop a test that compares competency areas of students in institutions. This information would be very useful in providing an understanding of their strengths and weaknesses. For example, an institution might discover they are very strong in design and practice, but weaker in basic engineering skills. This type of detailed information could provide significant information to program and faculty members when considering curriculum design\(^\text{18}\). Understanding of how they compared to other programs in the jurisdiction or internationally was the primary reason Ontario institutions chose to participate in the study. Despite not being able to receive this information, the potential to reshape the tests in order to do it is possible.

The context data collected at the institutional and program level also has the potential to hold valuable information. Ontario institutions also collect administrative information on faculty and students. Thus, much of this information is readily available in other sources, and the information on their own learning environments was not particularly new information to institutions or programs, nor is the within-jurisdiction comparison.

Yet, the point of interest came from seeing the institutions and programs in comparison to others and the international averages. It could be possible to compare institutions and programs around the world to see the commonalities and differences in learning environments and understand if they impact student capacities. Similarly, a pooled analysis of the international trends in program design and environmental characteristics could be provocative in providing information on trends of successful programs or in providing benchmarking through common structural traits.

Thus, while there is value in the institutional information and comparisons currently provided, greater insight would be gained from international contrasts, as well as pooled information on global trends.

Student level data

Student-level data collected in AHELO had the possibility to provide information on observable trends in student demographics and characteristics, and assessment scores.

However, within the feasibility study framework, it was expected that institutions would be provided a comparison of their students’ situations to those in local institutions, and that with that information they might be better able to support student success. For example, discovering that older students are less likely to be successful - perhaps due to outside employment - institutions could choose to develop different student support strategies. Providing comparative information might also encourage collaborations between programs grappling with similar issues. While not provided in the current AHELO framework, it would be interesting to identify international trends in habits of student success, thus supporting cooperation and collaboration regardless of jurisdictional boundaries.

The tests were not designed in a way to deliver student level feedback, as the intention was to combine student-level data at the institutional level. However it became clear in discussions that there is value and interest in providing student-level feedback: a way for students to demonstrate their capacities compared to their peers. In order to provide student level feedback the test would need to be significantly altered to ensure each student was writing a precisely equivalent (or same) test.

\(^{17}\)Engineering Design, Practice, and Analysis, and Generic and Basic Engineering skills.

\(^{18}\)It has been suggested that proper assessment of competency levels would require a re-design of the assessment, and would likely necessitate a longer test for students.
One aspect that was discussed at length by participating AHELO jurisdictions, the Consortium, and the OECD was the possibility of providing students with feedback on their assessment scores. The assessment frameworks were not designed in a way to provide reliable scores at the individual level, as it intended to provide institution level feedback. However, if students received their scores relative to others (either in the institution, jurisdiction, or internationally), there could be increased interest from students as it would provide a global and objective documentation of their abilities. This, in turn, could improve student recruitment and student effort in writing the test (creating more reliable information). This was not the intention of the AHELO feasibility study, but was of significant interest to many participating jurisdictions. This is one of many possible outputs of this type of assessment, but to doing so would necessitate a different assessment framework.

Conclusion

The AHELO feasibility study was a successful enterprise in building international relationships, aiding comparative system-level understanding, supporting institutional/programmatic understanding and exploring the potential of international student-level assessments. A great deal was learned from this research, both within jurisdictions and at the international level of the OECD. We know, for example, that it is possible to administer a standard test to students around the world, and it seems there is interest from a variety of stakeholders, particularly the engineering programs, to do so.

Within Ontario, we benefited from participating in the international conversation on the broader value of establishing and measuring learning outcomes. As Ontario is grappling with how to develop a system that incorporates learning outcomes, understanding how other systems are engaging with them has been appreciated. It was particularly interesting to see the tension between how various jurisdictions desire to use international learning outcomes assessments as either system level benchmarks or for institutional and program level improvement.

As a feasibility study, the AHELO work produced many lessons learned but raised even more questions. It was found that there is interest in international assessments from governments, institutions, programs, faculty members and students. It was also determined that it is possible to have agreement on expected learning outcomes and appropriate assessments from around the world. Furthermore, it was found that it is possible to test students around the world in compatible ways. These were the primary questions of the AHELO study.
References


Presenters’ Biographies

Robert Wagenaar

Prof. Robert Wagenaar is a historian and at present director of undergraduate and graduate studies at the Faculty of Arts of the University of Groningen in the Netherlands. He is also Joint Director of the Tuning Academy, located in Bilbao (Spain) and Groningen; and a member of the Editorial Board of Tuning Journal for Higher Education (TJHE). He is also coordinator and director of the Erasmus Mundus Master Course of Excellence Euroculture: Europe in the Wider World. Furthermore, he is an external expert on Higher Education for the European Commission and has been involved in main initiatives to harmonise European Higher Education, such as the development of a European Credit Transfer and Accumulation System (ECTS) since 1988 and the Qualifications Framework for the European Higher Education Area and a European Qualifications Framework for LLL. He also chairs the Dutch team of experts for the implementation of the ‘Bologna Process’ in Dutch Higher Education institutions. Together with Julia Gonzalez (University of Deusto, Bilbao, Spain), Wagenaar elaborated, designed and coordinates the large scale innovative project Tuning Educational Structures in the World. In 2009 the project developed the Tuning-AHELO Conceptual Framework for Expected / Desired Learning Outcomes in Economics and Engineering on invitation of the OECD.

Peter Ewell

Peter T. Ewell is Vice President of the National Center for Higher Education Management Systems (NCHEMS), a research and development center founded to improve the effectiveness of colleges and universities. A member of the staff since 1981, Dr. Ewell’s work focuses on assessing institutional effectiveness and the outcomes of college, and involves both research and direct consulting with institutions and state systems on collecting and using assessment information in planning, evaluation, and budgeting. He has directed many projects on this topic, including initiatives funded by the W. K. Kellogg Foundation, the Lumina Foundation for Education, the Ford Foundation, the Consortium for the Advancement of Private Higher Education, and the Pew Charitable Trusts. In addition, he has consulted with over 375 colleges and universities, twenty-four state systems of higher education, and twelve countries on topics including outcomes, program review, enrollment management, and strategic planning.

Dr. Ewell has authored six books and numerous articles on the topic of improving undergraduate instruction through outcomes-based approaches. In addition, he has prepared commissioned papers for many agencies, including the Education Commission of the States, the National Governors’ Association, the National Conference of State Legislators, the National Center for Public Policy in Higher Education, and the Organization for Economic Cooperation and Development (OECD). Widely sought as a speaker, in 1985 he gave the keynote address for the first national conference on Outcomes in American Higher Education, and has since spoken widely on this topic at both U.S. and international conferences. In 1996 he designed the curriculum for the outcomes-based Western Governors University (WGU) and in 1998 he led the design team for the National Survey of Student Engagement (NSSE) now used by more than 1200 colleges and universities in the U.S.

A graduate of Haverford College, Dr. Ewell received his Ph.D. in Political Science from Yale University in 1976 and was on the faculty of the University of Chicago.
Mary Catharine Lennon

Mary Catharine Lennon is a Senior Research Analyst at the Higher Education Quality Council of Ontario (HEQCO), an arms-length research agency of the Ontario government, where she is currently leading projects on establishing and measuring learning outcomes.

She has been involved in higher education policy development, advice and research in institutional, provincial, inter-provincial and international educational agencies including the Council of Ministers of Education, Canada, and the Association of Commonwealth Universities.

She also studied comparative and international higher education issues and policies in both her Masters and Doctoral programs where she is currently a PhD Candidate at the University of Toronto.

With this background, the majority of her work extends towards international and comparative system level policy issues including System Design, Accountability, Quality Assurance and Governance.

She is currently responsible for HEQCO’s suite of student learning outcomes projects including piloting the Collegiate Learning Assessment, facilitating Tuning projects with three sectors of academic disciplines, and acting as the Canadian National Project Manager for the OECD’s AHELO feasibility study.

Mary Catharine is also co-editor of ‘Measuring the Value of a Postsecondary Education’, a collection of international examples of learning outcome activities.

Daniel Edwards

Daniel Edwards is a Principal Research Fellow in the Higher Education research program at the Australian Council for Educational Research. He was appointed in January 2008 to help develop ACER’s higher education research.

Dr Edwards’ research encompasses a range of educational issues, with particular emphasis on higher education. He has explored issues relating to demand for higher education places (both amongst students and employers), student achievement, student aspirations and pathways, selection policies for entrance to university, and educational ‘choice’ theories. He also has experience researching wider social issues regarding social stratification and demographic change.

Dr Edwards’ leads a range of projects providing advice and insight into policy and practice across the spectrum of higher education for governments, universities, stakeholder organisations and other bodies. Projects he has led include the National Research Student Survey (NRSS) project, the National Project Management of AHELO in Australia, development of the AHELO Contextual Dimension. Daniel also edits and manages the ACER Joining the Dots, a policy resource for Australian Higher Education.

Dr Edwards is also an Adjunct Research Fellow at the Centre for Population and Urban Research at Monash University.
Tsutomu KIMURA
Advisor to the Ministry of Education, Culture, Sports, Science and Technology

Education:
Ph.D. (Engineering) Tokyo Institute of Technology, 1968
M.A. (Civil Engineering) The University of Tokyo, 1964
B.A. (Civil Engineering) The University of Tokyo, 1961

Professional Experience:
- Professor Emeritus and Specially Appointed Professor, National Institution for Academic Degrees and University Evaluation (NIAD-UE) (Apr. 2009 – present)
- President, Japan Accreditation Board for Engineering Education (Jun. 2009 – present)
- Advisor to the Ministry of Education, Culture, Sports, Science and Technology (Apr. 2009 – present)
- President, NIAD-UE (1998 – 2009)
- Professor Emeritus, Tokyo Institute of Technology (Mar. 1998 – present)
- Professor, School of Engineering, Tokyo Institute of Technology (1997 – 1998)
- President, Tokyo Institute of Technology (1993 – 1997)
- Dean, School of Engineering, Tokyo Institute of Technology (1992 – 1993)
- Professor, School of Engineering, Tokyo Institute of Technology (1982 – 1993)
- Researcher, University of Cambridge (1978 – 1979)
- Researcher, University of Strathclyde (1971 – 1973)
- Assistant Professor, School of Engineering, Tokyo Institute of Technology (1968 – 1982)
- Assistant, School of Science and Engineering, Tokyo Institute of Technology (1965 – 1968)
- NIPPO Corporation (1961 – 1965)

Awards:
- The Order of the Sacred Treasure, Gold and Silver Star, 2013
- Japan Academy Prize, 2012
- Most Excellent Order of the British Empire, CBE, 2004
- Japan Society of Civil Engineers Award (thesis prize), 1988
- Japanese Society of Soil Mechanics and Foundation Engineering Award (thesis prize), 1982
- Seiichi Tejima Memorial Research Award (writing prize), 1969
- Japan Society of Civil Engineers Award (thesis encouragement prize), 1966

External Appointments:
- Special Appointment Member, Central Council for Education (2007 – present)
- Member, Science Council of Japan (2005 – present)
- Chairman, Tokyo Metropolitan Board of Education (2004 – present)
- Member, National University Evaluation Committee (2003 – 2004)
- Vice Chairman, Central Council for Education (2001 – 2007)
- Chairman, Science Education and Industrial Education Council (1997 – 1999)

Selected Papers and Publications:
- Soil Mechanics (Civil Engineering Series, Vol. 8), Shokokusya Publishing, 1980
- Diffusion of Soil Stress, Kajima Institute Publishing, 1978
Dr. Kikuo Kishimoto is currently a Professor of the Department of Mechanical Sciences and Engineering and Dean of School of Engineering, Tokyo Institute of Technology. He received his B.S. degree in 1975, M.S. degree in 1977, and Doctor of Engineering degree in 1982 from Tokyo Institute of Technology. He worked as a Research Associate and Associate Professor at Tokyo Institute of Technology from 1977 to 1995. During this period, he was a Visiting Scholar at Cambridge University from 1987 to 1988. He previously served as Vice President for education in 2012.

He has published over 250 journal papers in the areas of applied mechanics, fracture mechanics, reliability of microelectronic devices, and others. He has also published authoritative and widely used reference books in these fields. In 1980 and 2000, he was awarded the JSME Medal for Best Paper for his contributions to dynamic fracture mechanics and interfacial mechanics, respectively. He also received the Best Paper Award from Japan Society of Corrosion Engineering in 1993, Society of Materials Science Award for Academic Contribution in 2006 and JSME Materials and Mechanics Division Achievement Medal in 2007.

He is a fellow of Japan Society of Mechanical Engineers, a fellow of Society of Automotive Engineers of Japan and a fellow of American Society of Mechanical Engineers. He is a member of Science Council of Japan and serving as a chairman of mechanical engineering committee. He has served and chaired on various scientific and technological committees promoting research and education. He is a vice-president of Japan Accreditation Board for Engineering Education (JABEE).
Motohisa KANEKO
Professor, Research Center for University Studies, University of Tsukuba
Professor Emeritus, The University of Tokyo

Education:
- Ph.D. The University of Chicago, 1985
- M.A. The University of Tokyo, 1974
- B.A. The University of Tokyo, 1972

Professional Experience:
- Professor, Research Center for University Studies, University of Tsukuba (Current Position)
- Professor and Research Director, Center for National University Finance and Management
- Dean, Graduate School of Education & Faculty of Education, University of Tokyo
- Professor, Graduate School of Education, University of Tokyo
- Associate Professor, University of Tokyo
- Associate Professor, Research Institute for Higher Education, Hiroshima University
- Visiting Assistant Professor, State University of New York at Albany
- Consultant, World Bank
- Researcher, Institute of Developing Economies, Japan External Trade Organization

External Appointments:
- Member, Central Council for Education
- Member, Science Council of Japan
- President, Japanese Association of Higher Education Research-Advising professor for Peking University, East China Normal University, Fudan University, and University of Science and Technology of China

Selected Papers and Publications:
- Reconstructing College Education in Japan, Tamagawa University Press, 2013
- Universities’ Capacities to Teach, Chikuma-Shobo, 2007
- Graduate Education in Engineering in the Age of Globalization, Report from the Science Council of Japan’s engineering education meeting, 2003
- Education, Economy, and Society, Course material for the Open University of Japan, 1996
Satoko FUKAHORI
Senior Researcher, Department for Higher Education Research, National Institute for Educational Policy Research of Japan

Education:
MA in Comparative Education, Kyoto University, Graduate School of Education. (1993).
Ph.D. in Sociology of Education, Columbia University, Graduate School of Arts and Sciences -Teacher College. (2000).

Academic Appointments:
Assistant Professor, Institute of Social Science, The University of Tokyo.
Associate Professor, Department of Education, Kyoto Women’s University and Junior College.
Senior Researcher, Department for Higher Education Research, National Institute for Educational Policy Research.

Major External Appointments:
Expert Group Member for the Government Official Examination, National Personnel Authority.
Advisory Board Member for the Management of the Super Science High School Program at The Tokyo Gagugei University Senior High School.
Research Committee Member, The Japan Society of Educational Sociology; Editorial Board Member.
Editorial Board Member, Tuning Journal for Higher Education.
Globalization Project Member, Liberal and General Education Society of Japan.
Research Institute for Higher Education at Hiroshima University, Visiting Scholar.

Major Publications: