ASSESSMENT OF HIGHER EDUCATION LEARNING OUTCOMES

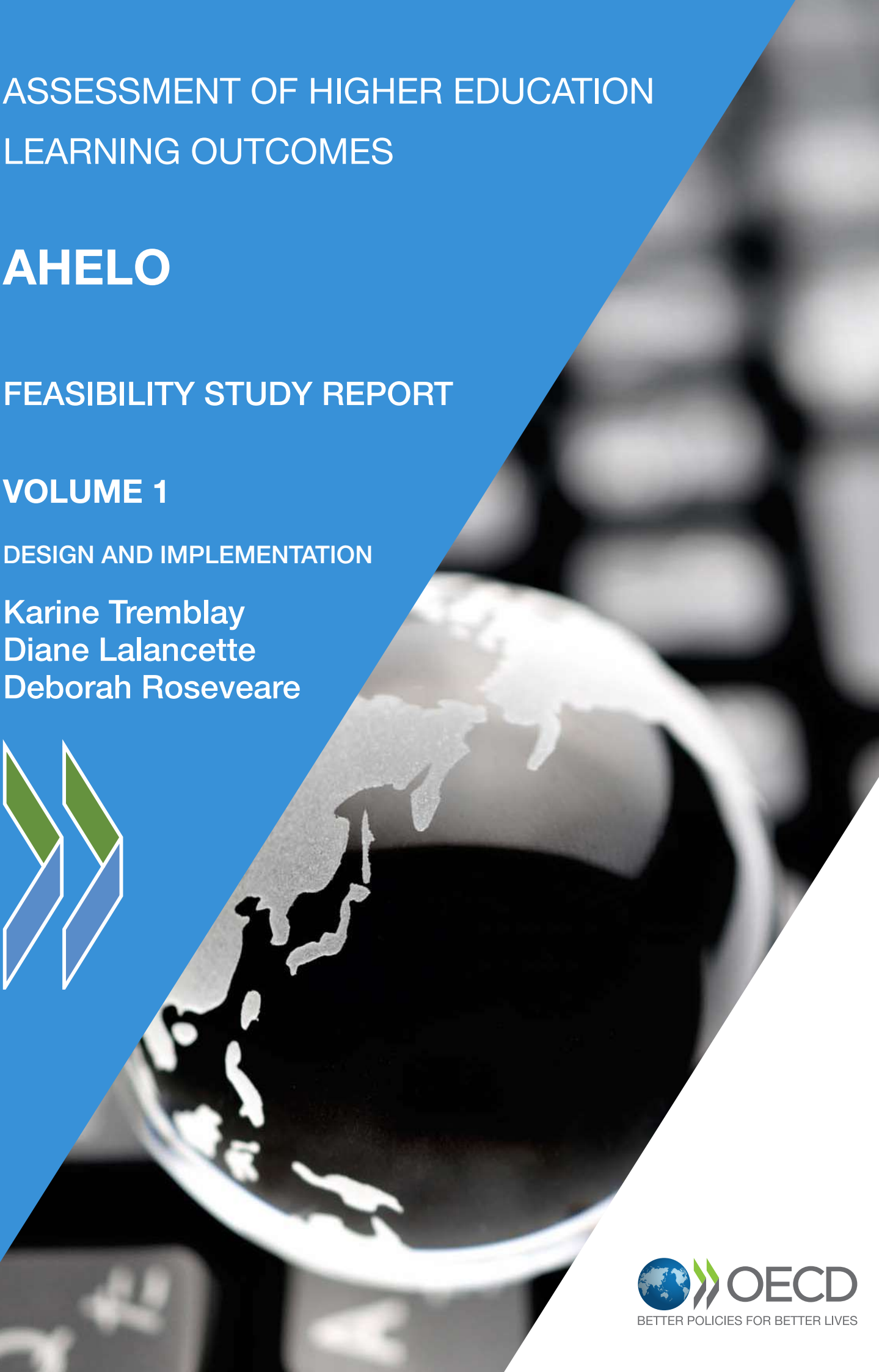
AHELO

FEASIBILITY STUDY REPORT

VOLUME 1

DESIGN AND IMPLEMENTATION

Karine Tremblay
Diane Lalancette
Deborah Roseveare



OECD
BETTER POLICIES FOR BETTER LIVES

Assessment of Higher Education Learning Outcomes

Feasibility Study Report

Volume 1 – Design and Implementation

Karine Tremblay
Diane Lalancette
Deborah Roseveare

© OECD 2012

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

© OECD 2012

ACKNOWLEDGEMENTS AND CREDITS

This report has been written by Karine Tremblay, Diane Lalancette and Deborah Roseveare of the OECD Directorate for Education.

The authors wish to acknowledge the extensive contributions from the AHELO Consortium report authors and contributors: Hamish Coates, Sarah Richardson, Yan Bibby, Stephen Birchall, Eva van der Brugge, Rajat Chadha, Steve Dept, Jean Dumais, Daniel Edwards, Thomas van Essen, Peter Ewell, Andrea Ferrari, Jacob Pearce, Claire Melican, Xiaoxun Sun, Ling Tan, Rebecca Taylor and Don Westerheijden.

The authors also wish to acknowledge the contributions of participating countries' National Project Managers and National Experts who drafted the boxes on national experiences.

Valuable comments on draft chapters were provided by members of the AHELO Group of National Experts and the Technical Advisory Group.

Special thanks are due to Leslie Diamond who edited the first drafts of this report, as well as to Cécile Bily who edited the final version, drafted the introduction, readers' guide and annexes, and prepared this report for publication.

Thanks are also due to the many other OECD colleagues who contributed to this project at different stages of its development including Barbara Ischinger, Andreas Schleicher, Richard Yelland, Fabrice Hénard, Valérie Lafon and Sabrina Leonarduzzi. Alenoush Saroyan also provided insightful feedback on the draft chapters during a sabbatical stay at the OECD. The authors are extremely grateful for her time and valuable input to the drafts. The AHELO feasibility study also benefited from the contributions of the following consultants, seconded staff and interns: Rodrigo Castañeda Valle, HoonHo Kim, Claire Leavitt, Eleonore Perez Duarte, Tupac Soulas, Takashi Sukegawa and Mary Wieder.

The Secretariat would also like to express its gratitude to the sponsors who, along with the participating countries, generously contributed to this project and without whom the AHELO feasibility study would not have been possible: Lumina Foundation for Education (USA), Compagnia di San Paolo (Italy), Calouste Gulbenkian Foundation (Portugal), Riksbankens Jubileumsfund (Sweden), the Spencer and Teagle Foundations (USA) as well as the higher Education Founding Council – HEFCE (England) and the Higher Education Authority – HEA (Ireland). The William and Flora Hewlett Foundation also provided support for U.S. participation in the study.

And finally a special word of thanks to Jan Levy, the Chair of the AHELO GNE, who provided invaluable guidance and support to the Secretariat throughout the feasibility study.

© OECD 2012

TABLE OF CONTENTS

INTRODUCTION 9

READERS' GUIDE 10

CHAPTER 1 – THE RATIONALE FOR AN AHELO: HIGHER EDUCATION IN THE 21ST CENTURY CONTEXT 15

Global trends in higher education 16
Expansion of higher education systems 16
Wider participation 18
Emergence of new players 19
More diverse profiles of institutions, programmes and students 20
Continuing advancement and rapid integration of new technology 22
Greater internationalisation 23
Increasing pressures on costs and new modes of financing 26
Growing emphasis on market forces: competition and signalling mechanisms 27
New modes of governance stressing performance, quality and accountability 29

The quality challenge and limitations of diverse attempts to fill the quality information gap 30
Concerns for drop-out and its implications for equity and efficiency 30
Insufficient information on higher education quality 32

The rationale for an AHELO 34
Moving beyond collegial approaches 34
Growing focus on student learning outcomes 35
Central emphasis on student centered learning and research on teaching-learning processes 37
AHELO within the broader movement towards competencies and learning outcomes 38

CHAPTER 2 – THE BEGINNING OF AHELO: DECISIONS AND CHALLENGES 55

Questioning feasibility: Key challenges in developing and operationalising an AHELO 57

Risk of ranking 57
Misuse of results 58
Complexity of comparisons across diverse institutions 59
Contesting of standardised tests 60
Potential impact for institutional autonomy and academic freedom 61
Relevance of generic skills in different systems 62
Methodological and practical questions 63

© OECD 2012

Initial expert meetings to frame a roadmap 64
Washington meeting 64
Paris meeting 67
Seoul meeting 69

CHAPTER 3 – DESIGN AND MANAGEMENT OF THE FEASIBILITY STUDY 77

Survey design 78
Study design emphasising a proof of the concept 78
Four distinct but coherent strands of work 80
Participating countries 82

Constraints and implications on the survey design 83
Implications of the global financial crisis and other constraints on the study 83
Closing the funding gap 84
Impacts of a phased approach on the AHELO Feasibility Study timeframe 86

Phases of work 88
Phase 1 instrumentation and initial proof of concept 88
Phase 2 - practical implementation and psychometric analysis of results 91
Phase 3 value-added methodologies and approaches 95

Study management and actors 96
International management 96
Expert groups involved 99
National and institutional coordination 100

CHAPTER 4 – INSTRUMENT DEVELOPMENT 105

The instrumentation process 106
Step A: Developing assessment frameworks 107
Step B: Developing assessment instruments 109
Step C: Translating and adapting assessment instruments and surveys 111
Step D: Small-scale validation of the assessment and survey instruments 111
Step E: Final review of the assessment and survey instruments 112

The AHELO feasibility study Generic Skills Assessment 112
Development of the Generic Skills assessment framework 113
Development of the Generic Skills instrument 114

The AHELO feasibility study Economics Assessment 116
Development of the Economics assessment framework 116
Development of the Economics instrument 118

The AHELO feasibility study Engineering Assessment 121
Development of the Engineering assessment framework 121
Development of the Engineering instrument 123

© OECD 2012

The AHELO feasibility study Contextual Surveys 128
Development of the Contextual dimension framework 129
Development of the Contextual dimension surveys 132

Localisation of assessment and survey instruments 135

CHAPTER 5 – IMPLEMENTATION 147

Management of AHELO implementations in countries and institutions 148

Selection of institutions and securing institutional engagement 151

Sampling of students and faculty 153
Student sampling 154
Faculty sampling 156

Electronic delivery 157
Study delivery mechanism and preparation 158

Response rates and incentives 161
Student response rates 161
Response rates and possible non-response bias 163
Student engagement and effectiveness of incentives 165
Faculty responses 172

Scoring student responses 173
Scorer technical standards and training 173
Scoring experience 174

Institution Context Instrument 175

CHAPTER 6 LESSONS LEARNT ON DESIGN AND IMPLEMENTATION 181

Purpose of AHELO – to support improvement in learning outcomes 182

Feasibility study design 183
Diversity 183
Higher education institutions as the unit of analysis 183
Engaging key stakeholders and experts 184
Assessment strands 184

Management of the feasibility study 184
Funding 185
Contracting arrangements 185
Timelines 186
Instrument development 186
Assessment frameworks 187

© OECD 2012

Matrix sampling 187
Multiple-choice questions and constructed response tasks 188
Contextual dimension 188
Instrument localisation 188

Field implementation 188
Timelines 189
Capacity, readiness and quality control 189
Electronic delivery 190
Sampling frames and student response rates 190
Scoring 191
Estimating costs of participation 192
Extra value for participating institutions and faculties 192

ANNEX A – LIST OF CONTRIBUTORS 197

ANNEX B – ILLUSTRATIVE ITEMS 219

Tables

Table 4.1 - Country localisation workflows 138
Table 5.1 - Key AHELO feasibility study implementation activities and timelines 150
Table 5.2 - Number of participating institutions per field implementation 152
Table 5.3 - Student response rates by strand and country 162
Table 5.4 - Comparison of population and response data (standard deviation) 164
Table 5.5 - Proportion of HEIs attaining the 50% response rate threshold, by strand 165
Table 5.6 - Institutional provision of incentives by strand and system 168
Table 5.7 - Types of incentives mentioned by ICs 168
Table 5.8 - HEIs characteristics by strand 176

Figures

Figure 1.1 - Trends in higher education enrolments worldwide, 1970-2025 17
Figure 1.2 - Growth in internationalisation of higher education (1975-2010, in millions) 24
Figure 1.3 - Proportion of students who enter higher education without graduating with at least a first degree at this level (2008) 31
Figure 3.1 - AHELO Feasibility Study strands of work 81
Figure 3.2 - Phases of the AHELO Feasibility Study 89
Figure 4.1 - Instrument development: from design to final review 106
Figure 4.2 - Developing frameworks 108
Figure 5.1 - AHELO feasibility study communications structure 148
Figure 5.2 - Comparison of male students in population and sample in 2 countries and their HEIs 164
Figure 5.3 - Institution response rate by incentive 171
Figure 5.4 - Effort put into taking the test by incentive 172

© OECD 2012

Boxes

Box 3.1 - Governance of the AHELO Feasibility Study 79
Box 3.2 - Countries finally participating in each skills strand 83
Box 3.3 - Contractual arrangements 87
Box 3.4 - AHELO Consortium partners 97
Box 4.1 - Australia - Students' feedback on the AHELO assessment instruments 128
Box 4.2 - Mexico - Students' feedback on the AHELO assessment instruments 135
Box 4.3 - Kuwait - Assessing students in a multicultural society 136
Box 4.4 - Country cooperation on translation 139
(Netherlands and Flanders, Colombia and Mexico)
Box 5.1 - Key players within countries 149
Box 5.2 - Italy - Engagement of institutions 153
Box 5.3 - A census approach in Abu Dhabi 156
Box 5.4 - IT strategy and preparation for AHELO tests in Egypt 158
Box 5.5 - The use of notebooks to limit IT preparation within institutions in the Slovak Republic 159
Box 5.6 - Student response rates in Norway and Finland 163
Box 5.7 - Student engagement strategies 166
(Slovak Republic, Colombia, Australia)
Box 5.8 - United States - Importance of an institutional culture of assessment 169
Box 5.9 - Colombia - Tight timelines 174
Box 5.10 - AHELO Scoring 175
(Finland, Australia)
Box 6.1 - Value to participating institutions and faculties 192
(Japan, Canada, Kuwait)

© OECD 2012

INTRODUCTION

In 2008, the OECD launched the AHELO feasibility study, an initiative with the objective to assess whether it is possible to develop international measures of learning outcomes in higher education.

Learning outcomes are indeed key to a meaningful education, and focusing on learning outcomes is essential to inform diagnosis and improve teaching processes and student learning. While there is a long tradition of learning outcomes' assessment within institutions' courses and programmes, emphasis on learning outcomes has become more important in recent years. Interest in developing comparative measures of learning outcomes has increased in response to a range of higher education trends, challenges and paradigm shifts.

AHELO aims to complement institution-based assessments by providing a direct evaluation of student learning outcomes at the global level and to enable institutions to benchmark the performance of their students against their peers as part of their improvement efforts. Given AHELO's global scope, it is essential that measures of learning outcomes are valid across diverse cultures and languages as well as different types of higher education institutions (HEIs).

The purpose of the feasibility study is to see whether it is practically and scientifically feasible to assess what students in higher education know and can do upon graduation within and across these diverse contexts. The feasibility study should demonstrate what is feasible and what could be feasible, what has worked well and what has not, as well as provide lessons and stimulate reflection on how learning outcomes might be most effectively measured in the future.

The outcomes of the feasibility study will be presented in the following ways:

- this **first volume** of the feasibility study Report focusing on the design and implementation processes;
- a **second volume** which will be published in February 2013 on data analysis and national experiences;
- the feasibility study **Conference** which will take place in Paris on 11-12 March 2013; and
- a **third** and final volume to be published in April 2013 on further insights (and which will include the conference proceedings).

© OECD 2012

READERS' GUIDE

Chapter 1 of the present report describes the general context which gave rise to the concept of an AHELO. It looks at the global trends in higher education and how the sector has evolved in last few decades and the challenges these present for the quality of higher education provision. It also points to the lack of data and information. It then describes how the concept of learning outcomes has gained importance and what current tools exist to assess higher education to date, thereby setting the scene for an AHELO.

Chapter 2 describes the early days of AHELO: the decision process and discussions which gave AHELO its original design as well as the early criticism received and the challenges which would have to be faced, and explains how these were taken into consideration.

Chapter 3 then presents the general design chosen for the feasibility study, how the financial constraints and the impact of the global financial crisis affected this design (including the phasing of the work). It also explains the management and running of the study at the national and international level, as well as the various groups involved in the decision-making or running of the feasibility study, and how they interact.

Chapter 4 goes into the detail of how the instruments for the assessment and contextual surveys were developed. It describes the typical instrument development process for large-scale international surveys/assessments and the importance of the various steps, before outlining the instrument development processes effectively adopted for the Generic Skills, Economics and Engineering assessments as well as for the Contextual Dimension surveys. Finally, it provides an overview of how the localisation process (i.e. translation and adaptation of the instruments) was undertaken for each of these instruments.

Chapter 5 details the process of implementation, i.e. the actual testing of the AHELO instruments in real-life conditions. It starts by providing an overview of the management of the AHELO feasibility study implementations in countries and institutions, and how participating institutions were selected. It also describes the approach to student and faculty selections and the sampling design adopted for the feasibility study, and the process of electronic delivery of the AHELO tests. A discussion of observed response rates and the effectiveness of student engagement strategies followed. Finally, the chapter describes the process for scoring constructed-response tasks in a consistent way.

Chapter 6 concludes with the lessons learnt from the feasibility study with respect to the design and implementation of an AHELO.

Annex A acknowledges all contributors to the feasibility study.

Annex B provides a list of illustrative items used in the actual test.

© OECD 2012

Note on terminology

The AHELO feasibility study involved the participation of 17 higher education systems. In most cases, participation was at the national level although a number of systems also participated in the feasibility study at the regional, provincial or state levels. This was the case for Abu Dhabi (United Arab Emirates), Belgium (Flanders), Canada (Ontario), and the United States (Connecticut, Missouri and Pennsylvania). For simplicity and ease of reading, all higher education systems are referred to as "countries" or "participating countries" in the report, irrespective of the national or sub-national level of participation.

Abbreviations used in the report

AACC	American Association of Community Colleges
AAC&U	Association of American Colleges and Universities
AASCU	American Association of State Colleges and Universities
AAU	Association of American Universities
ACE	American Council on Education
ACER	Australian Council for Educational Research
AERA	American Educational Research Association
AHELO	Assessment of Higher Education Learning Outcomes
AMAC	Australian Medical Assessment Collaboration
AMK	Ammattikorkeakoulu – Finnish institution of higher education comparable to a university of applied sciences
APA	American Psychological Association
APEC	Asia-Pacific Economic Cooperation
ATAV	Adaptation, Translation And Verification
BA	Bachelor of Arts
BMD	Bachelor-Master-Doctorate (degree structure)
CAE	Council for Aid to Education
cApStAn	Linguistic Quality Control Agency
CHEPS	Centre for Higher Education Policy Studies
CLA	Collegiate Learning Assessment
CPR	Indiana University Center for Postsecondary Research
CRT	Constructed-Response Task *Within the AHELO feasibility study, different types of constructed-response items were used entailing different types of responses (short and extended responses, performance tasks, etc.). For simplicity within the Report, constructed response items take the abbreviation of a constructed-response task, or CRT.*
DIF	Differential Item Functioning
ECTS	European Credit Transfer and Accumulation System
EDPC	Education Policy Committee
EHEA	European Higher Education Area
EI	Education International
EQF	European Qualifications Framework
ETS	Educational Testing Service

© OECD 2012

EU	European Union
EUA	European University Association
EUGENE	European and Global Engineering Education academic network
FCI	Faculty Context Instrument
GDP	Gross Domestic Product
GNE	Group of National Experts
GRE	Graduate Record Examination
HEI	Higher Education Institution
IAU	International Association of Universities
IC	Institution Coordinator
ICC	Item Characteristic Curves
ICI	Institution Context Instrument
IDP Australia	International Development Programme
IEA	International Association for the Evaluation of Educational Achievement
IEA DPC	International Association for the Evaluation of Educational Achievement Data Processing and Research Center
IMHE	OECD Higher Education Programme (formerly Programme on Institutional Management in Higher Education)
IMHE GB	IMHE Governing Board
INES	OECD's Indicators of Education Systems (framework)
IRT	Item Response Theory
ISCED	International Standard Classification of Education
IUT	Institut Universitaire de Technologie
JSON	JavaScript Object Notation
LEAP	Liberal Education and America's Promise
LS	Lead Scorer
MA	Master of Arts
MAPP	Motivational Appraisal of Personal Potential
MCQ	Multiple Choice Question
MOOC	Massive Open Online Courses
MSC-AA	Medical Schools Council Assessment Alliance
NAEP	National Assessment of Educational Progress
NAICU	National Association of Independent Colleges and Universities
NASULGC	National Association of State Universities and Land-Grant Colleges
NC	National Centre
NCME	National Council on Measurement in Education
NIER	National Institute for Educational Policy Research
NILOA	National Institute for Learning Outcomes Assessment
NPM	National Project Manager
NSSE	National Survey of Student Engagement
OECD	Organisation for Economic Co-operation and Development
PIAAC	OECD Survey of Adult Skills (formerly Programme for International Assessment of Adult Competencies)
PISA	OECD Programme for International Student Assessment
PPP	Purchasing Power Parity
PWB	Programme of Work and Budget

© OECD 2012

QAA	Quality Assurance Agency for Higher Education
SCG	Stakeholders' Consultative Group
SCI	Student Context Instrument
STEM	Science, Technology Engineering and Mathematics
TA	Test Administrator
TAFE	Technical And Further Education
TAG	Technical Advisory Group
TALIS	OECD Teaching and Learning International Survey
TECA	Tertiary Engineering Capability Assessment
TRP	Technical Review Panel
UAT	User Acceptance Testing
UCTS	UMAP Credit Transfer Scheme
UIS	UNESCO Institute for Statistics
UMAP	University Mobility in Asia and the Pacific
UNDP	United Nations Development Program
UNESCO	United Nations Education Science and Culture Organization

© OECD 2012

CHAPTER 1

THE RATIONALE FOR AN AHELO: HIGHER EDUCATION IN THE 21ST CENTURY CONTEXT

This chapter first provides a brief overview of the global trends that have shaped the development of higher education over the past half century. It then describes the emergence and growing emphasis in the policy debate of a quality challenge in the last few decades. Finally, it looks at the shift of focus in assessment and the growing recognition of the importance of learning outcomes, trends which, taken together, illustrate a paradigm shift for higher education.

© OECD 2012

The sections below describe the global trends that have shaped wide-ranging mutations in higher education, as well as the growing focus on the quality of higher education. Together, these patterns have shaped the context of the current OECD AHELO initiative.

Global trends in higher education

There is widespread recognition that skills and human capital have become the backbone of economic prosperity and social well-being in the 21st century. In contemporary knowledge-intensive economies and societies, individual and societal progress is increasingly driven by technological advances. Prosperity requires nations to retain their competitive edge by developing and sustaining a skilled workforce, maintaining a globally competitive research base, and improving the dissemination of knowledge for the benefit of society at large.

In this context, higher education represents a critical factor in innovation and human capital development and plays a central role in the success and sustainability of the knowledge economy (Dill and Van Vught, 2010). Hence, higher education has become increasingly important on national agendas and has undergone profound mutations and reforms worldwide over the past decades, as portrayed in a recent OECD review of tertiary education policies (OECD, 2008). As stated by Altbach et al., "an academic revolution has taken place in higher education in the past half century marked by transformations unprecedented in scope and diversity" (Altbach et al., 2009).

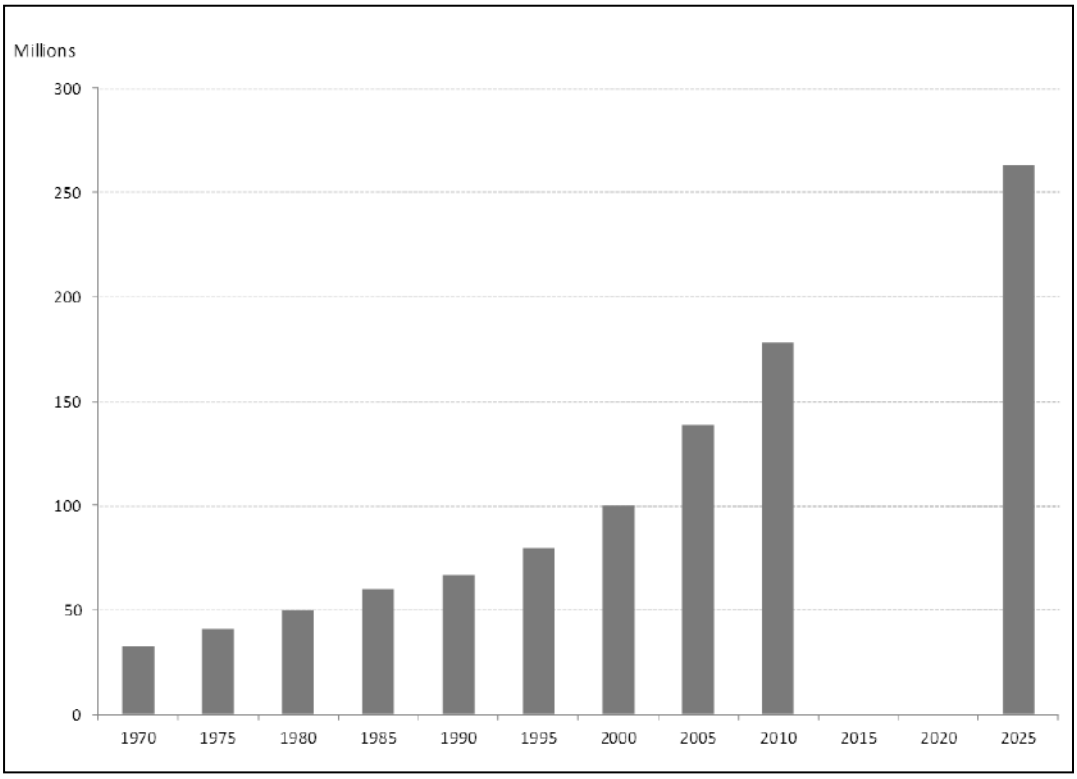
As recent as 40 to 50 years ago, higher education essentially referred to the traditional research universities. This picture is entirely different today. Several trends have contributed to reshaping the model of the collegial[1] and "ivory tower" university attended by the elite. Today HEIs are more diversified and are closer to a patchwork model attended by larger segments of the population. Thus, higher education today is characterised by massive expansion and wider participation; the emergence of new players; more diverse profiles of HEIs, programmes and their students; broader adoption and more integrated use of communications and educational technologies; greater internationalisation, competition and signalling mechanisms; growing pressures on costs and new forms of financing; as well as new modes and roles of governance, including increasing emphasis on performance, quality and accountability.

Expansion of higher education systems

In the last half century, the most salient of these trends is undoubtedly the dramatic expansion of higher education worldwide, as depicted in Figure 1.1. In 1970, the UNESCO Institute for Statistics (UIS) estimated that there were roughly 32.5 million students enrolled in higher education worldwide. In the year 2000, this estimation increased to nearly 100 million and in 2010 to 178 million. This translates into 4.3% average annual growth in tertiary enrolment, a very rapid growth when compared to the 1.6% average annual growth in the world population over the same period (UNDP, 2012). Figure 1.1 also reveals an accelerating expansion starting in the mid-1990s, with a 5.9% average annual growth of higher education enrolments in the first decade of the 21st century. The number of higher education students is forecast to further expand to reach 263 million by 2025 (British Council and IDP Australia, cited in Davis, 2003 and Daniel, 2009).

© OECD 2012

Figure 1.1 - Trends in higher education enrolments worldwide, 1970-2025



Source: UNESCO Institute for Statistics Data Centre for 1970-2010 and Daniel (2009) for 2025 forecast.

Growth has prevailed on all continents and constitutes a defining feature of global trends of the late 20th and early 21st centuries (Guri-Rosenblit et al., 2007). There are many underlying factors. First and foremost the public demand for higher education has soared, fuelled by growing upper-secondary completion rates. Additional factors include social mobility expectations, growing female participation, as well as democratisation and urbanisation processes and independence movements in the developing world. The shift towards post-industrial economies has also affirmed that an educated workforce is essential for economic development and has heightened the demand for white-collar workers in the public sector and service industries. Finally, the accelerated pace of technological change has further stimulated access to and participation in higher education (Gibbons, 1998; Schofer and Meyer, 2005 and Altbach et al., 2009).

Higher education participation has expanded in stages across countries and world regions. Altbach et al. (2009) note that the United States and Canada were first to achieve mass higher education in the 1960s, followed by Western Europe and Japan in the 1980s. This trend then spread towards emerging regions. The growth in tertiary enrolments over the past four decades was more obvious in emerging regions, notably Sub-Saharan Africa (8.4% average

© OECD 2012

annual growth), the Arab states (7.4%), East Asia and the Pacific (7%), Latin America and the Caribbean (6.4%). and South and West Asia (6%). More recent trends suggest that the greatest growth is now taking place in South and East Asia. China and India alone will account for over half of the global increase in student numbers in the years to come (Kapur and Crowley, 2008). Moreover, by 2020, they will account for 40% of young adults (aged 25-34) with a tertiary degree (OECD, 2012a).

Wider participation

The growth in absolute numbers of students is mirrored by trends in access to higher education. In just 14 years, the proportion of young adults entering undergraduate university programmes[2] has soared by 25 percentage points, from 37% in 1995 to 62% in 2010. Meanwhile, rates for those entering more vocationally-oriented programmes[3] have remained stable, at 17% (OECD, 2012b).

Comparable trend data are not available to examine changes in higher education participation over a longer period. It is possible, however, to capture the progress achieved indirectly, by comparing the attainment rates among different age groups. In this type of analysis the proportion of young adults currently holding a tertiary degree – i.e. aged 25-34 years in 2010 – is compared with those who completed their studies 30 years earlier – i.e. aged 55-64 years in 2010. The analysis shows that across the OECD, the proportion of tertiary degree holders has grown from 23 to 38% over three decades (OECD, 2012b).

Canada, Japan and Korea have already reached higher education attainment rates of over 50% and this is becoming the benchmark for OECD countries. Several OECD countries have indeed set ambitious goals, as illustrated by the European Commission's target of 40% higher education attainment among European younger generations by 2020, which 11 Member states have already surpassed[4] (European Commission, 2010; Roth and Thum, 2010). Likewise, President Obama's administration has established an ambitious goal of 60% attainment rates among 25-34 year-old Americans in 2020, so that *"*America will once again have the highest proportion of college graduates in the world*"* (Obama, 2009; Kanter 2011). Among non-OECD G20 countries, the Russian Federation has also achieved over 50% higher education attainment while China aims towards a 20% target by 2020 (China Government, 2010) and some leading Indian analysts call for 20-25% participation rates in the near future (Levy, 2008) and 30% by 2020 (Eggins and West, 2010).

A key feature behind this wider participation is the increasing female participation in higher education. According to data from the UNESCO Institute for Statistics, women made up 41% of higher education enrolments worldwide in 1970. They achieved parity with men in 2005 at the global level (despite some world regions lagging behind), and now slightly outnumber them with about 51% of global enrolments (UIS, 2011). The latest data from the OECD's *Education at a Glance* underlines that this trend is more marked within OECD countries, with significantly higher entry rates for women relative to men, both in undergraduate university programmes (69 vs. 55% on average) as well as vocationally-oriented programmes (19 vs. 16%). In 2010, women also reached parity with men with regard to access to advanced research programmes, at entry rates of 2.8% for both (OECD, 2012b).

© OECD 2012

Emergence of new players

As a corollary to the massification of access and participation over the past half century, higher education systems have experienced an increase in higher education providers, with a burgeoning of new HEIs established across the globe to respond to growing demand. As a matter of fact, the majority of HEIs operating today were established in the past century (Guri-Rosenblit and Sebkova, 2004).

To illustrate the magnitude of change, the Indian higher education system consisted of 27 universities and 695 colleges at the time of independence in 1949 (Agarwal, 2009). By 2006, the number of HEIs had sky-rocketed to 361 universities and thousands of colleges (Kapur and Crowley, 2008). This trend is unlikely to abate with the further growth in total enrolments projected for the next 15 years. To put figures in perspective, Daniel (2009) estimates that meeting the demand for an additional 85 million students worldwide between 2010 and 2025 will require accommodating an additional 109 000 students – i.e. the equivalent of four large universities – every week!

In many countries, the massification process has led to the emergence of new types of institutions within higher education, as alternatives to traditional universities. The growth of a strongly employer-oriented non-university sector, closely integrated with the labour market needs of each locality and region, is indeed one of the most significant structural changes in recent times for higher education systems (Grubb, 2003; OECD, 2005a). Within the OECD, this movement started in France in the mid 1960s with the creation of *Instituts Universitaires de Technologie* (IUTs) based on the model of some of the United States' vocationally-oriented junior and community colleges, followed in the early 1970s by Technical and Further Education Colleges (TAFE) in Australia, *Fachhochschulen* in Germany and *Distriktshøgskoler* in Norway. In the late 1970s, Portugal set up Polytechnic Institutes while the Netherlands created its *Hogescholen* (HBO) in the late 1980s. The 1990s saw the emergence of the Polytechnic sector (AMK) in Finland, the *Universidades Tecnológicas* in Mexico and the Swiss Universities of Applied Sciences. Finally, the *Universidades Politécnicas* and *Universidades Interculturales* emerged over the past decade in Mexico (OECD, 2008).

These new HEIs were often established to create training opportunities for mid-level professionals needed for post-industrial and increasingly knowledge-intensive economies (Altbach et al., 2009). By offering shorter programmes, they were better able to meet growing demands at a manageable cost (Trow, 1974). They were also able to respond to increasingly diverse needs of the labour market and regional development (Kyvik, 2004), and to accommodate the growing diversity of individual students' motivations, expectations and career plans (Goedegebuure et al., 1994).

Assessing the relative importance of each sector is not straightforward as there is no exhaustive register of HEIs worldwide, new providers are being established almost on a weekly basis, and the non-university sector is far from being homogenous. Nevertheless, the latest estimates of the UNESCO-affiliated International Association of Universities (IAU) list over 17 000 HEIs worldwide (IAU, 2013).

© OECD 2012

Several countries have also relied extensively on private providers to meet the growing demand for higher education, resulting in a massive expansion of the number of private HEIs. This trend has been most prevalent wherever there has not been a tradition of public funding of higher education, or resources have been limited to accommodate any additional demand through public higher education (Levy, 2008; Bjarnason et al., 2009). The fastest-growing systems have also been those in which private provision has expanded most rapidly. For instance, the private sector in India, which accounted for just 15% of the seats of engineering colleges in 1960, rose to nearly 87% of seats by 2003 (Kapur and Crowley, 2008). In Latin America, the past two decades have also seen the growing privatisation of higher education to balance resources with the need to satisfy increasing demand (Segrera, 2010).

In 2008, Gürüz estimated that the private sector accounted for some 30% of enrolments worldwide (Gürüz, 2008). However, this average reflects diverse country-specific realities. Within the OECD, Chile, Japan and Korea have the largest private university sectors with fewer than 30% of students enrolled in public HEIs. Mexico, Poland, Portugal and the United States also have sizeable private sectors operating with dominant private funding in the university sector, while Estonia, France, Norway and Switzerland have significant non-university private sectors (OECD, 2012b). Outside the OECD, the largest private sectors are found in Asia (Indonesia, Philippines, Malaysia) and Latin America (Brazil), and to a lesser extent in some post-communist countries (Altbach et al., 2009). Private enrolments are likely to expand further in the years to come given the large projected increases in higher education participation in China and India, and the reliance of these countries upon private providers to absorb excess demand.

More diverse profiles of institutions, programmes and students

A related trend is the growing diversity of higher education student bodies, HEIs and their educational offerings. This diversification is generally assumed to offer major advantages to the various stakeholders in higher education systems, like better addressing students' needs, enabling higher levels of higher education attainment, improving social mobility, better serving the needs of the labour market, increasing political legitimisation and more effective higher education systems (Birnbaum, 1983; Huisman, 1995; Van Vught, 2008).

A corollary of the shift from elite to mass – and now even universal in some countries – higher education access and participation is the growing heterogeneity of students in terms of their socio-economic background, academic ability and preparedness, career expectations, motivation and engagement. This diversification reflects the increasing social demand for tertiary education and the subsequent greater participation (OECD, 2008). Besides the rise of female participation, another prominent development is the growing participation of more mature students in search of a first degree, in pursuit of their studies after a period in the workforce, or attending while working in order to update or upgrade their skills.

HEIs today include an increasing number of non-traditional students, "those who did not enter directly from secondary school, are not from the dominant social groups in terms of gender, socio-economic status or ethnic background, or are not studying in a full-time, classroom-based mode" (Schuetze and Slowey, 2002). Increasingly, they are serving first-generation higher

© OECD 2012

education students seeking social mobility, representatives of economic or social minorities as part of equity policies, adults and lifelong learners looking to update/upgrade their skills, etc. These diverse student bodies each have their own constraints and challenges to overcome. HEIs are thus required to adapt their provision so as to respond to all needs and help all students thrive.

Student demands are also changing. Learners increasingly seek courses that enable them to update their knowledge throughout their working lives. In addition, as learners seek to acquire particular knowledge or skills to satisfy labour market needs, more and more prefer to pick and choose courses from the most suitable providers, rather than studying a traditional clearly defined programme at one institution (OECD, 2008).

As a result, there is ever more diversity within institutions. For instance, traditional universities are increasingly expanding their offer to include short-cycle courses and more vocationally-oriented degrees and vice versa. Modes of delivery have also considerably expanded. Indeed, the traditional mode of full-time and campus-based attendance is ill-suited to the needs of adults and lifelong learners who often undertake their studies while working and supporting a family. The development of more flexible ways of provision such as distance learning and e-learning has improved access to a wider range of student populations and contributed to meeting increasingly diverse demand (OECD, 2005b). These are also seen as more cost-effective alternatives to traditional modes of delivery (Salmi, 2000). New technologies have also brought about changes in approaches to teaching, especially at the under-graduate level, with standardised courses often delivered online, allowing for different use of classroom time with more small seminars and interactive discussions, and greater time spent with students on their individual projects. Finally, HEIs have started to extend their lifelong learning offerings. The organisation of learning is increasingly adapting to include: the assessment of prior learning; a wider range of programmes; part-time learning; module-based curricula and credit systems; competence-oriented, student-centred organisation of studies; and the provision of non-degree studies and continuing education (Schuetze and Slowey, 2002).

Higher education institutions have not only become more diverse in type, ownership and educational offerings, they have also diversified their missions, targeting specific groups of students (women, minorities, disadvantaged or students with special needs, adults and lifelong learners, international students, etc.), serving specific local or regional needs, specialising in some niche areas, or establishing close links with specific industries/corporations (Altbach et al., 2009). This suggests that HEIs, over time, have assumed responsibility for a far wider range of occupational preparation than in the past. Altogether, it has resulted in a strong institutional differentiation to meet the needs of increasingly diverse audiences.

However, this differentiation process has not unfolded clearly and without ambiguities. For instance, in some countries the more vocationally-oriented sector has faced an academic drift whereby seeking legitimacy as fully-fledged universities, despite formal institutional differentiation (Van Vught, 2008).

© OECD 2012

Continuing advancement and rapid integration of new technology

Communication and education delivery technologies are continuing to advance at accelerating rates. These advancements have had and will continue to have significant impact on the organisation and provision of higher education both within countries and worldwide (Johnson et al., 2012). This presents challenges for higher education in all countries including keeping pace with rapid advances in communications and social networking technologies; accommodating the increased costs of technology into existing mechanisms for financing higher education; and taking full advantage of the educational opportunities these technologies provide to expand student access and improve their success in higher education.

Many HEIs and programmes have successfully adapted and used a succession of technological advances in recent decades, including technology-assisted open universities, non-classroom-based modes of instructional delivery, and computer modelling and simulation as instructional tools. "Blended" instruction in which classroom time is augmented through internet-based student-faculty interaction or student-to-student networking is now the norm in many HEIs and programmes. Yet, research suggests that these steps are only early innovations in the transformation of both instruction and learning and that greater potential can be realised through the integration of technology (Norway Opening Universities, 2011; Johnson et al., 2012).

Internet-based online instructional delivery is now the fastest growing type or sector of higher education in many countries (Elaine Allen and Seaman, 2011). The recent and rapid emergence of Massive Open Online Courses (MOOCs) can potentially provide access to advanced courses taught by top faculty to hundreds of thousands of students. This has opened doors to even greater opportunities and at the same time has introduced new challenges for higher education. Several MOOC networks using similar technologies, and led by some of the world's leading universities, provide course-by-course access to students worldwide, raising questions about degree credit and credentialing or degree-granting for such course completion both within and across countries (OECD, 2007).

Online delivery of education is also expanding rapidly to meet the career-specific education and training needs of adult populations. While such educational opportunities, including many at the sub-degree or certificate level, are increasingly important for social advancement and economic development, they are often not effectively accommodated within traditional higher education governance, financing and quality control mechanisms.

Advances in data collection and warehousing technologies have given rise to additional opportunities to monitor and track student learning individually and longitudinally across as well as within HEIs and systems of higher education (Prescott and Ewell, 2009; Garcia and L'Orange, 2010). Advances in data analytics have also helped identify the impediments that students encounter and they offer considerable potential for conducting formative assessments of learning that can be helpful to both students and instructors.

© OECD 2012

Greater internationalisation

The internationalisation of higher education also features among the sector's key transformations in the past 25 years, especially in the European context. International activities and exchanges have long been bound to research – despite signs of student and academic mobility in medieval European universities – while teaching and learning remained essentially nationally-based (Scott, 2000). But internationalisation has widened in scope over the past three decades, and is now an essential dimension of national and institutional strategy and policy (OECD, 2008).

Internationalisation can be defined as *"*the process of integrating an international, intercultural or global dimension into the purpose, functions or delivery of tertiary education*"* (Knight, 2003). Although student and academic mobility are clearly the most observable features of internationalisation, they are not the only aspects. The internationalisation process manifests itself in various ways, and encompasses the full spectrum of educational programmes and activities that contribute to internationalised learning, ranging from the internationalisation of programmes' content and delivery to the mobility of students and scholars, in addition to intermediate forms of trans-national education such as the cross-border mobility of HEIs and/or their programmes. Another major form of internationalisation relates to the growing convergence of tertiary education systems (e.g. Bologna process), and curricula in some disciplines (Bennell and Pierce, 2003; Altbach, 2004).

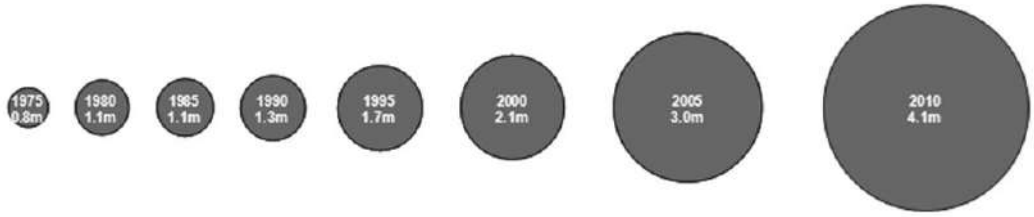
The internationalisation process has evolved in response to several trends. First, as world economies become increasingly inter-connected, international skills have become ever more important for operating successfully on a global scale. This has led to growing demands to incorporate an international dimension into education and training. Meanwhile the acceleration in global economic integration has hastened the internationalisation of the labour market for the highly skilled, and internationally recognised qualifications have risen in importance in some sectors (Bennell and Pierce, 2003; Peace Lenn and Campos, 1997). The rise of the new economy has provided additional stimulus since employers in OECD countries increasingly need to look abroad for talent as their own graduates are insufficient to replace those going into retirement. Internationalisation has thus become a part of a longer term skill development strategy. Demographic trends have also triggered internationalisation. In some OECD countries, faced with decreasing domestic enrolments after unprecedented expansion in tertiary education provision in the 1980s, internationalisation is increasingly seen as a way to compensate losses and ensure the viability of some HEIs. By contrast, in many emerging countries internationalisation offers a cost-effective alternative to national provision to increase capacity and meet growing demand on short notice (Van der Wende, 2001). The match between the insufficient capacity of emerging countries and the oversupply of some OECD tertiary education systems has been facilitated by the trend towards deregulating higher education in many OECD countries enabling the emergence of new educational offerings, including distance learning and cross-border operations of HEIs. Lastly, the emergence and rapid expansion of higher education export industries in some OECD countries has heightened awareness on the trade value of internationalisation from a macroeconomic perspective (OECD, 2008).

© OECD 2012

All forms of internationalisation have developed and grown in importance over the past three decades. Student mobility is an important and relatively well-documented aspect of this growth (Tremblay, 2002), but other forms of internationalisation have also gained momentum.

Figure 1.2 - Growth in internationalisation of higher education (1975-2010, in millions)

Long-term growth in the number of students enrolled outside their country of citizenship



Data on foreign enrolment worldwide comes from both the OECD and the UNESCO Institute for Statistics (UIS). UIS provided the data on all countries for 1975-95 and most of the non-OECD countries for 2000, 2005 and 2010. The OECD provided the data on OECD countries and the other non-OECD economies in 2000 and 2010. Both sources use similar definitions, thus making their combination possible. Missing data were imputed with the closest data reports to ensure that breaks in data coverage do not result in breaks in time series.

Source: OECD and UNESCO Institute for Statistics, published in *Education at a Glance 2012.*

International student mobility has increased tremendously over the past three decades, from 0.8 million students worldwide in 1975 to 4.1 million in 2010 (Figure 1.2). This trend has been accelerating in recent years, driven by large increases in student mobility from China, India and European countries in particular. Growth is projected to continue in the future to reach approximately 5.8 million around 2020 (Böhm et al., 2004) and 8 million by 2025 (Altbach and Bassett, 2004). A noteworthy development is the new European mobility strategy launched at the last Bologna Ministerial conference in Bucharest which sets the specific target of 20% of graduates in Europe to have studied or been trained abroad by 2020 (Bologna Secretariat 2012). While not representative of global trends, the Bologna developments are nevertheless important drivers of student mobility given the geographic scope of the Bologna process.

Unlike student mobility, data are scarce on the international mobility of academic staff. This is further complicated by the multiple forms of mobility, from short-term moves of a few days/weeks to longer movements of one year or more. Available evidence suggests that the main internationalisation of faculty consists of short-term leave, exchange visits and research collaborations (Enders and de Weert, 2004). The proportion of academics involved in longer-term exchanges is considerably lower than it is for short stays abroad. Marginson and Van der Wende (2007a) consider that it is not clear that longer term academic migration is increasing, although mobility to the United States is the exception. Other reports also stress limited longer-term academic mobility in Europe, although the United Kingdom, Scandinavia and the Netherlands seem more open in their recruitment (Mora, 2004; Musselin, 2004; Jacobs and Van der Ploeg, 2006).

© OECD 2012

The internationalisation of curricula gained momentum in the past decade with greater emphasis on the teaching of foreign languages and broader international perspectives in the substantive content of programmes and curricula (Van der Wende, 2001; Harman, 2006).

Another significant trend relates to the profound changes in the organisation and structure of national higher education systems to improve their transparency and inter-operability. This phenomenon has been most evident in Europe with the Bologna Process aimed at establishing a European Higher Education Area (EHEA) and enhancing the comparability and compatibility of higher education structures and degrees in Europe by 2010 (Bologna Secretariat, 1999). The Bologna Process is far-reaching, insofar as a number of non-EU countries have endorsed the Bologna declaration and joined its convergence process, to reach 47 participants spread geographically between Iceland, Portugal, Turkey and the Russian Federation.

The Bologna declaration proposed to adopt a system of easily readable and comparable degrees based on a common degree structure, often referred to as the BMD[5] structure. Several authors have noted the resemblances between the Bologna degree structure and the American model, hence broadening the convergence movement to intercontinental scale and global relevance (Douglass, 2006; Tapper and Palfreyman, 2005).

In addition, the Bologna declaration engaged signatory countries to develop instruments to translate and recognise credits and qualifications earned elsewhere, including in other countries. A major development has been the establishment of the European Credit Transfer and Accumulation System (ECTS), a student-centred system based on the student workload required to achieve the objectives of a programme in terms of the learning outcomes and competencies to be acquired. Meanwhile, the Diploma Supplement was developed as a follow-up tool for the implementation of the Lisbon Recognition Convention (Council of Europe, 2005). This is intended to enhance transparency and to facilitate academic and professional recognition of higher education qualifications.

Similar developments are taking place in other world regions with the development of the University Mobility in Asia and the Pacific (UMAP[6]) Credit Transfer Scheme (UCTS) to promote university student mobility in the Asia Pacific region (Mongkhonvanit and Emery, 2003). To the extent that the key players in international education are involved in either the ECTS or UCTS schemes, the influence of ECTS and UCTS on other countries' practices is likely to increase in the future. The Diploma Supplement also extends beyond Europe and has been piloted in Australia (OECD, 2008).

The past 15 years have also seen the emergence and growing development of off-shore delivery of education by HEIs. Australia and the United Kingdom pioneered this movement in the 1990s, but the United States has also become a major force in this area and a number of other countries – especially European countries – have joined this trend and, since 2000, set up campuses abroad and other trans-national arrangements (OECD, 2008, OECD, 2004; McBurnie and Ziguras, 2001). Given the obstacles as well as the risks faced by HEIs in setting up campuses in foreign countries, the majority of trans-national operations take the form of joint programmes with foreign HEIs through franchise arrangements with a partner institution in the

© OECD 2012

international students' home country. More recently, virtual HEIs which operate exclusively online have emerged.

Recent trends suggest that although programme and institution mobility is not yet that important, these forms of internationalisation are expected to outpace the delivery to international students onshore.

Increasing pressures on costs and new modes of financing

Another prominent trend of higher education over the past few decades relates to the growing pressure of its cost, and the adoption of new modes of financing in many countries.

The first phenomenon of rising costs is a direct consequence of the expansion of higher education systems and wider participation, which have increased the financial burden of higher education as most countries have tried to expand their systems while limiting the adverse impact on unit costs and expenditure to maintain quality. Indeed higher education provision offers limited scope for economies of scale. At the aggregate level, for the 25 OECD countries with trend data, the cost of higher education has risen from 1.3 to 1.5% of GDP between 1995 and 2009. Moreover, unit costs have also increased since 2000 by 9% on average across the OECD (OECD, 2012b).

In addition, a second phenomenon of fiscal pressure to curb costs has arisen in many countries. Indeed, economic growth over the past two decades has been insufficient to sustain the rising costs of higher education resulting from massification in most countries across the globe (Altbach et al., 2009; Sanyal and Johnstone, 2012). This mismatch has put increasing pressure on public budgets, especially in those countries with a strong tradition of public financing of higher education (e.g. most of Europe). Many countries have thus adopted new modes of financing over the past 15 years to foster cost-sharing. In 2011, *Education at a Glance* reported that more than half of the 25 countries with available information had, since 1995, undertaken system reforms of tuition fees and financial support for students, and most had introduced or increased students' contribution to the cost of their higher education (OECD, 2011). As a result, among OECD countries with trend data the public share of higher education expenditure has decreased from 78% in 1995 to 73% in 2009 (OECD, 2012b).

Overall, reforms of higher education finances over the past 15 years have focused on three main directions, although the relative importance of each depends on the specific country (OECD, 2008):

- First, there has been a diversification of funding sources. This reflects, in part, an overall trend of shifting the cost burden to students and away from public subsidies through greater contributions by students and their families. Private resources have also been mobilised through the commercialisation of research and other private uses of institutional facilities and staff.
- Second, the allocation of public funding for tertiary education is increasingly characterised by greater targeting of resources, performance-based funding, and competitive procedures. The basis for allocating core funding to HEIs has become more output-oriented. In a number of countries, formulas to allocate public funds to

© OECD 2012

HEIs are now related to performance indicators such as graduation or completion rates. Research funding has also increasingly been allocated to specific projects through competitive processes rather than block grants. A number of countries have also linked the allocation of research funds to assessments of research quality.

- Third, a number of countries have expanded their student financial support systems. In some countries, loans have gained in importance relative to grants in overall financial aid packages. Repayable types of aid have also increased in some countries.

More recently, the global financial crisis and the current economic downturn in the world economy have further fuelled concerns for the sustainability of higher education funding models. While costs of provision have continued to increase, students and systems have had less money to invest. Eggins and West (2010) suggest that recent trends in the financing of higher education are likely to exacerbate in the years to come:

> After many years of running hard to stay still, and taking advantage of all the readily available means of cutting costs and raising income, a tipping point is being reached in many systems, where more radical steps are unavoidable. […] Once it is accepted that states will no longer be able to fund universities at previous levels, it follows that other sources must be found. With them come new stakeholders, whether fee-paying students, alumni, business, professional practices, city authorities or regional economic development agencies. […] Public accountability, along with service to the community, will remain valued aspects of HE provision. The surest route to success will be to identify a unique mission and pursue it tenaciously.

Growing emphasis on market forces: competition and signalling mechanisms

Competition and signalling mechanisms have also become more prominent in the higher education sphere over time, as a corollary to the growth in educational offerings, private provision, rising costs, fiscal pressure and the internationalisation of higher education. These mechanisms were triggered by a number of different factors:

- First, since the 1980s, there has been a rise of New Public Management approaches to public services provision in many OECD countries which has put ever more emphasis on market mechanisms and principles borrowed from the private sector. In particular, leadership, incentives and competition among public sector agencies and private providers have been promoted to enhance the outcomes and cost-efficiency of public services (Parker and Gould, 1999; Marginson and Van der Wende, 2007; Lebeau et al., 2012). Instruments used in this context range from competitive research grants, contract research, performance-based funding formulas for teaching and learning activities, and public funding based on the number of students (OECD, 2008).

- A number of countries have also strengthened market mechanisms with a view to reducing the reliance of HEIs on public funding in a context of tight education budgets (Kaiser et al., 1999). HEIs have indeed been under growing pressure to diversify their revenues over the past two decades, and market mechanisms have

© OECD 2012

been introduced or reinforced with this specific goal in mind, e.g., through policies on tuition fee deregulation, policies on marketing institutional research, etc.

- Competition has also been advocated as a tool to improve teaching quality, as well as institutions' responsiveness to the needs of society, the labour market and students (Kaiser et al., 1999). The underlying assumption is that students would "vote with their feet" if dissatisfied with the quality or relevance of service provision in a given institution. Likewise, relevance to the needs of the local community and labour market becomes critical as soon as HEIs compete with each other for additional sources of funding.
- Meanwhile, the emergence of new providers in the public or private sectors has been possible thanks to deregulation policies in many countries which have lifted entry barriers to the higher education market. But in a context of imperfect information on teaching quality, these new players have often suffered from a bad reputation and their inability to demonstrate a track record on their outcomes (Altbach et al., 2009). They have thus placed increased emphasis on signalling mechanisms to ascertain their value and build their legitimacy to attract the best students.
- The asymmetric information problem is even more acute whenever a border is crossed. Building and maintaining an international reputation is an essential part of institutions' and systems' strategies to attract international students and faculties (Nelson, 2003; Douglas and Edelstein, 2009). Indeed, surveys of prospective international students confirm that reputation is a key driver of mobility flows (Kemp et al., 1998; Pimpa, 2005).

Signalling is an integral aspect of these increasingly competitive markets. Indeed, higher education has many features of an experience good, i.e. a service whose characteristics such as quality are difficult to observe prior to consumption (Nelson, 1970). With such information asymmetry, consumers struggle with their consumption choices and reward reputation in the absence of more transparent measures of quality. Accordingly, HEIs as well as systems use a range of signals and marketing tools to showcase their worth and build their reputation. For instance, Kwieck (2001) observes that HEIs in the Western world are increasingly borrowing marketing methods from the private sector. The emergence of global rankings and their growing importance is illustrative of this compelling search for signals and recognition (Van Vught and Ziegele, 2012). In reference to British business schools, Wilkins and Huisman (2012) assert that they operate in a market-driven environment and rankings are very much part of that environment to the extent that they have a significant impact on the ability of schools to attract top scholars, the most able students, and research funding.

The emergence of the first global ranking of research universities in 2003, and the burgeoning of global rankings ever since, have provided an unprecedented opportunity for HEIs and systems to rely on these external and transparent instruments to support claims on their quality and win high-stakes reputational capital (Shanghai Jiao Tong University, 2003; Usher and Savino, 2006; Rinne, 2008). While rankings are less of a concern among the current elite world and historical HEIs, they garner more attention from recently-established HEIs within the

© OECD 2012

emerging global marketplace, especially so in the case of research universities. Moreover, rankings draw much media and public attention. In this context, they have started to guide the strategic behaviours of university leaders, governments, students and employers (Hazelkorn, 2008; Van Vught and Ziegele, 2012). Nowadays, HEIs just like systems increasingly use rankings to support claims of excellence and "world class" status (Wilkins and Huisman, 2012; Song and Tai, 2007).

Marginson (2009) notes in this respect that the global referencing of research universities has gathered prominence in just half a decade, which Salmi and Saroyan (2007) attribute to the absence of a single global quality assurance agency, giving ranking systems the role of a quality regulator for international students. Rinne (2008), by contrast, expresses disquiet with the notion that nation states are losing their power to define standards and to control the key features of educational selection. But as Marginson contends, the external drivers sustaining the momentum for rankings and other comparisons of countries and HEIs are so powerful that "never again will higher education return to the state of innocence prior to the Internet and the Jiao Tong" (Marginson, 2009).

Rankings' key contribution is their focus on transparency. Rankings underline the principles of powerful signalling tools, i.e. the measurement of real outputs rather than reputation, and transparent and accurate data collection (Marginson, 2009; Van Vught and Ziegele, 2012).

However, this is not to say that global rankings have not been contested. Several authors have stressed their biases, noting that the resulting perceptions of reputation are all too often based upon imperfect proxies of quality such as input factors or the research performance of HEIs (Siow, 1997; Dill and Soo, 2005; Griffith and Rask, 2007; Salmi and Saroyan, 2007; Hazelkorn, 2008; Van Vught, 2009). Others have offered suggestions to improve them (Swail, 2011) and have designed new conceptual approaches (Van Vught and Ziegele, 2012). And yet some also denounce their distorting effects (Hazelkorn, 2007).

New modes of governance stressing performance, quality and accountability

In parallel with the above trends, the governance and steering of higher education has also evolved in the last two decades, spurred on by the growing number of HEIs, the diversification of their missions, student bodies and roles, and the rise of New Public Management principles as well as private providers (Dobbins et al., 2011). Overall, the growing complexity of higher education systems has made central oversight increasingly inadequate, and most countries in which it was prevalent have engaged in reforms to revisit higher education steering mechanisms (Dobbins et al., 2011).

Reviewing these steering mechanisms has generally involved a reappraisal of the purposes of higher education and the setting by governments of new strategies for the future. It has also entailed granting greater autonomy and room for manoeuvre to HEIs to encourage them to be more responsive to the needs of society and the economy, but with clearer accountability to governments and society at large to ensure quality and cost-efficient operation (OECD, 2008).

Institutional autonomy has been widened – e.g. by authorising HEIs to be established as legal persons (foundations, not-for-profit corporations) and allowing private providers. The trade-off

© OECD 2012

is that HEIs have become increasingly accountable for their use of public funds and are required to demonstrate "value for money" (OECD, 2008). They are under pressure to improve the quality of their teaching and research despite decreasing resources due to mounting funding constraints. Important changes have occurred to address these new challenges through renewed academic leadership and new ways of organising the decision-making structure. Academic leaders are increasingly seen as managers, coalition-builders or entrepreneurs (Askling and Stensaker, 2002) while governing bodies composed of internal and external stakeholders have been established to provide strategic guidance and steering.

Performance indicators and external quality evaluations are an integral aspect of the new model of distant steering, and a number of countries have introduced some degree of performance-based funding (Borden and Banta, 1994; OECD, 2008). In this context, in recent years there has been a growing appetite for performance and quality measures from both higher education public and private stakeholders, as well as from the HEIs themselves.

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

Concerns for drop-out and its implications for equity and efficiency

In the context of massive expansion of higher education systems and wider participation, there are persistent concerns related to the quality and relevance of students' preparation for higher education. These concerns stem from the greater heterogeneity of students than in the past, with respect to their abilities or expectations, and the multiplication of new HEIs – including some rogue providers – as discussed above. Of particular concern to policy makers is the magnitude of non completion, often perceived as a waste of financial and human resources. Indeed, in spite of the adoption and development of sophisticated quality assurance systems in most OECD countries over the past two decades, failure and inefficiencies in the learning process have not been eradicated (OECD, 2008).

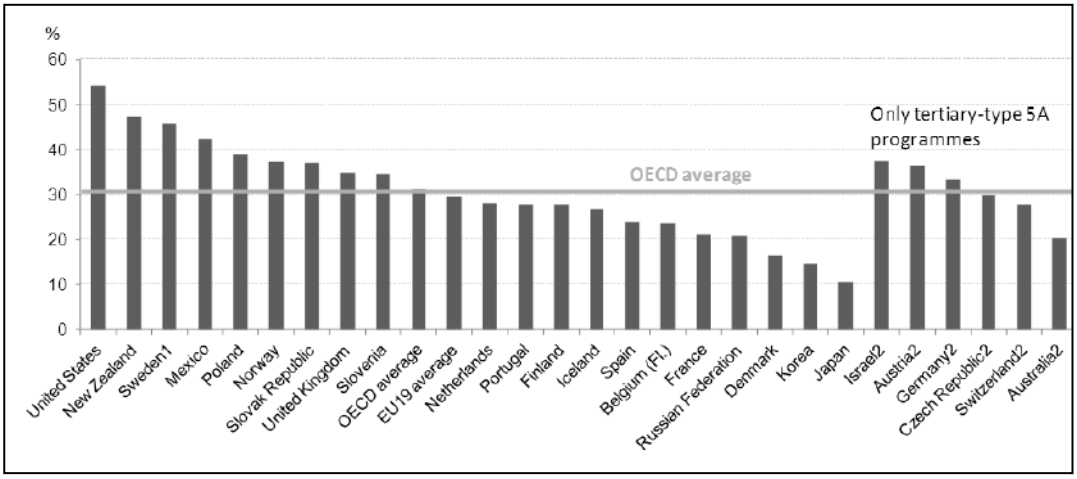
As depicted in Figure 1.3, on average three out of ten students entering a higher education programme of some kind in OECD countries will drop out without obtaining at least a first degree (OECD, 2010). These figures, however, should be interpreted cautiously:

- First, because most national data collections do not track students across HEIs. The number of dropouts therefore includes some students transferring to and graduating from other HEIs.
- In addition, research shows that there are sizable economic returns from higher education even when degrees are not achieved, so completion rates are not all that matters (National Research Council, 2012).

© OECD 2012

Figure 1.3 - Proportion of students who enter higher education without graduating with at least a first degree at this level (2008)

On average, in the 18 OECD countries for which data are available, some 31% of tertiary students enter tertiary education without graduating from a programme equivalent to this level of education



1. Some of the students who have not graduated may be still enrolled, or may have finished their education at a different institution than the one they started at, like in the United States.
2. Includes students entering single courses who may never intend to study all courses needed for a degree.
3. Tertiary-type A only.

Source: OECD. *Education at a Glance* 2010.

Even with these caveats in mind, drop-out and underachievement incur economic costs as a result of lower returns to non-degree higher education compared to full degrees, while costs per student are the same in both cases – at USD 13 700 per year on average across the OECD in 2009[7] (OECD, 2012b). This raises questions about the scope for improving the "productivity" of higher education through targeted policies to enhance the quality of service provision, and as a consequence, students' retention and success. Some countries are indeed concerned with inefficiencies in their systems, including not only high student drop-out rates, but also excessive time to completion, programme duplication, programme under-enrolment, and insufficient use of cross-institution collaboration (see OECD, 2008 for examples).

The social costs of underachievement are equally high given the social outcomes of higher education and its contribution to social mobility and equity (d'Addio, 2007; Blanden et al., 2007). Available evidence suggests that disadvantaged students are not only less likely to aspire to higher education studies and to access higher education programmes, but they are also more prone to dropping out, thereby reinforcing inequality (Mateju et al., 2003; Koucky et al., 2008, OECD, 2008; Brock, 2010). Given these social costs and consequences, efforts to improve student completion and institutional productivity must be carefully undertaken so that they do

© OECD 2012

not further inhibit access and success for sub-populations already underrepresented in higher education.

Insufficient information on higher education quality

Ever more stakeholders have an interest in higher education quality and outcomes as a result of the shift from "elite" to "mass" – and in certain countries to "universal" – systems of higher education. Consequently, greater interest and scrutiny – and hence greater transparency – is a natural corollary of this growth. This issue was taken up at the Meeting of OECD Education Ministers held in Athens in June 2006. Ministers committed their countries to the goal of raising the quality of higher education:

> At our meeting, we agreed on a new task: to go beyond growth, by making higher education not just bigger but also better.
>
> […] Ministers recognised that key stakeholders – including students, families and governments – must have better information about topics such as quality and cost to make decisions and hold HEIs accountable for their performance.
>
> […] In Particular, measuring the quality of higher education outcomes is needed both to justify the allocation of public resources and the effectiveness with which they are used by increasingly autonomous HEIs, and also to pursue enhancements in the quality and relevance of educational outcomes more broadly and systematically, so that HEIs serve economies and local communities effectively (Giannakou, 2006).

This affirmed accent on quality does not underestimate the progress achieved already in this area. In fact, the strong development of quality assurance systems is one of the most significant developments of higher education since the early 1980s. Quality assurance expanded in response to the massification of participation, the growing diversity of educational offerings and the expansion of private provision (El-Khawas et al., 1998, Dill and Beerkens, 2010). While traditional, often informal quality assurance procedures were suited to systems with a small number of HEIs and students, they are no longer adapted to contemporary complex and diversified systems. Van Vught and Westerheijden (1994) note that the development of quality assurance also responded to cost pressures: "Budget-cuts and retrenchment operations automatically lead to questions about the relative quality of processes and products in higher education." Finally, increased market pressures also fostered greater focus on accountability in higher education. In the United States for instance, higher education has become more consumer-driven with students and parents resisting tuition hikes and calling for more accountability with regard to quality and cost-effectiveness of HEIs.

Yet, despite the rapid development of quality assurance systems and their growing sophistication in the last two decades, little is known of their actual impact (OECD, 2008). Indeed, the overarching goal of quality assurance processes is to ensure that minimum standards are met and to improve the quality of higher education outcomes over time. In this context, impact is the yardstick by which to gauge the effectiveness and success of the systems put in place.

© OECD 2012

There is evidence of effects on academics' behaviours and management within HEIs, but scepticism surfaces in a number of studies examining the impact of quality assurance on teaching and learning, (Christensen, 2010; Ewell, 2010; Harvey 2006; Newton 2000, 2001), although some authors are more optimistic and report a more concrete impact of evaluations on teaching practices (Brennan, 1997; Silva et al., 1997). Likewise, some of the countries who participated in the OECD Thematic Review of Tertiary Education offered some evidence of the positive impact of quality assurance mechanisms on the quality of teaching and learning, although this impact was measured in terms of the number of negative accreditation evaluations, student satisfaction ratios or the acceptance and recognition of HEIs – i.e. by very rough proxies of impact on student learning. In fact the OECD Review concluded with a recommendation to increase the focus on student outcomes to alleviate this problem (OECD, 2008).

In assessing higher education student outcomes, policy makers face a considerable information gap. Ewell (1999) notes the remarkable development worldwide of performance indicators in higher education. The most commonly used indicators are completion rates and time to completion indicators, drop-out rates, especially after the first year, and graduation rates as well as destinations and employment rates of graduates in specific fields of study. These performance indicators provide a valuable information base to understand and monitor the performance of higher education at institutional level. However, for other aspects, and most notably for the learning outcomes of higher education, while it would appear to be among the most important pieces of information on higher education, available data remains scarce in many systems (see Nusche, 2007, for a summary of existing assessments).

There are proxies of higher education quality in addition to quality assurance systems and existing performance indicators, but each has significant limitations:

- There is strong evidence that students still base their enrolment decisions largely on the reputations of HEIs, particularly when engaging in international mobility. Yet, reputations are highly subjective and tend to reinforce existing hierarchies. Marginson (2009) notes for instance that the peer survey reputation proxy used in the Times Higher/QS ranking has only a 1% return rate and tends to over-represent the former British Empire countries.
- Likewise, there has been extensive research on the limitations of international rankings to understand aspects of quality that are most relevant to prospective students (Hazelkorn, 2008; Rauhvargers, 2011). Indeed, to the extent that rankings apply a transparent and uniform methodology to a large number of HEIs and/or programmes, they are constrained by data availability (Van Vught and Ziegele, 2012). This results in a bias towards available data on inputs and research excellence which may be adequate to capture research excellence but is ill-suited to provide an accurate picture of teaching and learning quality in various HEIs and/or programmes (Dill and So, 2005; Astin and Lee, 2003).
- Student surveys and student satisfaction surveys are commonly used to capture the quality of teaching and learning (OECD, 2008; Kuh, 2009; McCormick, 2009; Radloff

© OECD 2012

and Coates, 2009; Radloff, 2011). They can indeed provide valuable information to HEIs when applied at national level, providing comparative insights on their strengths and weaknesses as well as an opportunity to benchmark their performance with the achievement of others on a number of criteria. But their applicability across national boundaries is more delicate, especially for international satisfaction surveys whose cultural sensitivity is well-documented. In addition, satisfaction is not a measure of learning.

- Other proxies used or advocated to capture the quality of teaching and learning are students' self-reports on their learning gain (Douglass et al., 2012). Recent research suggests however that students' self-reports on their learning are poorly correlated with objective measures of learning, which calls into question their use in studies of higher education quality (Bowman, 2010).
- Finally, graduate surveys are commonly used to provide insights into the success of graduates in the competitive labour market as well as the adequacy of provision in relation to the needs of employers. While appealing and useful from a qualitative standpoint, the use of labour market outcomes as proxies for higher education quality is problematic to the extent that labour market outcomes of higher education graduates are highly sensitive to conjuncture and local economic conditions. Their use in comparative quality assessments could result in a bias against HEIs located in economically-depressed areas or preparing students for occupations in sectors facing a difficult conjuncture.

In this context, a growing need has emerged to develop direct measures of learning outcomes to overcome the limitations of current proxies for the quality of teaching and learning.

The rationale for an AHELO

The proposition to explore the development of an international AHELO emerged during the Athens Meeting of OECD Education Ministers in 2006 (Giannakou, 2006). This concept was put forth at a time of great pressure to develop better performance metrics in higher education, whether to support accountability reporting requirements or to assist leaders, teachers and learners to understand and position their work in increasingly complex environments. This idea illustrates the broader trends reshaping higher education as it shifts into the 21st century. Indeed, the AHELO proposition is illustrative of a paradigm shift for higher education, which manifests itself in several ways.

Moving beyond collegial approaches

As described in the above sections, the expansion and mutations of higher education have brought about a shift of emphasis from collegial approaches of horizontal governance by communities of scholars towards models combining greater autonomy with top-down features such as increased transparency and accountability.

This shift has been examined extensively. In the United States, Liu (2011a) observes that issues related to improved higher education performance and effective accountability have received

© OECD 2012

unprecedented attention from stakeholders[8]. Hursh and Wall (2011) deplore the growing pressure on HEIs to adopt neoliberal principles (privatisation, entrepreneurship, standardisation, assessment, and accountability) which, in their view, undermine the historical purposes of higher education and reduce faculty autonomy. Segrera (2010) makes similar observations in the Latin American and Caribbean contexts and asserts that while the traditional values of universities are still valid (autonomy, academic freedom, research, students' work, assessment), they should be viewed within the context of new global norms. Finally in Europe, Bologna developments are also illustrative of this shift. Within this framework, however, a deliberate attempt is made to reconcile accountability with institutions' autonomy and academic freedom, as expressed in the Bucharest Communiqué of Bologna Ministers in 2012:

> […] We confirm our commitment to maintaining public responsibility for higher education and acknowledge the need to open a dialogue on funding and governance of higher education. We recognise the importance of further developing appropriate funding instruments to pursue our common goals. Furthermore, we stress the importance of developing more efficient governance and managerial structures at HEIs. We commit to supporting the engagement of students and staff in governance structures at all levels and reiterate our commitment to autonomous and accountable HEIs that embrace academic freedom (Bologna Secretariat, 2012).

The growing emphasis on transparency and accountability that characterises the new paradigm has led to increased demands for colleges and universities to engage in outcomes assessment for accountability purposes (Secolsky and Denison, 2011). As a matter of fact, the increase in the various global ranking tools can be seen as a reaction to this rising demand for transparency (Van Vught, 2009; Van Vught and Ziegele, 2012).

While the assessment of learning outcomes has traditionally been an internal matter for HEIs, there are now signs that traditional collegial approaches are no longer sufficient. Indeed, internal assessments are often tailored to be as locally relevant as possible. With the shift towards more universal and internationally-oriented higher education systems, using internal assessments to yield more general information might, without some external check, spark concerns about grade inflation. This is not to say that localised assessments of student learning do not remain important, but recent trends show the emergence of more generalisable forms of assessment as a complement, such as common examinations, shared item libraries, the application of graduate tests or professional licensing examinations (for instance, GAMSAT, 2012; Ideal Consortium, 2012; MSC-AA, 2012; USMLE, 2012).

Given its international scope, the AHELO concept reflects the most advanced manifestation of a generalisable test that could potentially provide independent insights into learners' knowledge, capacity, and their ability to apply knowledge and skill to solve real-world problems.

Growing focus on student learning outcomes

A second paradigm shift relates to the conception of higher education outcomes. Increasingly, the focus is moving away from input-based conceptions (number of classes taken, study time

© OECD 2012

and student workload) towards outcome-based notions of higher education throughput, as described by Chung (2011) in the context of engineering education:

> Under the impact of globalization and the coming of the Information Age, there is a paradigm shift occurring in the engineering curriculum and academic structure. Apart from the creation of new programs for the emerging fields in engineering, the approach and orientation have also been shifted from objective-based/input-based education to outcome-based education. The criteria for the new generation of quality engineering graduates have been much broadened (Chung, 2011).

This shift has been most evident in Europe where the Bologna Declaration of 29 European ministers of education in June 1999 stated as a key objective for Europe to establish a European Higher Education Area (EHEA) by 2010 and committed to write all higher education modules and programmes in terms of learning outcomes by that date (Bologna Secretariat, 1999). The Bologna Declaration has now been endorsed by 47 countries around the world. In addition, many countries outside the Bologna Process are aligning their higher education systems to be Bologna-compatible so as to facilitate description of qualifications, mutual recognition of degrees, and student mobility (Kennedy, 2008). Ten years down the road, Reichert (2010) praises the visionary goals of using learning outcomes and competencies as the structuring principle of all curricula in Europe, but laments that only few countries and HEIs have embraced this approach. It is thus no surprise that Bologna ministers have recently reaffirmed the importance of moving forward in the implementation of a learning outcomes' approach:

> […] To consolidate the EHEA, meaningful implementation of learning outcomes is needed. The development, understanding and practical use of learning outcomes is crucial to the success of ECTS, the Diploma Supplement, recognition, qualifications frameworks and quality assurance – all of which are interdependent. We call on institutions to further link study credits with both learning outcomes and student workload, and to include the attainment of learning outcomes in assessment procedures. We will work to ensure that the ECTS Users' Guide fully reflects the state of on-going work on learning outcomes and recognition of prior learning.
>
> […] We welcome the clear reference to ECTS, to the European Qualifications Framework and to learning outcomes in the European Commission's proposal for a revision of the EU Directive on the recognition of professional qualifications. We underline the importance of taking appropriate account of these elements in recognition decisions (Bologna Secretariat, 2012).

Meanwhile, a similar shift is also underway across the Atlantic (Johnstone et al., 2002). Loris (2010) refers to a new vision for college learning at the beginning of the second decade of the 21st century in the United States context, and describes the Liberal Education and America's Promise (LEAP) initiative launched by the Association of American Colleges and Universities (AAC&U) to outline the essential learning outcomes that contemporary college students need to master in both general education and the major. In addition to and related to LEAP, there is also growing interest among U.S. HEIs and States in applying European-based "Tuning"

© OECD 2012

practices (described below) within a number of major academic disciplines, and in adopting Degree Qualification Profiles, a U.S. version of Bologna-based degree frameworks, as part of accreditation or other quality assurance and public accountability provisions (Lumina Foundation, 2011).

With higher education systems being increasingly interconnected and operating on a global scale, such a paradigm shift is not without implications for student mobility, transferability of credentials and degree recognition. A growing body of research over the past 10 to 15 years has examined the implications of this shift towards learning outcomes at both national and international levels:

- defining desired learning outcomes across institutional and national boundaries and in various disciplines (the “Tuning” process described below);
- integrating learning outcomes perspectives in quality assurance processes; and
- measuring learning outcomes, first at national level (Nusche, 2007) and then across borders with the AHELO initiative.

Central emphasis on student centered learning and research on teaching-learning processes

In 1995, Barr and Tagg launched a thought-provoking discussion in higher education circles by stating that there was a paradigm shift in American undergraduate education. In their view, higher education is shifting from an ”instruction paradigm” – characterised by an emphasis on delivering lectures and providing students with the means to learn – towards a ”learning paradigm” in which the emphasis is no longer on the means but on the end, i.e. supporting the learning process of students:

> A paradigm shift is taking hold in American higher education. In its briefest form, the paradigm that has governed our colleges is this: A college is an institution that exists to provide instruction. Subtly but profoundly we are shifting to a new paradigm: A college is an institution that exists to produce learning. This shift changes everything. It is both needed and wanted.
>
> We are beginning to recognize that our dominant paradigm mistakes a means for an end. It takes the means or method-called "instruction" or "teaching" - and makes it the college's end or purpose. […] We now see that our mission is not instruction but rather that of producing learning with every student by whatever means work best (Barr and Tagg, 1995).

Associated with the move towards a learning paradigm, the dominant pedagogy has also shifted to a learner-centred focus (Cornelius-White, 2007; Weimer, 2002). There is some evidence that faculties have started embracing the principles of a learning-centred philosophy in the United States and are willing to change their practices to adopt new classroom strategies (Webber, 2012; Scott et al., 2009). In the Asia-Pacific region by contrast, Cheng (2009) reports more difficulties in implementing effective reforms and stresses the importance of teachers’ management and professional development (Cheng, 2009). Learner-centred approaches are

© OECD 2012

also prominent within European policy agendas, as affirmed in the Bucharest Bologna Communiqué:

> […] We reiterate our commitment to promote student-centred learning in higher education, characterised by innovative methods of teaching that involve students as active participants in their own learning. Together with institutions, students and staff, we will facilitate a supportive and inspiring working and learning environment (Bologna Secretariat, 2012).

A corollary of this strong emphasis on student-centred learning is to better understand the interplay between teaching and learning so as to identify effective teaching strategies and test new ideas to enhance students' learning outcomes (American Psychological Association, 1997; McCombs and Miller, 2007). Research scrutiny on teaching-learning processes is also prompted by pressures on costs which call for improving efficiency of higher education provision and heighten the need for evidence-based research on effective teaching and learning strategies.

In reviewing systematic research findings on the effectiveness of various interventions designed to help at-risk students remain in college, Brock (2010) shows that some student-centred programmes and interventions can improve student outcomes. But Liu (2011b) stresses the importance of standardised outcomes assessment for the evaluation of instructional effectiveness of higher education. In this respect, the AHELO initiative has the potential to open a range of new research avenues by providing objective measures on learning outcomes, as well as a wealth of contextual and background information. It is therefore fully in line with the new paradigm of student-centred learning.

AHELO within the broader movement towards competencies and learning outcomes

While AHELO is the first international attempt at measuring higher education student learning outcomes across borders, languages and cultures, it is by no means unique or isolated. On the contrary, it is part of a broader context of distinct initiatives converging in their focus on performance, competencies and learning outcomes. This final section briefly outlines the various forms of this movement and related international initiatives before situating AHELO within this complex field. In framing its relevance, it is indeed useful to review how AHELO is situated alongside related international developments.

In an effort to develop cross-institutionally generalisable information on educational outcomes, HEIs and systems have turned to proxy information. Accordingly, a first group of initiatives focus on the collection of data to be used as indirect proxies of outcomes and performance:

- Student engagement and satisfaction surveys

  While student surveys abound in higher education, particularly those that measure "satisfaction", only a few have sought to move beyond the bounds of single systems and to measure learning as opposed to satisfaction with provision (AHELO Consortium, 2011). The data collection on student engagement is probably the most sophisticated international contextual assessment to date. It was initiated in 1999 by Indiana University through the United States' National Survey of Student Engagement (NSSE), and has been replicated since 2006 in Australia, Canada, China, Japan,

© OECD 2012

Mexico, New Zealand and South Africa. Rather than directly assessing learning outcomes using a test, the assessment of student engagement records the extent to which students have participated in defined learning activities and the amount of teacher and institutional support they received in doing so (Kuh, 2009). These data have stimulated important discussions in several countries about learning processes, contexts, and outcomes. Such collections can provide useful insights, however the data collected are subjective in nature and focused on educational processes rather than learning outcomes.

- Rankings

  The proliferation of global rankings illustrates the need for systems, HEIs and individuals to gain access to internationally comparable data and information on university achievement. Prominent initiatives include the Times Higher Education World University Rankings (TSL Education, 2011), the Academic Ranking of World Universities (Shanghai Jiao Tong University, 2011), QS World University Rankings, and the US News and World Report Best Colleges (US News and World Report, 2011). With numerous others, these developments have driven a "rankings movement" that has considerably sharpened focus on data-driven cross-institutional comparison. The impact and visibility of these rankings emphasises the demand for this type of information, but it also brings to fore the limitations of these tools. Both the rankings and the discourse that surround them are a direct trigger for AHELO inasmuch as they highlight a need to compensate for the shortcomings of existing metrics.

- U-Multirank

  By far the most significant attempt to overcome many limitations of prevailing rankings is the U-Multirank project (Van Vught and Ziegele, 2012). Following on from the U-Map project (Van Vught, 2009), the U-Multirank has set out to compare peer HEIs, using the U-Map classification tool to construct groups of similar HEIs. Within these groups, the project then attempts to apply a multi-dimensional ranking that takes into account teaching and learning, research, knowledge transfer, internationalisation (including mobility), and community outreach. Ultimately, the project aims to deliver a transparency tool that can be used by multiple stakeholders. U-Multirank is, in many ways, a natural companion project for AHELO, aspiring to increase the validity, scope, diversity and transparency of information on higher education.

A second group of initiatives focus on the definition of expected learning outcomes for different qualifications or degrees:

- Qualifications frameworks

  Qualification frameworks are grids developed to enhance the transparency and recognition of qualifications. They show the desired learning outcomes associated with a particular qualification, what an individual holding this qualification should know, understand and be able to do, as well as the interaction between various

© OECD 2012

qualifications within a system. Qualification frameworks have developed rapidly over the past two decades, and more than 70 countries now have developed national qualification frameworks. In Europe, the overarching Framework of Qualifications of the European Higher Education Area (EHEA-QF) was adopted by the ministers of education of the Bologna Process in May 2005, through the Bergen Communiqué. Descriptors for each cycle are based on learning outcomes and competencies, and signatories committed to develop compatible national frameworks by 2010. In addition, the European parliament adopted in 2010 a recommendation establishing a European Qualifications Framework (EQF) which acts as a translation device to make national qualifications more readable across Europe, promote workers' and learners' mobility and facilitate their lifelong learning. Starting in 2012, all qualifications are expected to explicitly refer to an appropriate EQF level. Their main drawback is that they are unfamiliar to non-specialists, and prescriptive rather than empirically descriptive.

- Tuning

  "Tuning" (Tuning Association, 2011) is a process initiated in 2000 in Europe which involves describing and aligning degree outcomes on the basis of competencies and learning outcomes. It was initially developed as a project to link the higher education sector to the Bologna Process and Lisbon Strategy. Working with devolved communities of scholars, Tuning reflects the need for universities to sustain their unique missions within collaboratively determined frames of reference and to retain their diversity while looking for points of reference, convergence and common understanding. Its underlying rationale is to enable the comparability of degrees. Reference points have been developed for first and second cycle programmes, covering generic and subject-specific competencies for a number of subject areas. The work, now expanded into the Americas, Africa, Central Asian Republics and Russia (with feasibility studies in Australia, Canada and China), provides important preconditions for AHELO, as it initiates conversations about learning outcomes, encourages scholars to consider curricula and qualification comparisons, and produces cross-national communities of practice. As a matter of fact, a Tuning-AHELO project was undertaken at the outset of the AHELO feasibility study to reflect on expected learning outcomes for the two disciplines chosen (see Chapters 3 and 4).

Finally, a third group of initiatives go one step further and attempt to measure learning outcomes achieved by students and learners, as opposed to learning outcomes expected or desired.

- National assessments

  Various national approaches to assessing learning outcomes exist (Nusche, 2007), yet few span national boundaries. Hence a particularly important rationale for a study such as AHELO is to produce consistent information on learning outcomes that is international, and potentially global, in scope.

© OECD 2012

- International assessment collaborative initiatives

  International assessment collaborations, such as those of the United Kingdom's Medical Schools Council Assessment Alliance (MSC-AA, 2012) or the Australian Medical Assessment Collaboration (AMAC, 2012) constitute promising initiatives for yielding generalisable information on learning outcomes. They concentrate on the collaborative development of common examinations, question items and formative assessments. Partner schools can then access them through a bank of high-quality, valid and reliable items in a variety of formats. These initiatives go one step further and deliver learning outcomes data which can be generalised beyond local contexts. Work is underway to extend and bolster this model in cross-national contexts.

This snapshot review has outlined key initiatives collecting generalisable data on students' learning outcomes. It has highlighted the complexity of attempts to undertake meaningful international comparative studies which are relevant to a variety of stakeholders and which facilitate transparency in increasingly diverse higher education systems. As illustrated, there are solid foundations, but to date they have not explored system, institutional and discipline-specific performance.

Students' learning outcomes are a key factor of institutional performance, and hence of aggregate system performance. While some indirect evidence can be gained from graduate destination surveys and student engagement surveys, to date, there are no instruments for international measurement. AHELO has the potential to fill this gap by directly assessing student performance.

As such, AHELO can play a unique and increasingly significant role within this global dynamic. But the prerequisite is to ensure that AHELO is feasible. This has been the focus of the feasibility study whose processes and initial findings are described in the following chapters. Interestingly, as the AHELO feasibility study has unfolded over the past few years, some of the shaping rationales which led to AHELO have become even more important. Learning outcomes have gained in importance in higher education discussions and debates. Budgets are tighter than ever, placing growing pressures on the sector and students, and raising interest in higher education productivity. New HEIs have emerged. And the ever more diverse profiles of student bodies require increasingly individualised and student-centred teaching and learning strategies. All of this makes the AHELO feasibility study work particularly timely.

© OECD 2012

REFERENCES

d'Addio, A.C. (2007), "Intergenerational Transmission of Disadvantage: Mobility or Immobility across Generations? A Review of the Evidence for OECD Countries", OECD Social Employment and Migration Working Papers No. 52, Paris, OECD Publishing.

Agarwal, P. (2009), *Indian Higher Education, Envisioning the Future*, Sage Publishers, New Delhi.

AHELO Consortium (2011), AHELO Contextual Dimension Assessment Framework, Melbourne, ACER, CHEPS & Indiana University, *http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=edu/imhe/ahelo/gne(2011)21/ann1/final&doclanguage=en*.

Altbach, P. (2009), "The Giants Awake: the Present and Future of Higher Education Systems in China and India", *Economic and Political Weekly*, Vol. 44(23), Mumbai, pp. 39-51.

Altbach, P. (2004), "Higher Education Crosses Borders", *Change*, Vol. 36(2), Philadelphia, pp. 18-24.

Altbach, P. et al. (2009), *Trends in Global Higher Education: Tracking an Academic Revolution*, Report prepared for the UNESCO 2009 World Conference on Higher Education, UNESCO Publishing, Paris. *http://unesdoc.unesco.org/images/0018/001831/183168e.pdf*.

Altbach, P. and R. Bassett (2004), "The Brain Trade", *Foreign Policy*, September-October, Washington, DC, pp. 30-31.

American Psychological Association Work Group of the Board of Educational Affairs (1997), *Learner-Centered Psychological Principles: Guidelines for School Reform and Redesign*, American Psychological Association, Washington.

Askling, B. and B. Stensaker (2002), Academic Leadership: Prescriptions, Practices and Paradoxes", *Tertiary Education and Management*, Vol. 8(2).

Astin, A. and J. Lee (2003), "How Risky are One-Shot Cross-Sectional Assessments of Undergraduate Students?", *Research in Higher Education*, Vol. 44(6), Jacksonville.

Australian Medical Assessment Collaboration (2012), Australian Medical Assessment Collaboration, *www.acer.edu.au/amac*.

Barr, R. and J. Tagg (1995), "From Teaching to Learning - A New Paradigm for Undergraduate Education", *Change Magazine*, Vol. 27(6), Taylor & Francis Group, Philadelphia, pp. 12-25 *http://ilte.ius.edu/pdf/BarrTagg.pdf*

Bennell, P. and T. Pierce (2003), "The Internationalisation of Tertiary Education: Exporting Education to Developing and Transitional Economies", *International Journal of Educational Development*, Vol. 23, Amsterdam, pp. 215-232.

Birnbaum, R. (1983*), Maintaining Diversity in Higher Education*, Jossey-Bass, San Francisco.

© OECD 2012

Bjarnason, S. et al. (2009), *A New Dynamic: Private Higher Education* Report prepared for the UNESCO 2009 World Conference on Higher Education, UNESCO Publishing, Paris. *http://unesdoc.unesco.org/images/0018/001831/183174e.pdf*.

Blanden, J. et al. (2007), "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education", *The Economic Journal*, Vol. 117, Wiley Online Library.

Böhm, A., et al. (2004), *Forecasting International Student Mobility – a UK Perspective*, British Council, Universities UK and IDP Education Australia, London.

Bologna Secretariat (2012), Making the Most of Our Potential: Consolidating the European Higher Education Area - Bucharest Communiqué of the European Ministers of Education, Bucharest, *http://www.ehea.info/Uploads/%281%29/Bucharest%20Communique%202012.pdf*

Bologna Secretariat (1999), *Joint Declaration of the European Ministers of Education*, 19 June 1999 *http://ec.europa.eu/education/policies/educ/bologna/bologna.pdf*

Borden, V. and T. Banta (1994), *Using Performance Indicators to Guide Strategic Decision Making*, New Directions for Institutional Research No.82, Jossey-Bass, San Francisco.

Bowman, N. (2010), Can 1st Year College Students Accurately report their Learning and Development?", *American Educational Research Journal*, Vol. 47(2), Washington, DC, pp. 466-496.

Brennan, J. (1997), "Authority, Legitimacy and Change: the Rise of Quality Assessment in Higher Education", *Higher Education Management*, Vol. 9(1).

Brock, T. (2010), "Young Adults and Higher Education: Barriers and Breakthroughs to Success", *Future of Children*, Vol. 20(1), Princeton, pp. 109-132.

China Government (2010), *Outline of China's National Plan for Medium and Long-Term Education Reform and Development 2010-2020*, Beijing, Government. *http://planipolis.iiep.unesco.org/format_liste1_fr.php?Chp2=Chine*

Cheng, Y.C. (2009), "Teacher Management and Educational Reforms: Paradigm Shifts", *Quarterly Review of Comparative Education*, Vol. 39(1), Springer, Netherlands/UNESCO, pp. 69-89.

Chung, C. (2011), "Changing Engineering Curriculum in the Globalised World", New Horizons in Education, Vol. 59(3), Hong Kong, pp. 59-70.

Christensen, B. (2010), "Has External Quality Assurance Actually Improved Quality in Higher Education Over the Course of 20 Years of the 'Quality Revolution'?", *Quality in Higher Education*, Vol 16(2) *www.tandfonline.com/doi/pdf/10.1080/13538322.2010.485732*

Council of Europe (2005), *Standards for Recognition: the Lisbon Recognition Convention and its Subsidiary Texts*, Higher Education Series No. 3, Council of Europe Publishing, Strasbourg.

© OECD 2012

Cornelius-White, J. (2007), "Learner-Centered Teacher-Student Relationships are Effective: A Meta-Analysis", Review of Educational Research, Vol. 77(1).

Daniel, J.S. (2009), *Highlights of the UNESCO Global Forum on Rankings and Accountability: Uses and Misuses*, closing presentation, Paris. *www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/ED/pdf/RANKINGS/Stamenka-JohnDaniel.pdf*

Davis, T.M. (2003), *Atlas of Student Mobility*, Institute of International Education, New York.

Dill, D. and F. Van Vught (eds) (2010), National Innovation and the Academic Research Enterprise; Public Policy in Global Perspective, the Johns Hopkins University Press, Baltimore.

Dill, D. and M. Beerkens (2010), Public Policy for Academic Quality: Analyses of Innovative Policy Instruments, Springer, New York.

Dill, D. and M. Soo (2005), "Academic Quality, League Tables, and Public Policy: A Cross-national Analysis of University Rankings", *Higher Education*, Vol. 49, pp. 495-533.

Dobbins, M., C. Knill and E.M. Vögtle (2011), "An Analytical Framework for the Cross-Country Comparison of Higher Education Governance", *Higher Education*, Vol. 62(5) *http://download.springer.com/static/pdf/651/art%253A10.1007%252Fs10734-011-9412-4.pdf?auth66=1354724592_67f6e71ec96a1bbced55eec25da48d42&ext=.pdf*

Douglass, J. (2006), "The Waning of America's Higher Education Advantage: International Competitors Are No Longer Number Two and Have Big Plans in the Global Economy", *Research and Occasional Paper Series*, CSHE.9.06, Center for Studies in Higher Education, Berkeley.

Douglass, J. et al. (2012), "The Learning Outcomes Race: the Value of Self-Reported Gains in Large Research Universities", *Higher Education*, Vol. 64, pp. 317-335.

Douglass, J. and R. Edelstein (2009), "The Global Competition for Talent: The Rapidly Changing Market for International Students and the Need for a Strategic Approach in the U.S.", *Research and Occasional Paper Series*, CSHE.8.09, Center for Studies in Higher Education, Berkeley.

Eggins, H. and P. West (2010), "The Global Impact of the Financial Crisis: Main Trends in Developed and Developing Countries", *Higher Education Management and Policy*, Vol. 22(3), OECD Publishing, Paris.

Elaine Allen, I. and J. Seaman (2011), Going the Distance: Online Education in the United States 2011, , Babson Survey Research Group, Babson Park. *www.babson.edu/Academics/centers/blank-center/global-research/Documents/going-the-distance.pdf*

El Khawas, E. et al. (1998), *Quality Assurance in Higher Education: Recent Progress; Challenges Ahead*, World Bank, Washington, DC. *www.eric.ed.gov/PDFS/ED453720.pdf*

© OECD 2012

Enders, J. and E. de Weert (2004), "Science, Training and Career", *Higher Education Policy*, Vol. 17(2), pp. 129-133.

European Commission (2010), *Europe 2020 - A European Strategy for Smart, Sustainable and Inclusive Growth*, European Commission, Brussels. *http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf*

Ewell, P. (2010), "Twenty Years of Quality Assurance in Higher Education: What's Happened and What's Different?", *Quality in Higher Education*, Vol. 16(2) *www.tandfonline.com/doi/pdf/10.1080/13538322.2010.485728*

Ewell, P. T. (1999), "Linking Performance Measures to Resource Allocation: Exploring Unmapped Terrain", *Quality in Higher Education*, Vol. 5(3), Taylor & Francis Online.

GAMSAT (2012), Graduate Medical School Admissions Test *www.gamsat-ie.org/*

Garcia, T. and H. L'Orange (2010), *Strong Foundations: The State of State Postsecondary Data Systems*, State Higher Education Executive Officers (SHEEO), Boulder.

Giannakou, M. (2006), Chair's Summary, Meeting of OECD Education Ministers: *Higher Education – Quality, Equity and Efficiency*, Athens. *www.oecd.org/site/0,3407,en_21571361_36507471_1_1_1_1_1,00.html*

Gibbons, M. (1998), *Higher Education Relevance in the 21st Century*, UNESCO World Conference on Higher Education, 5-9 October 1998, Paris

Goedegebuure, L. et al. (1994), Higher Education Policy, an International Comparative Perspective, Pergamon, London.

Griffith, A. and K. Rask (2007), "The Influence of the US News and World Report Collegiate Rankings on the Matriculation Decision of High-ability Students: 1995-2004", *Economics of Education Review*, Vol. 26, Elsevier, pp. 244-255.

Grubb, N. (2003), "The Roles of Tertiary Colleges and Institutes: Trade-offs in Restructuring Postsecondary Education", Commissioned paper prepared during a sabbatical period at the OECD Directorate for Education, Paris.*www.oecd.org/dataoecd/50/55/35971977.pdf*

Guri-Rosenblit, S. et al. (2007), "Massification and Diversity of Higher Education Systems: Interplay of Complex Dimensions", *Higher Education Policy*, Vol. 20, Palgrave Macmillan, Basingstoke, pp. 373-389.

Guri-Rosenblit, S. and H. Sebkova (2004), "Diversification of Higher Education Systems: Patterns, Trends and Impacts", UNESCO Forum Occasional Forum Series No. 6 - Diversification of Higher Education and the Changing Role of Knowledge and Research, Paris, pp. 40-69.

Gürüz, K. (2008), Higher Education and International Student Mobility in the Global Knowledge Economy, SUNY Press, Albany.

© OECD 2012

Harman, G. (2006), "Australia as a Higher Education Exporter", *International Higher Education*, Vol. 42, CIHE, Chestnut Hill.

Harvey, L. (2006), *Impact of Quality Assurance – Overview of Discussions*, presentation at the INQAAHE workshop on Trans-national cooperation between agencies and institutions, The Hague, 17-19 May 2006. *www.nvao-event.net/inqaahe/proceedings.php*

Hazelkorn, E. (2008), "Learning to Live with League Tables and Ranking: The Experience of Institutional Leaders", *Higher Education Policy*, Vol. 21, Palgrave Macmillan, Basingstoke, pp. 193-215.

Hazelkorn, E. (2007), "Impact and Influence of League Tables and Ranking Systems on Higher Education Decision-Making", *Higher Education Management and Policy*, Vol. 19(2), pp. 87-110.

Huisman, J. (1993), Differentiation, Diversity and Dependency in Higher Education, Lemma, Utrecht.

Hursh, D. and A. Wall (2011), "Repoliticizing Higher Education Assessment within Neoliberal Globalization", *Policy Futures in Education*, Vol. 9(5), Symposium Journals, Oxford, pp. 560-572.

Ideal Consortium. (2012). IDEAL Consortium: Sharing Medical Student Assessment Banks *http://temporary.idealmed.org/wp/*

International Association of Universities (2013), *International Handbook of Universities 2013*, Palgrave Macmillan, London.

Jacobs, B. and F. Van der Ploeg (2005), "Guide to Reform of Higher Education: A European Perspective", *CEPR Discussion Paper No. 5327*, Centre for Economic Policy Research, London.

Johnson, L. et al. (2012), *The NMC Horizon Report: 2012 Higher Education Edition*, The New Media Consortium, Austin. *http://www.nmc.org/pdf/2012-horizon-report-HE.pdf*

Johnstone, S., P. Ewell, K. Paulson (2002), Student Learning as Academic Currency, American Council on Education *http://books.google.com/books/about/Student_learning_as_academic_currency.html?id=5c9KAAAAYAAJ*

Kaiser, F. et al. (1999), "Market-type Mechanisms in Higher Education: A Comparative Analysis of Their Occurrence and Discussions on the Issue in Five Higher Education Systems", *Higher Education Monitor Thematic Report*, Center for Higher Education Policy Studies (CHEPS), Twente.

Kanter, M. J. (2011), "American Higher Education: First in the World", *Change Magazine*, Vol. 43(3), Taylor & Francis Group, Philadelphia, pp. 7-19.

© OECD 2012

Kapur, D. and M. Crowley (2008), "Beyond the ABCs: Higher Education and Developing Countries", *Working Paper No. 139*, Center for Global Development, Washington D.C. *www.cgdev.org/content/publications/detail/15310/*

Kennedy, D. (2008), "Linking Learning Outcomes and Assessment of Learning of Student Science Teachers", *Science Education International*, Vol. 19(4), pp. 387-397.

Kemp, S. et al. (1998), "Emerging Australian Education Markets: A Discrete Choice Model of Taiwanese and Indonesian Student Intended Study Destination", *Education Economics*, Vol. 6(2), Taylor & Francis Online, pp. 159-169.

Knight, J. (2003), "Updated Internationalisation Definition", *International Higher Education*, Vol. 33, CIHE, Chestnut Hill, pp. 2-3.

Koucký, J. et al. (2008): Inequality and Access to European Tertiary Education During Last Fifty Years. Education Policy Centre, Charles University, Prague.

Kuh, G. (2009), "The National Survey of Student Engagement: Conceptual and Empirical Foundations", *New Directions for Institutional Research*, No. 41, Wiley InterScience, Malden *http://onlinelibrary.wiley.com/doi/10.1002/ir.283/pdf*

Kwieck, M. (2001), "Social and Cultural Dimensions of the Transformation of Higher Education in Central and Eastern Europe", *Higher Education in Europe*, Vol. 26(3), Taylor & Francis Online, pp. 399-410.

Kyvik, S. (2004), "Structural Changes in Higher Education Systems in Western Europe", *Higher Education in Europe,* Vol. 29, Taylor & Francis Online, pp. 393-409.

Lebeau, Y. et al. (2012), "Who Shall Pay for the Public Good? Comparative Trends in the Funding Crisis of Public Higher Education", *Compare: a Journal of Comparative and International Education*, Vol. 42(1), Taylor & Francis Online, pp. 137-157.

Levy, D. C. (2008), "Access through Private Higher education: Global Patterns and Indian Illustrations", *PROPHE Working Paper No. 11*, Program for Research on Private Higher Education, Albany. *www.albany.edu/dept/eaps/prophe/publication/paper/PROPHEWP11_files/PROPHEWP11_Levy.pdf*

Liu, O.L. (2011a), "Examining American Post-Secondary Education", *Research Report ETS RR-11-22*, Educational Testing Service, Princeton.

Liu, O.L. (2011b), "Outcomes Assessment in Higher Education: Challenges and Future Research in the Context of Voluntary System of Accountability", *Educational Measurement: Issues and Practice*, Vol. 30(3), Wiley-Blackwell, pp. 2-9.

Loris, M. (2010), "The Human Journey: Embracing the Essential Learning Outcomes", *Liberal Education*, Vol. 96(1), AAC&U, Washington, DC, pp. 44-49.

© OECD 2012

Lumina Foundation (2011), *The Degree Qualifications Profile*, The Lumina Foundation, Indianapolis.

Marginson, S. (2009), "The Knowledge Economy and Higher Education: Rankings and Classifications, Research Metrics and Learning Outcomes Measures as a System for Regulating the Value of Knowledge", *Higher Education Management and Policy*, Vol. 21(1), OECD Publishing, Paris, pp. 31-46.

Marginson, S. and M. Van der Wende (2007), "Globalisation and Higher Education", *OECD Education Working Paper No. 8*, OECD Publishing, Paris.

Matějů, P. et al. (2003), "Transition to University under Communism and after Its Demise: The Role of Socio-economic Background in the Transition between Secondary and Tertiary Education in the Czech Republic, 1948-1998", *Czech Sociological Review*, Vol. 39 (3), Prague, pp. 301-324.

Medical Schools Council Assessment Alliance (2012), *Medical Schools Council Assessment Alliance*, *http://www.medschools.ac.uk/MSC-AA/Pages/default.aspx*

McBurnie, G. and C. Ziguras (2001), "The Regulation of Transnational Higher Education in Southeast Asia: Case Studies of Hong Kong, Malaysia and Australia", *Higher Education*, Vol. 2, Springer, the Netherlands, pp. 85-105.

McCombs, B. and L. Miller (2007), Learner-Centered Classroom Practices and Assessments: Maximizing Student Motivation, Learning, and Achievement, Corwin Press, Thousand Oaks.

McCormick, A. (2009), Toward Reflective Accountability: Using NSSE for Accountability and Transparency, New Directions for Institutional Research, No. 41, Wiley InterScience, Malden
http://onlinelibrary.wiley.com/doi/10.1002/ir.289/pdf

Mongkhonvanit, P. and S. Emery (2003), "Asian Perspectives on European Higher Education", *Higher Education in Europe*, Vol. 28(1), pp. 51-56.

Mora, J. (2004), "Academic Staff in Spanish Universities: Country Report Spain", in J. Enders and E. de Weert (eds.), *The International Attractiveness of the Academic Workplace in Europe*, Herausgeber und Bestelladresse, Frankfurt, pp. 395-414.

MSC-AA (2012), Medical Schools Council Assessment Alliance
http://www.medschools.ac.uk/MSC-AA/Pages/default.aspx

Musselin, C. (2004), "Towards a European Academic Labour Market? Some Lessons Drawn from Empirical Studies on Academic Mobility", *Higher Education*, Vol. 48, Springer, New York, pp. 55-78.

National Research Council (2012), *Improving Measurement of Productivity in Higher Education*, The National Academies Press, Washington DC.
www.nap.edu/catalog.php?record_id=13417

© OECD 2012

Nelson, B. (2003), Engaging the World through Education: Ministerial Statement on the Internationalisation of Australian Education and Training, Australia Department for Education, Science and Training, Canberra.

Nelson, P. (1970), "Information and Consumer Behavior", *Journal of Political Economy*, Vol. 78(2).

Newton, J. (2001), *Views From Below: Academics Coping with Quality*, Keynote presentation at the 6th QHE Seminar: The End of Quality? Birmingham, 25-26 May 2001.

Newton, J. (2000), "Feeding the Beast or Improving Quality?: Academics' Perceptions of Quality Assurance and Quality Monitoring", *Quality in Higher Education*, Vol. 6(2), Taylor & Francis Online.

Norway Opening Universities (2011), ICT Monitor
http://norgesuniversitetet.no/files/ict_status_2011_0.pdf

Nusche, D. (2007), "Approaches to Learning Outcomes Assessment in Higher Education", *OECD Education Working Paper No. 15*, OECD Publishing, Paris.

Obama, B. (2009), *State of the Union Address to Joint Session of Congress*, February 24, 2009.
www.whitehouse.gov/the_press_office/Remarks-of-President-Barack-Obama-Address-to-Joint-Session-of-Congress/

OECD (2012a), "How is the Global talent Pool Changing?", *Education Indicators in Focus*, Issue 2012/05, May 2012, OECD Publishing, Paris.

OECD (2012b), *Education at a Glance 2012: OECD Indicators*, OECD Publishing, Paris.

OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/eag-2011-en.

OECD (2010), *Education at a Glance 2010: OECD Indicators*, OECD Publishing, Paris.
www.oecd.org/edu/eag2010

OECD (2008), *Tertiary Education for the Knowledge Society*, OECD Publishing, Paris.
www.oecd.org/edu/tertiary/review

OECD (2005a), "Alternatives to Universities Revisited", Chapter 1 in *Education Policy Analysis 2004*, OECD Publishing, Paris.

OECD (2005b), *E-learning in Tertiary Education: Where do We Stand?*, Centre for Educational Research and Innovation (CERI), OECD Publishing, Paris.

OECD (2007), *Giving Knowledge for Free: the Emergence of Open Educational Resources*, OECD Publishing, Paris.
www.oecd.org/edu/ceri/38654317.pdf

OECD (2004), *Internationalisation and Trade in Higher Education: Opportunities and Challenges*, OECD Publishing, Paris.

© OECD 2012

Parker, L. and G. Gould (1999), "Changing Public Sector Accountability: Critiquing New Directions", *Accounting Forum*, Vol. 23(2), Elsevier.

Peace Lenn, M. and L. Campos (1997), Globalisation of the Professions and the Quality Imperative. Professional Accreditation, Certification and Licensure, Magna Publications, Madison.

Pimpa, N. (2005), "Marketing Australian Universities to Thai Students", *Journal of Studies in International Education*, Vol. 9(2), Sage Journals, pp. 137-146.

Prescott, B. and P. Ewell (2009), *A Framework for a Multi-state Human Capital Development Data System*, Western Interstate Commission for Higher Education *www.wiche.edu/info/publications/FrameworkForAMultistateHumanCapitalDevelopmentDataSystem.pdf*

Pusser, B. (2003), "Beyond Baldridge: Extending the Political Model of Higher Education Organization and Governance", *Educational Policy*, Vol. 17(1) *http://epx.sagepub.com/content/17/1/121.full.pdf*

Radloff, A. (2011), Student Engagement in New Zealand's Universities, Australasian Survey of Student Engagement *http://www.acer.edu.au/documents/aussereports/AUSSE_New_Zealand_Report_Web_Version.pdf*

Radloff, A. and H. Coates (2009), Doing More for Learning: Enhancing Engagement and Outcomes, Australasian Student Engagement Report *http://www.acer.edu.au/documents/aussereports/AUSSE_Australasian-Student-Engagement-Report-ASER-2009.pdf*Rauhvargers, A. (2011), Global University Rankings and Their Impact, EUA Report on rankings 2011, *www.eua.be/pubs/global_university_rankings_and_their_impact.pdf*

Reichert, S (2010), "The Intended and Unintended Effects of the Bologna Reforms", *Higher Education Management and Policy*, Vol. 22(1), pp. 59-78, OECD Publishing, Paris.

Rinne, R. (2008), "The Growing Supranational Impacts of the OECD and the EU on National Educational Policies, and the Case of Finland", *Policy Futures in Education*, Vol. 6(6), pp. 665-680.

Roth, F. and A.-E. Thum (2010), "The Key Role of Education in the Europe 2020 Strategy", *Working Document No. 338,* October 2010, Centre for European Policy Studies, Brussels. *www.ceps.eu/book/key-role-education-europe-2020-strategy*

Salmi, J. (2000), "Tertiary Education in the Twenty-First Century: Challenges and Opportunities", *LCSHD Paper Series*, The World Bank, Washington, DC.

Salmi, J. and A. Saroyan (2007), "League Tables as Policy Instruments", *International Higher Education*, Vol. 47, CIHE, Chestnut Hill, pp. 2-4.

© OECD 2012

Sanyal, B. and B. Johnstone (2012), International Trends in the Public and Private Financing of Higher Education", *Quarterly Review of Comparative Education*, Vol. 41(1), Springer, Netherlands/UNESCO, pp. 157-175.

Schofer, E. and J.W. Meyer (2005), "The Worldwide Expansion of Higher Education in the Twentieth Century", *American Sociological Review,* Vol. 70, American Sociological Association, Washington, DC.

Schuetze, H.G. and M. Slowey (2002), "Participation and Exclusion: A Comparative Analysis of Non-Traditional Students and Lifelong Learners in Higher Education", *Higher Education,* Vol. 44, Springer, the Netherlands.

Scott, P. (2000), "Globalisation in Higher Education: Challenges for the 21st Century", *Journal of Studies in International Education*, Vol. 4(1), pp. 3-10.

Scott, W. et al. (2009), "Learning Centered Universities: the Changing Face of Higher Education", Journal of Faculty Development, Vol. 23(1), New Forums Press, Stillwater, pp. 14-23.

Secolsky, C. and B. Denison Eds (2011), Handbook on Measurement, Assessment, and Evaluation in Higher Education, Taylor & Francis Group.

Segrera, F.L. (2010), "Trends and Innovations in Higher Education Reform: Worldwide, Latin America and the Caribbean", *Research and Occasional Paper Series No.12.10*, Center for Studies in Higher Education, Berkeley.

Shanghai Jiao Tong University (2011), *Academic Ranking of World Universities 2011*, Institute of Higher Education, Shanghai Jiao Tong University. *www.shanghairanking.com/ARWU2011.html*

Shanghai Jiao Tong University (2003), *Academic Ranking of World Universities 2003*, Institute of Higher Education, Shanghai Jiao Tong University. *www.arwu.org/ARWU2003.jsp*

Silva, M., et al. (1997), "Effects of External Quality Evaluation in Chile: A Preliminary Study", *Quality in Higher Education*, Vol. 3(1), Taylor & Francis Online.

Siow, A. (1997), "Some Evidence on the Signalling Role of Research in Academia", *Economics Letters*, Vol. 54(3), Elsevier, pp. 271-276.

Song, M.M. and H.H. Tai (2007), "Taiwan's responses to Globalisation: Internationalisation and Questing for World Class Universities", *Asia Pacific Journal of Education*, Vol. 27(3), Taylor & Francis Online, pp. 323-340.

Swail, W.S. (2011), "In Search of a Better Mousetrap: A Look at Higher Education Ranking Systems", *College and University*, Vol. 86(4), AACRAO, Washington, DC., pp. 28-33.

Tapper, T. and D. Palfreyman (2005), Understanding Mass Higher Education: Comparative Perspectives on Access, Routledge Falmer, London.

© OECD 2012

Tremblay, K. (2002), "Student Mobility Between and Towards OECD Countries: A Comparative Analysis", in *Trends in International Migration – SOPEMI 2001*, OECD Publishing, Paris, pp. 93-117.

Trow, M. (1974), "Problems in the Transition from Elite to Mass Higher Education", in *Policies for Higher Education*, OECD Publishing, Paris.

Tuning Association (2011), Tuning Educational Structures in Europe, *www.unideusto.org/tuningeu/home.html*.

TSL Education (2011). *World University Rankings 2011-2012* *www.timeshighereducation.co.uk/world-university-rankings/*.

UNDP (2012), World Population Prospects: The 2010 Revision *http://esa.un.org/unpd/wpp/index.htm*.

UNESCO Institute for Statistics (2011), *Global Education Digest 2011*, UNESCO Publishing, Montreal. *http://stats.uis.unesco.org*

Usher, A. And M. Savino (2006), *A World of Difference: A Global Survey of University League Tables*, Education Policy Institute: Canadian Education Report Series *www.educationalpolicy.org/pdf/world-of-difference-200602162.pdf*

USMLE (2012), United States Medical Licensing Examination *http://www.usmle.org/*

US News and World Report (2011), *Best Colleges* *www.usnews.com/education*.

Vught, F. Van (2008), Mission Diversity and Reputation in Higher Education, *Higher Education Policy*, Vol. 21(2).

Vught, F. Van, Ed. (2009), Mapping the Higher Education Landscape, towards a European Classification of Higher Education, Springer, Dordrecht.

Vught, F. Van and D. Westerheijden (1994), "Towards a General Model of Quality Assessment in Higher Education", *Higher Education*, Vol. 28.

Vught, F. Van and F. Ziegele (2012), Multidimensional Ranking: The Design and Development of U-Multirank, Springer, Dordrecht.

Webber, K. (2012), "The Use of Learner-Centered Assessment in US Colleges and Universities", *Research in Higher Education*, Vol. 53(2), Springer, pp. 201-228.

Weimer, M. (2002), Learner-Centered Teaching: Five Key Changes to Practice, Jossey-Bass, San Francisco.

Wende, M. Van der (2001), "Internationalisation Policies: About New Trends and Contrasting Paradigms", *Higher Education Policy*, Vol. 14, Palgrave Macmillan, Basingstoke, pp. 249-259.

Wilkins, S. and J. Huisman (2012), "UK Business Schools Rankings over the Last 30 years (1980-2010): Trends and Explanations", *Higher Education*, Vol. 63(3), pp. 367-382.

© OECD 2012

NOTES

1 Pusser (2003) defines the collegial model of university governance as follows: "Within the collegial dimension, organizations are viewed as entities based on collective action and widely shared values. Decision making is seen as a participatory, pluralist, and democratic process within a collegium or community of scholars. Here goals are shaped by collective deliberation, action is guided by consensus, and power and decision making are shared throughout the organization".

2 Entry rates to tertiary type 5A programmes as per the international standard classification of education (ISCED) taxonomy. Entry rates represent the proportion of people in a single age-cohort who enter a given level of higher education at some point in their lives. The OECD average is processed between OECD countries with trend data only.

3 Entry rates to tertiary type 5B programmes. Average processed between OECD countries with trend data.

4 EU Member states who had already surpassed the EU 2020 goal by 2009: Belgium, Cyprus *(see notes 4a and 4b below)*, Denmark, Finland, France, Ireland, Italy, Luxembourg, the Netherlands, Sweden and the United Kingdom.

4a. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southernpart of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

4b. Note by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

5 Bachelor-Master-Doctorate. In fact the Bologna declaration only envisaged a 2-cycle structure but a third cycle was added in 2003 in Berlin (OECD, 2008).

6 UMAP is an association of government and non-government representatives of the higher education sector in the Asia-Pacific region.

7 Cost per student is calculated as the ratio of public and private expenditure by HEIs to the number of full-time equivalent students enrolled at that level. Expenditure in national currencies is converted using purchasing power parities (PPP) to enable international comparisons. See OECD (2012b) for further details on methodology.

© OECD 2012

8 These include the federal government, accrediting and higher education organisations, and the public.

© OECD 2012

CHAPTER 2

THE BEGINNING OF AHELO: DECISIONS AND CHALLENGES

This chapter provides an overview of the questioning that surrounded the concept of an international Assessment of Higher Education Learning Outcomes (AHELO) when the idea was first discussed in a meeting of higher education ministers in June 2006 in Athens. The chapter then describes the approach chosen by the OECD to carry out the AHELO feasibility study.

© OECD 2012

The notion of exploring the development of an international Assessment of Higher Education Learning Outcomes (AHELO) emerged within the context of the Meeting of OECD Education Ministers held in Athens in 2006, as part of a broader commitment to raise the quality of higher education in their countries (Giannakou, 2006):

> […] measuring the quality of higher education outcomes is needed both to justify the allocation of public resources and the effectiveness with which they are used by increasingly autonomous institutions, and also to pursue enhancements in the quality and relevance of educational outcomes more broadly and systematically, so that higher education institutions (HEIs) serve economies and local communities effectively.

However this initiative – initially referred to as a "PISA for higher education" – generated much discussion throughout 2006-2008 in higher education circles between policy makers, experts and stakeholders. It was received very differently by different constituencies or their members, with some stakeholders embracing the idea enthusiastically and participating actively in its development, while others being more critical of the approach and vocal in pointing to its potential risks. In many cases, the AHELO idea also raised different viewpoints within countries and stakeholder groups.

The first chapter has described in detail the potential value of assessing learning outcomes in higher education in the 21st century. As skills and human capital are gaining in importance for economic prosperity and social well-being, growing cohorts of students worldwide enrol in higher education programmes with the hopes of preparing adequately to meet the challenges of contemporary knowledge-intensive economies and societies. There is already ample evidence of the existence of economic and social returns from higher education, but also scope to make their higher education experience more meaningful. Focusing on learning outcomes can help HEIs and systems move further in this direction.

Indeed, a major benefit of working towards developing comparative measures of learning outcomes is to keep the spotlight on teaching and learning within HEIs. Measures of learning outcomes also hold important promises for higher education faculties and leaders in providing evidence-based diagnosis tools on the strengths and weaknesses of their courses and programmes to be used as part of their quality improvement efforts. Developing benchmarks on learning outcomes attained by students in different HEIs is also essential to better understand the interplay between teaching and learning, and support HEIs and faculties in moving towards more student-centred pedagogies. As higher education is going through a fundamental shift from an instruction paradigm to a learning paradigm – partly as a result of increasingly diverse missions, roles and student bodies – the value of AHELO types of measures will only grow in importance.

This is not to say that the concept of developing an AHELO is not without questions and challenges. As a matter of fact, in the early years of the AHELO feasibility study there was much discussion between experts and stakeholders on the AHELO initiative and the challenges to develop and operationalise such an assessment. This chapter therefore portrays the key challenges identified by stakeholders in developing and operationalising an AHELO, before

© OECD 2012

outlining the experts' advice on addressing these issues and the implications they had on the design of the AHELO feasibility study.

Questioning feasibility: Key challenges in developing and operationalising an AHELO

The first few years of the AHELO feasibility study were characterized by much discussion within some stakeholder circles on the nature and purposes of AHELO, and the perceived challenges and risks associated with the development of international measures of learning outcomes. Overall, concerns focused on the risk of rankings based on AHELO data, possible misuse by governments and policy makers, the complexity of comparisons across diverse HEIs, the risks inherent in the use of standardised tests, the potential impact of such a tool on institutional autonomy and academic freedom, the focus of AHELO on generic skills as well as some more methodological and practical questions. These issues are outlined in this section for the sake of providing an overall picture of the context in which the AHELO feasibility study unfolded, even though some of the challenges below are sometimes based on misconceptions or opinions. Accordingly, with a few exceptions, the references in this section essentially consist of opinion pieces reported in various media rather than scholarly research articles.

Risk of ranking

In its early months, the AHELO idea was received with caution and scepticism in some higher education circles, not least due to the erroneous yet common comparison of this initiative with rankings of HEIs whose distorting effects were widely commented on and criticized at the time (Hazelkorn, 2007; Van Vught, 2009). For example, *The Economist* reported in 2007 that AHELO would:

> […] sample university students to see what they have learned. Once enough universities are taking part, it may publish league tables showing where each country stands, just as it now does for compulsory education. That may produce a fairer assessment than the two established rankings.

Despite the fact that AHELO was never envisaged as a ranking tool within OECD circles, the risk for AHELO data to yield another ranking – or to be used in existing rankings – has featured prominently among the concerns expressed by stakeholders. In a 2008 letter to OECD senior officials, the president of the American Council on Education (ACE) – speaking on behalf of other U.S. higher education associations[1] – indicated that although OECD's "intention is not to develop another ranking system, it is highly likely that the results will indeed be used to rank HEIs" (Ward, cited in Lederman, 2008a).

As described by Ewell (forthcoming), "the AHELO proposal came at a sensitive time for the United States. The negative atmosphere around standardized testing as an element of accountability generated by the Spellings Commission remained palpable and the Voluntary System of Accountability and No Child Left Behind kept the topic alive and controversial in policy circles." As put by one of ACE's officials, "the notion of measuring students' achievement within the United States has been very controversial [...] the notion of developing a mechanism

© OECD 2012

to do it across the world seems orders of magnitude more controversial" (Hartle, cited in Labi, 2007).

But while the U.S. context provides a specific spin on initial reactions in North America, similar concerns were also prevalent elsewhere. Education International (EI) – a global federation of teacher unions – issued a statement in 2007, outlining its methodological and political reservations on AHELO's development. Among them, the organisation voiced concerns from academic communities "that a PISA for higher education could easily be transformed into a simplistic ranking or league table of HEIs" (Education International, 2007). Likewise, a January 2008 meeting of OECD senior officials with the Council of the European University Association (EUA) also revealed worries by some members that AHELO might produce a one-dimensional ranking of the world's universities that would promote a competitive "World Championship" behaviour of HEIs over more diverse measures in an "Olympic Games" spirit, as one speaker put it (EUA, 2008).

According to Achard (2010), the resistance against rankings in general – and any tool which is believed to be possibly used in rankings – derives from the fact that rankings are viewed as "the dragon of New Public Management[2] and accountability assaulting the ivory tower of disinterested knowledge". In his view, rankings "certainly participate in a global shift in the contract between society and universities".

Wildavsky provides however a more strategic interpretation to this defiance:

> So far, United States colleges have little to fear from the currently available international rankings, which focus heavily on the research and reputation measures at which the long-established and top tier of American schools excel. But new rankings that shine a spotlight on student learning as well as research could deliver far less pleasant results, both for American universities and for others around the world that have never put much focus on classroom learning. The truth is that we don't know how we'll stack up – and not everybody wants to find out. Some in the American higher education community have been deeply resistant to the prospect of AHELO (Wildavsky, 2009).

Misuse of results

Some in the higher education sector also pointed out that AHELO results could be misused by governments. According to Olds and Robertson (2011), "there is a political economy to world university rankings, and these schemes [...] are laden with power and generative of substantial impacts; impacts that the rankers themselves often do not hear about, nor feel (e.g. via the reallocation of resources)".

In fact, the emergence of the AHELO initiative within an intergovernmental organisation and the interest it has generated from a number of governments have heightened the sector's anxiety as to the stated (or possibly unstated) goals of AHELO. Green (2011) reports:

> that governments have provided financial support for the first phase and secured institutional participation indicates that policymakers see such a tool as useful. Their very interest, however, creates anxiety among some institutional leaders and

© OECD 2012

> representatives, who fear government-led efforts […] Concerns center on […] the potential for policymakers to use the results as a tool for decision-making rather than for analysis and improvement.

The bulk of concerns related to the possible misuses of results have focused on possible reallocations of resources based on AHELO findings. Notwithstanding the OECD policy recommendation on higher education quality assurance to disconnect assessment results and public funding decisions (OECD, 2008), ACE's president David Ward feared in 2008 that "policymakers will undoubtedly be inclined to use the results in ways that will disadvantage those HEIs that do not perform well against others, possibly for reasons over which they have no control" (cited in Lederman, 2008b). Similar uneasiness was voiced by EI which, in a letter to OECD officials, warned against the "danger that AHELO could be used as a crude accountability mechanism by governments to reward good performers and punish the poor ones". Some apprehensions centred on the possibility that low performing HEIs could suffer funding cutbacks – thereby accentuating the problems. But concerns also pointed to the risk of reallocations of funding within HEIs towards teaching to the detriment of their other missions (e.g. equity, regional role).

Another set of worries address the risk that AHELO data could distort public policy and possibly run counter to AHELO's stated goal to enhance higher education quality. In this spirit, Hazelkorn (2011) argues that "many countries [...] have used rankings to spearhead profound restructuring of their higher-education systems [...] are choosing to reward the achievements of elites and flagship HEIs rather than improve the capacity and quality of the whole system". In her view, AHELO results could similarly be used to strengthen high-performing HEIs at the expense of enhancing quality across the board.

The OECD position on this issue has been clearly outlined in the quality assurance chapter of the Thematic Review of Tertiary Education Policies, which recommended disconnecting assessment results and public funding decisions to avoid such perverse effects (OECD, 2008).

Complexity of comparisons across diverse institutions

There are also a range of arguments against the AHELO concept due to the diversity of HEIs' missions, profiles and student bodies, and the resulting complexity to engage in fair comparisons. The various implications of institutional diversity for developing a comparative assessment tool can be summarised as follows:

- With respect to the diversity of institutions' / programmes' profiles and the determination of learning outcomes to be assessed, EI (2007) notes that "there is a notable lack of consensus on what should be the appropriate practices and outcomes of higher education [...] There is considerable variation between and within HEIs with respect to what specialized programs students pursue and even what is taught within specific subject areas". In the context of greater specialisation of degree programmes at the higher education level, "a standardized international test of higher education outcomes makes little sense".

© OECD 2012

- In relation to variations in student intake and characteristics, critics have also challenged the impossible interpretation of differences in performance between HEIs given that: "attendance at a college or university is not universal but is based on selective standards that vary markedly between countries and even between HEIs within countries […] Test results would therefore not necessarily reflect the 'quality' of education students receive at community colleges or elite HEIs, but would rather be more indicative of differences in admission standards and the socio-economic status of students" (EI, 2007).
- In a letter to OECD senior officials, ACE stressed the need to factor in the profile of students (in terms of goals, admissions criteria, age, academic preparation, socio-economic background and financial means) to ensure meaningful comparisons of institution performance.
- A similar issue relates to the diversity of institutions' missions and that an AHELO – by placing greater emphasis on learning outcomes' performance and results – might result in shifting priorities and goals, and ultimately run counter to the equity mission of some HEIs to train disadvantaged groups of students. A more philosophical concern from some in the higher education community is the fear that institutions would increasingly be seen as producers of manpower skills, thereby fuelling pressures for uniformity to the detriment of other HEI missions.
- Finally, stakeholders have also warned of the risks inherent in comparing HEIs or programmes with varying levels of resources, whether in terms of funding, faculty numbers and qualifications, and other resources.

While dealing with these various dimensions of institutional diversity is far from trivial in developing an international comparative assessment, the collection of detailed information on these important contextual factors as part of an AHELO would enable the identification of peer institutions with similar characteristics in terms of mission, resources, faculty and student characteristics, and emphasis of programmes. An important step in this respect has been the development and feasibility test of the "U-Map" tool, which provides information about the activity profiles of HEIs with the explicit intention to create the conditions to compare HEIs with comparable profiles (Van Vught, 2009).

Contesting of standardised tests

Directly related to institutional diversity, AHELO detractors have also contested the use of standardised tests and a "one-size-fits-all" approach to assessing higher education learning outcomes. As put by one of ACE's officials in 2007, "The conversations in the last year have underscored […] the difficulty of finding a common instrument for measuring [learning outcomes] in the United States […] If we haven't been able to figure out how to do this in the United States, it's impossible for me to imagine a method or standard that would work equally well for Holyoke Community College, MIT and the Sorbonne" (Hartle, cited in Lederman, 2007).

A first set of arguments against standardized tests focus on the inevitably limited information that an instrument, designed to cover the broad spectrum of higher education programmes,

© OECD 2012

can yield for individual HEIs and faculties. Banta and Pike (2012) argue that "the college outcomes assessed by standardised tests represent a very small slice of what is important in education and almost certainly in graduates' post-college lives."

But more than standardised tests, it is often their indiscriminate use which is considered to be problematic as Rhodes (2012) explains: "When decisions are made based on standardised test results of a very few learning outcomes, as we have done in the schools in this country, virtually every other critical learning outcome disappears from practice. [...] Our graduates need more than the limited range of competencies easily measured by standardised tests." In the same spirit, EI argues that "standardised tests almost invariably result in oversimplified measurements of the 'quality' of education" and there is a danger in them leading to "very simplistic conclusions about the quality of complex and highly differentiated systems, processes and HEIs." The organization of faculty unions concludes that "standardized testing can and should never replace the central role that higher education teaching personnel play in assessing students" (EI, 2007).

Others point to the amount of information lost in using standardised tests and claim that other data sources may actually yield more informative data. For instance, Douglas et al. (2012) contend that student self-assessments can provide "an opportunity to learn more than what is offered in standardized tests […] especially in large, complex institutional settings." They argue that properly-designed student surveys "offer a valuable and more nuanced alternative in understanding and identifying learning outcomes in the university environment."

This is not to say that critics of standardised tests see no value in them. Banta (2007) recognises that they "can be helpful in initiating faculty conversations about assessment." But in EI's view, for standardized tests to be useful, they will need to provide useful information on the educational context, resources, programmes and pedagogical processes.

The criticisms forwarded above are somewhat exaggerated, to the extent that as AHELO has never been intended to replace other forms of classroom and institutional assessments or student surveys, nor has it meant to provide the sole yardstick by which to judge higher education quality. Nevertheless, it is important to take notice and have a good understanding of the underlying fears behind some common arguments against AHELO.

Potential impact for institutional autonomy and academic freedom

The questions surrounding AHELO have also pointed to its potential impact on institutional autonomy and academic freedom with fears within higher education circles that participation in the AHELO assessment might be forced on HEIs as an accountability requirement, and could over time contribute to homogenising higher education programmes and constraining academic freedom. Representative organisations of HEIs and faculty unions have been most vocal about these implications.

In a letter to OECD senior officials (cited in Lederman, 2012), representatives from the three institution member organisations involved in the Transatlantic Dialogue[3] argued strongly that "it is possible to measure outcomes and enhance quality without promoting standardisation of

© OECD 2012

educational programmes and homogenisation of institutional missions." EI's position is also illustrative of similar concerns:

> External standardised assessments raise important issues around professional autonomy for academic staff. […] Traditionally, the quality of HEIs has been assessed through rigorous and regular peer reviews[4]. […] The quality of the educational experience students receive simply cannot be quantified in a performance-based test. Quality is a function of the "lived experience" of higher education including the conditions and activities of teaching and free enquiry. (EI, 2007)

The origins of these concerns are somewhat unclear since AHELO has always been envisaged as a tool which would complement – but not replace – existing classroom and institutional assessments and which would be designed precisely to account for and protect the diversity of HEIs' missions, profiles and programmes so as to avoid standardisation and homogenization.

Relevance of generic skills in different systems

Another criticism of the AHELO concept has derived from its focus on generic skills. As Peter Ewell has aptly articulated,

> to academics in the U.S., the existence and importance of "general education" abilities like oral/written communication, critical thinking, and quantitative reasoning go without saying. […] Undergraduate programs elsewhere for the most part have no general education component and students enter directly into study in depth. So it is not surprising that including a Generic Skills assessment in AHELO was questioned in some offshore academic communities. (Ewell, forthcoming)

This issue of different academic traditions in different parts of the world was exacerbated by the choice of the U.S. Collegiate Learning Assessment (CLA) as the basis for the generic skills instrument for the AHELO feasibility study. While the AHELO feasibility study only tests one component of the CLA, and moreover an adapted version of it, it has often been mistakenly interpreted as if the AHELO feasibility study was testing the CLA for broader adoption in an AHELO main study context should the feasibility study be successful. In fact, all the instruments used for the AHELO feasibility study (whether for generic skills of the disciplines) were selected for the purpose of exploring the possibility of validly and reliably measuring learning outcomes and proving the concept. They do not necessarily reflect what a future AHELO might look like.

The use of a CLA-based instrument for generic skills in the AHELO feasibility study has generated three distinct ranges of objections against the "generic skills" component of AHELO:

- Within U.S. circles, the selection of the CLA has somewhat transposed on AHELO fierce methodological discussions on the merits of this specific instrument relative to other approaches, as illustrated by Banta (2007) or Douglass et al. (2012).
- Outside of the United States by contrast, concerns have focused essentially on whether a future AHELO might impose a U.S.-based model on the rest of the world, as reported by Green (2011): "other objections to AHELO derive from criticisms of the CLA, the US instrument that is being adapted for the generic strand. That it is a US

© OECD 2012

test alarms some, even though testing experts and faculty will adapt it to different countries and languages".

- And finally at a conceptual level, the choice of the CLA has raised much discussion on the relevance of assessing generic skills independently and decontextualized from the subject matters. Indeed, while there is wide acceptance in the academic global community that the development of generic skills and competencies should be an integral part of academic studies, there is however much debate as to whether generic skills can and should be assessed out of disciplinary context – as done in the CLA approach. Herl et al. (1999) argue that "in order to be an effective critical thinker/problem solver, one must have knowledge about the issue at hand" whereas Banta and Pike (2012) contend that "skills like written communication, problem solving, and analytic reasoning are learned – and assessed – best as they are applied in a discipline." In addition, Baird (1988) notes that "if we look for evidence about critical thinking and problem solving within discipline or programme areas, the results will be much more acceptable and meaningful to faculty."

Notwithstanding the fact that all instruments chosen for the feasibility study were only intended to prove a concept and do not pre-judge what an eventual future AHELO might look like, they have certainly fuelled anxieties in some academic circles.

Methodological and practical questions

Finally, the concept of an AHELO has raised some additional methodological or practical questions which can be summarised as follows:

- In a review of existing ranking and benchmarking tools, the EUA noted in relation to AHELO that "using learning outcomes for comparing the performance of teaching and learning processes can only be successful if the participating HEIs and countries have actually adopted a learning outcomes-based approach in both the teaching process and in student assessment" which – he argues – is not yet the case as "recent progress reports of the Bologna Process demonstrate that it will take time before this practice becomes general" (Rauhvargers, 2011).

  While the insufficient development of a learning outcomes approach in some national and institutional settings may explain part of the resistance against AHELO types of initiatives, it can be argued that this should not uphold the development of learning outcomes measures which, on the contrary, could contribute to shifting emphasis from staff-centred to student-centred learning in those countries and HEIs where the learning outcomes approach has not yet been adopted.

- Some concerns also focused on how an AHELO could address the issue of student motivation to ensure that student response rates would be sufficient to ensure that the results are representative of an institution's performance. As described by Ewell (forthcoming), "a major challenge that AHELO faces is getting students to exert sufficient effort in completing (or even showing up for) an examination that does not count", and controlling for possible differences between countries in terms of level of

© OECD 2012

student effort and engagement. While some HEIs have become accustomed to addressing this issue through various incentives for student respondents, some stakeholders have questioned the potential impact of student self-selection and the potential of compromising results. Banta and Pike (2012) illustrate those concerns: "Even if samples of test takers could be standardized across colleges and universities, issues of motivation will remain. Motivation to do well on the test also varies enormously, from those who just show up because they are asked to do so to those who are given a vested interest such as earning extra credit in a course or ensuring that their college looks as good as possible in a comparison."

This issue is addressed at some depth in the AHELO feasibility study (see Chapter 5).

As outlined in the paragraphs above, initial discussions around the idea of an AHELO have underlined a number of key challenges that would need to be addressed in developing a valid and meaningful assessment.

Initial expert meetings to frame a roadmap

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

In forming these groups of 10 to 20 experts (see Annex A), a deliberate decision was made to keep the meetings small enough to remain focused, and to ensure the presence of a range of expertise including policy makers, higher education researchers as well as assessment specialists with hands on experience of assessment techniques and challenges.

This pragmatic arrangement made it possible to address a range of issues and provide guidance for a strategy within a limited timeframe but it excluded some stakeholders from these initial discussions. This led to the perception by some that AHELO was being developed without input from certain segments (or groups). Moreover, the presence of experts from the testing industry – while crucial to address some more technical issues and challenges – may have conveyed the impression that the expert task force was skewed towards advocates of standardised testing (Adelman and Hartle, cited in Lederman, 2007; EI, 2007).

Washington meeting

The first expert meeting in Washington considered the usefulness and desirability of an international AHELO (OECD, 2007a). The experts identified considerable challenges to developing internationally comparative measures of higher education learning outcomes and acknowledged that it was not clear how these could be overcome. However, none considered the goal unattainable and all recognised that reliable information on learning outcomes would only increase in importance, as higher education would continue to diversify, internationalise, and as new virtual ways of delivery and provision would make physical space as a unit of service provision less relevant.

© OECD 2012

Determining relevant units of data collection and analysis

Experts acknowledged that measures of higher education learning outcomes hold out important promises for students, HEIs, and public policy, more generally. However the extent to which the information needs of the various stakeholders could be met would depend on the units from which data are collected and for which outcomes are reported.

The experts suggested that at least initially, it would not be feasible to develop internationally comparative information on higher education learning outcomes at the system-level for several reasons:

- variation in institutional structures challenges the establishment of cross-nationally comparable classes of HEIs;
- governments in many countries have limited options to incentivise HEI participation in assessments, particularly those carried out at an international level;
- cross-country variation in enrolment rates and selectivity; and
- mandated assessment, in the form required to obtain a system-wide representative sample of HEIs, would tend to lead to administrative response, not faculty response and thus was not effective as a tool for improvement.

Subsequent discussions, therefore, centred on establishing measures of learning outcomes at the level of HEIs, departments or faculties. The idea was to combine agreed measures of quality with reliable assessment methods to which HEIs could, with an appropriate set of incentives, voluntarily subscribe. The experts considered two kinds of outcome measures at the level of institutions or sub-institutions:

- Individuals, whether prospective students or employers, would want to know the raw scores enrolled students would attain on an assessment, recognising that such an assessment would not only measure the quality of educational services provided but also other facets, such as the effects of selection and the socio-economic aspects.
- Individuals, HEIs and public policy makers would primarily be interested in the "value added" contributed by the HEIs, i.e. the scores an institution would attain after accounting for the quality of prior schooling or the degree of selectivity of the higher education programmes and institutions.

The experts also considered what would constitute meaningful target populations for an international AHELO and rejected the option of comparable age bands as too complicated. They considered an assessment towards the end of a three or four-year degree as a more practical solution for a feasibility study.

Using the results

There were diverging views on how the knowledge about learning outcomes in HEIs can and should be used. Some saw such information primarily as a tool to reveal best practices, to identify shared problems among HEIs and to encourage collaboration and lateral capacity

© OECD 2012

building among research and teaching personnel. This approach emphasises the relevance of performance information for the institutions themselves and on contextualising performance data with other information on the learning environment in HEIs. Other views extended the purpose of learning outcome measures to support contestability of public services or market-mechanisms in the allocation of resources, e.g. by making comparative results of HEIs publicly available to facilitate choice.

The experts agreed that there was little point in reporting data at the level of individual students. Therefore, the assessment could use matrix sampling techniques to widen the coverage of the assessment without unduly extending the response time demands on individuals.

Defining and operationalising higher education learning outcomes

The experts acknowledged that there was no generally accepted definition of what higher education outcomes ought to be, but considered that there were promising ways underway to examine various facets of learning outcomes.

The experts reviewed existing experience on establishing measurable criteria for the quality of learning outcomes and suggested that the feasibility study should encompass two strands, which together would embrace a wide spectrum of learning outcomes:

- A first strand could assess transversal higher-order competencies, such as critical thinking, analytic reasoning, problem-solving, or generating knowledge and the interaction between substantive and methodological expertise. These competencies are widely viewed as critical for the success of individuals and ever more relevant in the information age.
- A second strand would seek to assess discipline-related competencies, starting with one or two disciplines. This approach would require highly differentiated assessment instruments and exclude competency areas that are not easily amenable to large-scale assessment or that are not sufficiently invariant across cultures. The experts suggested that, whatever the disciplines chosen, the aim would be to assess competencies that are fundamental and "above content", i.e. with the focus on the capacity of students to extrapolate from what they have learned and apply their competencies in unfamiliar contexts, an approach similar to PISA.

The experts acknowledged that the outcomes from these two strands would speak in different ways to different stakeholders.

The experts suggested that simplistic assessment tools and an impoverished definition of assessment would pose significant threats to the credibility of the exercise. They were also concerned about the "tunnel vision" that could result from narrow assessments driving policy and practice. To counter such tendencies they suggested the feasibility study focus on a few HEIs, rather than aiming initially at large scale assessment and use measures and instruments that engage students and faculties. This would be easier for discipline-specific measures.

© OECD 2012

Establishing a "quality space" in which HEIs and systems can be meaningfully situated

The OECD made it clear that it does not aim to establish a single performance measure that will then be used for a uni-dimensional ranking of HEIs or countries. The OECD also acknowledged that any effort to bring together all HEIs on one standard would risk driving the assessment down to the lowest common denominator. Its aim is rather to establish a "multi-dimensional quality space", in which quantifiable quality criteria establish the dimensions of the space. If the concept of the "quality space" proves possible, higher education systems, institutions, departments and faculty could then be situated within this space depending on the prevalence of the different quality attributes. Consumers would then be able to choose programmes and institutions depending on the configuration of the quality attributes most relevant to them, rather than depend on uni-dimensional ratings that combine quality attributes in predefined ways, which may not necessarily be the most relevant ways for either consumers or providers.

The key question would be: which dimensions of diversity would an international assessment by the OECD capture? In this context, the experts recognised that learning outcomes are only one component of the quality of HEIs. Therefore, the "quality space" would need to recognise other dimensions of the quality of outcomes as well, such that HEIs and systems could be appropriately represented in the space, in accordance with their respective missions.

A critical question would always be how to report outcomes from such assessments, i.e. what types of reporting would prove most effective for the various stakeholders, in terms of raising outcomes and engaging personnel and HEIs in improvement. This would also include the extent to which the information individuals and HEIs receive should go beyond the performance of their own institution. One view was that all evidence from the evaluation of public policy should be made public (with appropriate analyses) to provide evidence to taxpayers and the users of higher education as to whether HEIs are delivering the expected results, to provide a basis for intervening across the systems where results in priority areas are unsatisfactory, and to improve the quality of policy debate. Other views were that the publication of outcome data may be counterproductive as it could detract from the diversity of HEIs and bias institutional behaviour.

Paris meeting

The main purpose of the second meeting of experts in Paris (OECD, 2007b) was to consider whether an AHELO would be conceptually possible, i.e. how an AHELO study might be organised and what research and policy questions it might address. The experts were tasked to review existing approaches to assessing higher education learning outcomes and to explore the design of a study to assess the feasibility of pursuing such work in an international comparative context.

The experts agreed that it was important to develop any international AHELO in a transparent and participatory manner. HEIs and their representative bodies, such as the IAU and the EUA, should be informed of ongoing work. The experts stressed that it was also important to consult relevant professional bodies, employers and experts working in the area of assessing learning outcomes.

© OECD 2012

Objectives of an international AHELO

The experts reviewed the design and operational implications of the various potential objectives for an assessment of learning outcomes, namely: *i)* to inform national policy development, *ii)* to compare learning outcomes across HEIs, *iii)* to inform institutional strategic planning, and *iv)* to inform consumer choice.

The design of an international assessment would depend on the objectives chosen and no assessment could achieve all of the goals equally well. Although establishing the objective(s) would be a policy choice, the experts considered comparing learning outcomes across HEIs and/or programmes both a relevant and, at least in principle, feasible goal and that a feasibility study would need to cover aspects of both transversal and discipline-related competencies. This would provide institutions with useful information on their strengths and weaknesses compared to objective benchmarks.

Operationalising the assessment of higher education learning outcomes internationally

The experts stressed that a fully developed assessment of different kinds of learning outcomes was beyond the scope of this initial project. They reviewed a number of existing assessments for transversal competencies including the CLA, MAPP, iSkills and the GRE and preferred the CLA approach because it tests high-level critical thinking and communicating competencies with a high degree of face-validity but suggested adding some shorter test items to the CLA type performance tasks.

The experts agreed that an assessment of subject-specific competencies would also be needed, while the diversity within subjects posed major challenges, recommended conducting the feasibility study in subject areas that have a stable core of methodologies such as engineering or economics and possibly expand the range of covered subjects over time. They noted that the Tuning project as well as the subject-specific tests in Brazil or Mexico might provide insights for this process.

The experts noted that transversal and subject-specific competencies are just two dimensions of the many learning outcomes of HEIs. Non-cognitive learning outcomes and labour market outcomes could not be assessed in the feasibility study due to time and resource constraints.

The experts stressed that assessments of higher education learning outcomes have to be sensitive to social and cultural differences and testing material should thus be checked carefully for cultural differences. But the experts did not see a general problem with cross-cultural validity.

Determining relevant units of data collection

The experts suggested establishing a sample of countries with diverse educational systems and positions on measuring higher education learning outcomes.

It was agreed that the feasibility study should involve a limited set of volunteer universities or colleges.

© OECD 2012

Most experts argued that testing should be done shortly before the end of the first academic degree, which in most countries would be a Bachelor's degree. The necessary sample size would depend on the desired statistical significance, the study design and the choice of method.

Most experts recommended that results from outcomes assessments should not be compared to an absolute quality criterion, because it is difficult to establish quality benchmarks and HEIs might oppose such a criterion-referenced approach. Instead the results should be interpreted by comparing institutional performance with benchmarks based on the performance of other participants.

The experts agreed that, in order to provide for appropriate coverage of the competency domains, the matrix sampling technique should be used, which would preclude the use of the generation of individual student scores.

Seoul meeting

The main purpose of the third expert meeting held in Seoul in October 2007 (OECD, 2007c) was to discuss the design and implementation of a feasibility study and the structure suggested for a feasibility study by the OECD and Educational Testing Service (ETS) was broadly endorsed by the experts.

Communication with stakeholders

The experts recommended that the work of assessing higher education outcomes should be viewed as a process, which includes not only designing and implementing the study, but also communicating with stakeholders to build awareness of the assessment and an understanding of its value.

The purposes of the feasibility study

The experts recommended that the two purposes of the feasibility study should be to test whether reliable cross-national comparison of higher education outcomes is scientifically possible; and whether a valid assessment can be practically implemented within HEIs.

The experts advised that the proposed feasibility study should not be regarded as a pilot study. Fuller field trials would be required in a subsequent stage. It was also agreed that in order to test the assessment effectively, a diverse selection of HEIs should participate in the feasibility study.

Construction of the assessment instrument

The experts discussed three options for constructing an assessment for a feasibility study, namely to construct an entirely new instrument for the assessment; to "internationalise" one of the existing instruments (for example the CLA); or to compose an instrument by selecting items/questions from a pool of existing instruments.

None of these approaches had unanimous support: constructing a new instrument would take a long time, internationalising an existing instrument had its difficulties, and mixing existing instruments might lose the core qualities of the separate instruments.

© OECD 2012

However, the experts agreed that it would be possible to have different approaches for the feasibility study and a subsequent main study. Since the instrument for the feasibility study needed to answer whether it is possible to measure and compare cross-national outcomes of higher education, it could perhaps be constructed from existing instruments. For a future, fully-fledged study a new instrument could be constructed.

Number of countries/languages and HEIs in the feasibility study

The experts agreed that the feasibility study should cover a minimum of three countries, include at least three languages and define the desirable outcomes of higher education (skills and knowledge) with representatives of different cultures, to find the common characteristics and a cross-national consensus.

It was also suggested that three to five HEIs per country would be enough to conduct the feasibility study. In considering whether to include more countries or HEIs in the study, the information gain was taken into account and the recommendation was to ensure that a balance be maintained in complexity and cost. This would be a necessary criterion for the successful testing of the concept.

Subjects in the feasibility study

The experts agreed that engineering, economics and biotechnology were the most interesting subjects for a feasibility study. It was agreed that one subject could be enough for the purpose of a feasibility study and that, if costs and practical concerns would allow, two subjects could be included.

Timing and duration of testing

Because of the national differences in academic year, experts agreed that a window of two months of testing would be precise enough, while still giving flexibility to the HEIs.

Experts estimated that a reasonable time length for the assessment for a student to be 1.5 to 3 hours and agreed that two hours was reasonable duration for individual students.

Using a matrix sampling approach, different sections of the assessment would be given randomly to students and no student would take the full assessments. All the test items would then be aggregated to give the complete results per institution. However, using matrix sampling means that the results are no longer meaningful for individual students.

Computer delivered assessment

The experts recommended that the assessment should be computer delivered. It was made clear that this would be an irrevocable decision.

For comparability reasons, it was essential that the sampled student take the assessment because the assessments results would be associated to background information on the students. To ensure that it was the sampled students who took the assessment, it would be advantageous if the assessment would take place at the institution, for instance in a computer lab.

© OECD 2012

Motivating institutions to participate

In the opinion of the experts, the information that the feasibility study would provide HEIs on how to improve their own teaching and learning processes should be an appealing incentive for them to participate. While some HEIs would want to participate to *show* how well they do, some would want to participate to *know* how well they do. For the purpose of the feasibility study it would not be necessary to reveal the results of the individual HEIs unless they so agreed.

Motivating students to participate

Motivating students to participate is critical to a successful feasibility study, for two reasons: to have students participate in sufficient numbers, and to ensure they make the effort to perform their best in the assessment. The experts expected that it would be harder to motivate students in the start-up phase, but it would become easier once the assessment was established.

The experts thought it important to give individual students the results on their performance in the assessment. Because of the matrix sampling design of the assessment it would not be possible to give the entire assessment score as comparison (since each student only would only do a selection of the entire assessment). The experts still deemed it important to give the corrected assessment results to the students. It was suggested by the experts to be sensitive about comparisons in order not to discourage any students.

Next steps

The insights and recommendations of the three groups of experts were subsequently shared with OECD education ministers during an informal ministerial meeting on evaluating the outcomes of higher education held on 11-12 January 2008 in Tokyo. As reported by Japan's Minister for Education, Culture, Sports, Science and Technology who chaired the meeting (Tokai, 2008), the ministers:

> Underlined the importance of establishing valid and reliable measures of learning outcomes and welcomed the initiative led by the OECD to assess the feasibility of an international study on assessment of learning outcomes [...] Underlined the need to develop and implement the work in open and transparent ways, to involve HEIs and relevant agencies in the process, and to document the conceptual underpinning of the proposed feasibility study, the criteria for success and the process to assess the validity of the measures [...] and noted that countries would base decisions on further steps on the outcomes of the feasibility study.

© OECD 2012

REFERENCES

Achard, P. (2010), "Rankings: a case of blurry pictures of the academic landscape?", *Inside Higher Ed*, Issue 21 September 2010, Washington, DC. *http://www.insidehighered.com/blogs/globalhighered/rankings_a_case_of_blurry_pictures_of_the_academic_landscape*

Baird, L. (1988), "Diverse and Subtle Arts: Assessing the Generic Outcomes of Higher Education", In C. Adelman (Ed.), *Performance and Judgment: Essays on Principles and Practice in the Assessment of College Student Learning*, U. S. Government Printing Office, Washington.

Banta, T. (2007), "A Warning on Measuring Learning Outcomes", *Inside Higher Ed*, Issue 26 January 2007, Washington, DC. *www.insidehighered.com/views/2007/01/26/banta#ixzz27ZMS4500*

Banta, T. And G. Pike (2012), "Making the Case Against – One More time", Commentary on R. Benjamin., (2012), *The Seven Red Herrings About Standardized Assessments in Higher Education*, September 2012, National Institute for Learning Outcomes Assessment, *http://learningoutcomesassessment.org/documents/HerringPaperFINAL.pdf*

Douglass, J., G. Thomson and C. Zhao (2012), "Searching for the Holy Grail of learning outcomes", *Inside Higher Ed*, Issue 20 February 2012, Washington, DC. *www.insidehighered.com/blogs/globalhighered/searching-holy-grail-learning-outcomes*

Education International (2007), *Assessing higher education learning outcomes: "PISA" for Higher Education?*, Malaga statement, November 2007, *http://download.ei-ie.org/docs/IRISDocuments/Education/Higher%20Education%20and%20Research/Higher%20Education%20Policy%20Papers/2008-00036-01-E.pdf*

European University Association (2008), *OECD project on "Assessing Higher Education Learning Outcomes": Update on developments and validation of key questions*, Barcelona meeting, January 2008.

Ewell, P. (forthcoming), "A World of Assessment: OECD's AHELO Initiative", *Change Magazine*, Vol. 44(5), pp. 33-42.

Giannakou, M. (2006), Chair's Summary, Meeting of OECD Education Ministers: *Higher Education – Quality, Equity and Efficiency*, Athens. *www.oecd.org/site/0,3407,en_21571361_36507471_1_1_1_1_1,00.html*

Green, M. (2011), Lost in Translation: Degree Definition and Quality in a Globalized World, *Change Magazine*, Issue September-October 2011, Taylor & Francis Group, Philadelphia. *www.changemag.org/Archives/Back%20Issues/2011/September-October%202011/lost-translation-full.html*

© OECD 2012

Hazelkorn, E. (2007), "Impact and Influence of League Tables and Ranking Systems on Higher Education Decision-Making", *Higher Education Management and Policy*, Vol. 19(2), OECD Publishing, Paris.

Hazelkorn, E. (2011), "It's Time to Move Beyond Rankings", *Chronicle of Higher* Education, Issue 24 October 2011, Washington, DC. http://chronicle.com/blogs/worldwise/its-time-to-move-beyond-rankings-2/28830

Herl, H., H. O'Neal Jr., G. Chung, C. Bianchi, S. Wang, R. Mayer, C. Lee, A. Choi, T. Suen and A. Tu (1999), *Final Report for Validation of Problem-Solving Measures* (CSE Technical Report 501), UCLA Center for the Study of Evaluation, Los Angeles.

Labi, A. (2007), "International Assessment Effort Raises Concerns Among education Groups", *Chronicle of Higher Education*, Issue 28 September 2007, Washington, DC.

Lederman, D. (2012), "Rifts Over Global Test of Learning", *Inside Higher Ed*, Issue 20 September 2012, Washington, DC. www.insidehighered.com/news/2012/09/20/oecd-test-student-learning-raises-hackles-us-canada-and-europe#ixzz2EVGdRknd

Lederman, D. (2008a), "Quick Takes: Caution on Plans for World Test", *Inside Higher Education*, Issue 4 April 2008,Washington, DC. www.insidehighered.com/news/2008/04/04/qt#ixzz252NtfjCn

Lederman, D. (2008b), "A Test the Education Department Doesn't Love", *Inside Higher Ed*, Issue 21 April 2008, Washington, DC. www.insidehighered.com/news/2008/04/21/oecd#ixzz252BkZiCC

Lederman, D. (2007), "A Worldwide Test for Higher Education?", Inside Higher Ed, Issue 19 September 2007, Washington, DC. www.insidehighered.com/news/2007/09/19/international

OECD (2008), *Tertiary Education for the Knowledge Society*, OECD Publishing, Paris. www.oecd.org/edu/tertiary/review

OECD (2007a), *Assessing Higher Education Learning Outcomes – Summary of a First Meeting of Experts*, Washington, 18 April 2007, OECD, Paris. www.oecd.org/edu/highereducationandadultlearning/39117243.pdf

OECD (2007b), *Assessing Higher Education Learning Outcomes – Summary of the Second Meeting of Experts*, Paris, 5-6 July 2007, OECD, Paris. www.oecd.org/edu/highereducationandadultlearning/39117295.pdf

OECD (2007c), *Assessing Higher Education Learning Outcomes – Summary of the Third Meeting of Experts*, Seoul, 26-27 October 2007, OECD, Paris. www.oecd.org/edu/highereducationandadultlearning/41059347.pdf

Olds, K. and S. Robertson (2011), On being seduced by The World University Rankings, *Inside Higher Ed*, Issue 5 October 2011, Washington, DC. http://www.insidehighered.com/blogs/globalhighered/on_being_seduced_by_the_world_university_rankings_2011_12#ixzz2528eObsp

© OECD 2012

Rauhvargers, A. (2011), *Global university rankings and their impact*, EUA Report on Rankings 2011, Brussels. *www.eua.be/Libraries/Publications_homepage_list/Global_University_Rankings_and_Their_Impact.sflb.ashx*

Rhodes, T. (2012), "Getting Serious About Assessing Authentic Student Learning", Commentary on R. Benjamin , (2012), *The Seven Red Herrings About Standardized Assessments in Higher Education*, September 2012, National Institute for Learning Outcomes Assessment, *http://learningoutcomesassessment.org/documents/HerringPaperFINAL.pdf*

The Economist (2007), "University Rankings: Measuring Mortarboards - A New Sort of Higher Education Guide for Very Discerning Customers", *The Economist*, Issue 15 November 2007. *www.economist.com/node/10143217*

Tokai, K. (2008), "Chair's Summary of the Informal OECD Ministerial Meeting on Evaluating the Outcomes of Higher Education", presented at the OECD Ministerial Meeting, Tokyo, 11-12 January 2008. *www.oecd.org/education/highereducationandadultlearning/name,70067,en.htm*

Vught, F. Van, Ed. (2009), Mapping the Higher Education Landscape, towards a European Classification of Higher Education, Springer, Dordrecht.

Wildavsky, B. (2009), "How America's mania for college rankings went global", *Washington Monthly*, Issue September-October 2009. *www.washingtonmonthly.com/college_guide/feature/international_studies.php?page=all*

© OECD 2012

NOTES

1 The letter was sent on behalf of the following organisations: the American Association of Community Colleges (AACC), the American Association of State Colleges and Universities (AASCU), the American Council on Education (ACE), the Association of American Colleges and Universities (AAC&U), the Association of American Universities (AAU), the National Association of Independent Colleges and Universities (NAICU) and the National Association of State Universities and Land-Grant Colleges (NASULGC).

2 See Chapter 1 for a description of the New Public Management theory and its implications for higher education steering and governance.

3 The series of Transatlantic Dialogue events bring together the American Council on Education (ACE), the Association of Universities and Colleges of Canada (AUCC), and the European University Association (EUA). They involve some thirty presidents, rectors, and vice-chancellors from universities in Canada, Europe and the United States.

4 It has been argued, however, that those peer reviews through academic senates or councils traditionally took place without agreed reference points among the academic community. As a result, they were often based on personal opinions of peers, which proved to be rather problematic at times because of lack of objectivity.

© OECD 2012

CHAPTER 3

DESIGN AND MANAGEMENT OF THE FEASIBILITY STUDY

This chapter first describes the general AHELO feasibility study design and how this was adapted in the context of the global financial crisis. It then goes through the details of the phases of the work and concludes by describing the various groups involved in the management of the feasibility study and providing an overview of the management activities at international, national and institutional levels.

© OECD 2012

Survey design

Following the endorsement of the OECD ministers, the AHELO feasibility study was launched in early 2008, with work effectively starting in May following the recruitment of a Senior Survey Manager at the OECD. The Senior Survey Manager's first task was to draw up a road map for the feasibility study and seek interest in participation from countries.

Study design emphasising a proof of the concept

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

First, the research approach focused on gathering evidence in support of the AHELO concept by building, as much as possible, upon existing tools and instrument materials rather than developing new assessment tools. With this *modus operandi*, it was evident that the instruments and tools chosen for the feasibility study would not in any way prejudge the instruments and tools that might be developed in the context of an eventual AHELO follow-up main survey. This point was underlined by the AHELO Group of National Experts (see Box 3.1) – which was responsible for the technical conduct of the feasibility study – when decisions on instruments and contractors were made.

Second, the research approach also allowed to artificially break down the work into several distinct strands in order to examine different types of higher education learning outcomes – e.g. generic and discipline-specific skills – as well as different approaches to assessment and testing – e.g. through performance tasks, other constructed-response tasks (CRTs) or multiple choice questions (MCQs). This purely artificial approach would yield insights on the relative merits and drawbacks of various methodologies and assessment domains and the design of an eventual AHELO follow-up main survey would most likely consider combining different types of learning outcomes and testing approaches.

Third, the research approach guided the selection of disciplines. Experts suggested engineering, biotechnology and economics for a feasibility study. However conversations with stakeholders revealed an overall belief that developing an international assessment would be far more challenging in the social sciences than STEM and it was therefore decided to focus on two contrasting disciplines from the STEM and social sciences areas:

- Engineering was chosen for the STEM areas given the large amount of work and research done to date on defining learning outcomes of engineering programmes, the reasonably well-defined occupational destinations of engineering graduates, and the well-articulated expectations of employers in the engineering sector with well established standards by professional associations.
- Economics was chosen as a field of study that leads to a broader range of occupational destinations, and with less guidance from employers on the skills and competencies which they expect economics graduates to have acquired during the course of their education.

© OECD 2012

Box 3.1 - Governance of the AHELO Feasibility Study

Governance of the AHELO feasibility study was shared between the OECD Education Policy Committee (EDPC) and the Governing Board of the OECD Programme for Institutional Management of Higher Education (IMHE GB).

The EDPC is made up of representatives of OECD member countries and it provides strategic direction for all OECD work on education. The EDPC decided to embark on the feasibility study and will take the strategic policy decision on whether to take AHELO forward beyond the feasibility study.

The IMHE Programme is open to all recognised higher education institutions (HEIs) as well as associations of HEIs and government representatives and the IMHE GB was responsible for the oversight of the AHELO feasibility study. The IMHE GB thus provided a platform for HEIs to engage with governments in AHELO so that the approaches adopted took account of institutional needs and concerns. In particular, the IMHE GB:

- determined the AHELO feasibility study policy objectives;
- ensured compliance with the policy objectives and design parameters at milestones during implementation; and
- enabled participating countries, agencies and HEIs to be fully informed of all aspects of the implementation.

However the technical nature of the project led the EDPC and the IMHE GB to establish an AHELO Group of National Experts (GNE) to oversee decisions on the methods, timing and principles of the AHELO feasibility study. The AHELO GNE acted as a subsidiary body to the IMHE GB and its members were nominated by countries (see Annex A for the list of members). The AHELO GNE also allowed countries that were not directly involved in the feasibility study to have their say and monitor progress as members of the GNE.

The AHELO GNE was the main steering mechanism for technical aspects of the feasibility study and:

- provided guidance on directions for the project, in line with the policy objectives established by the IMHE GB;
- made key decisions related to the design and conduct of the project, including tendering processes as well as the phasing and financing of the work to bridge the financing gap;
- provided input into the development of the assessment instruments and contextual surveys to ensure that the diverse cultural and curricular contexts of participating countries were reflected in the assessments;
- established priorities for indicators, analysis and data-collection instrument development;
- ensured compliance with the policy objectives established by the IMHE GB at key milestones during the implementation of the feasibility study; and
- guided the preparation, review, and completion of all documentation and reports from the AHELO feasibility study.

In order to make the feasibility study process as open and transparent as possible, an AHELO Stakeholders Consultative Group (SCG) was established to provide additional scope for consultation and dialogue with key groups interested in higher education quality. As outlined in Annex A, this group comprises representative organisations of HEIs, students, faculty unions, employers, quality assurance agencies, professional associations of engineers and economists, as well as AHELO sponsors.

© OECD 2012

While the AHELO feasibility study design essentially followed the recommendations of the 2007 expert groups presented in Chapter 2, it also embraced some of the recommendations made by stakeholders, for instance detailed advice from Education International (2007).

At their suggestion, the OECD:

- Undertook initial work prior to launching the AHELO Call for Tenders to assess whether it would be possible to agree on the learning outcomes to be assessed. To this end, a Tuning-AHELO project was undertaken in early 2009 which consisted of applying the Tuning methodology developed in the European context to a broader range of countries more reflective of global higher education systems. The outputs of this work were two reports describing the Tuning-AHELO conceptual frameworks of expected/desired learning outcomes in the Science of Economics and in Engineering (Tuning Association, 2009a, 2009b).
- Incorporated a strong contextual dimension in the AHELO feasibility study design to enable detailed analyses of the factors associated with enhanced learning outcomes. It also follows the EI recommendation that the contextual dimension be applied across the board rather than as a separate strand of work.
- Established a mechanism for consultations with stakeholders by forming the Stakeholders' Consultative Group (SCG) that met regularly throughout the course of the feasibility study (see Box 3.1).
- Designed the feasibility study in such a way that it would examine whether the AHELO test applied to diverse HEIs in terms of their mission, curriculum and student intake and reveal possible biases.

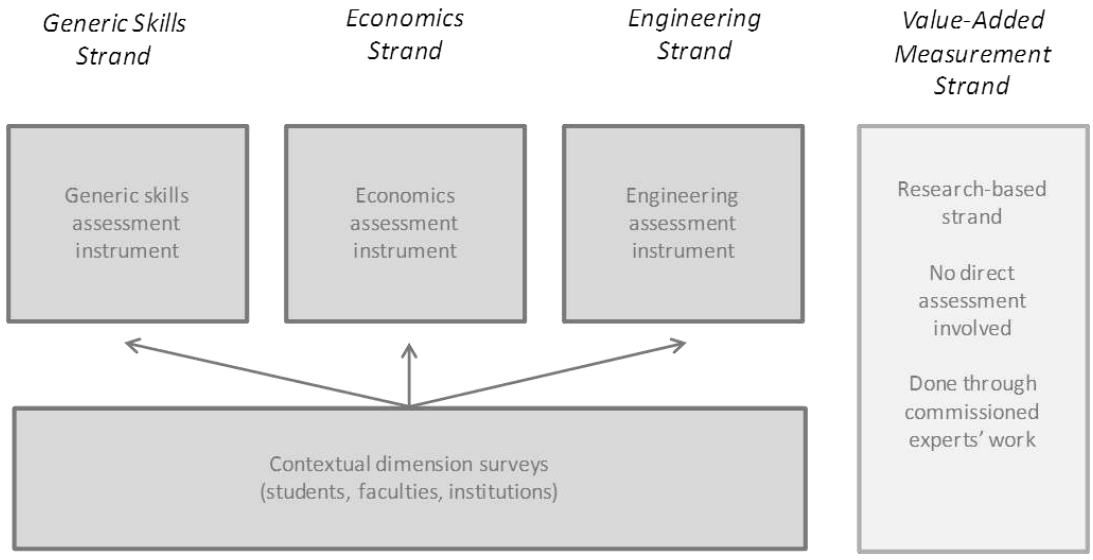
The AHELO feasibility study design made higher education institutions (HEIs) the main units of analysis and reporting, in accordance with the recommendations of the 2007 groups of experts. This meant that no attempt would be made to sample HEIs nor to develop country-level performance measures. At the same time, it was important to gauge whether an AHELO could provide valid and reliable results across the spectrum of diverse HEIs types. Thus, each participating country was encouraged to select a convenient group of volunteer HEIs that was sufficiently different in terms of mission, geographic location, student characteristics, selectivity etc. that it could reflect the range of HEIs in their system.

Four distinct but coherent strands of work

The study design of the feasibility study therefore consisted of four distinct strands of work to be developed separately but coherently (see Figure 3.1). The first three strands consisted of an assessment in three different domains: generic skills (such as critical thinking, analytical reasoning, problem-solving and written communication), economics and engineering. The last strand of work focused on the issue of value-added measurement but did not involve direct assessments of learning outcomes and value-added. Instead, it was decided that within the scope of a feasibility study, this issue would be addressed from a research perspective, through a reflection on methodological approaches, data needs of various value-added strategies, and pros and cons of different approaches.

© OECD 2012

Figure 3.1 - AHELO Feasibility Study strands of work



While it was envisaged that the three different strands of work would be undertaken separately, with different groups of countries involved in each of them, the feasibility study design also envisaged that they would be carried out coherently in terms of processes, test administration and analysis, so as to maximise synergies across the different strands, streamline communications and generate economies of scale. An additional benefit was to obtain results which would be as comparable as possible to gauge the relative merits and drawbacks of each approach.

In order to achieve coherence across the three assessment strands, it was decided to bring them under the umbrella of a single Consortium with responsibility for international project management, test administration and analysis of the feasibility study findings, while leaving scope for different organisations to be responsible for the development of instruments in each strand.

This approach had the advantage of streamlining the feasibility study processes for greater comparability of its findings, capitalising on various Consortium partners' expertise, and pooling expertise to reach superior analytical capability. The success of this strategy, however, would depend critically on the level of co-operation between the Consortium partners and the ability of the Consortium leading organisation to establish trust between the various groups involved. This was not a trivial concern due to the nature of the project which deliberately attempted to explore different types of approaches to testing which could have implications for the longer term commercial interests of partner organisations.

An additional challenge of the AHELO feasibility study design was that the instrument for the Generic Skills strand was selected early in the process, following the recommendation of the

© OECD 2012

2007 groups of experts who preferred the CLA approach because it tested high-level critical thinking and communicating competencies with a high degree of face-validity. This was considered the only available instrument with these properties at the time. The agreement in principle to go ahead with an instrument based on the CLA for the Generic Skills strand was reached in early 2009, after the AHELO Group of National Experts had verified that no Generic Skills instrument with similar properties had emerged since 2007. As a result, work on the adaptation of some components of the CLA instrument for international use started in 2009, before the AHELO Call for Tenders for the other strands was awarded. This timing mismatch limited the scope for building synergies between this strand of work and the others, whose development did not start until later in 2010 (see below). More particularly, it led to additional operational and co-ordination challenges in the development of Generic Skills instrumentation and test administration procedures.

The other key features of the AHELO feasibility study design can be summarised as follows,[1]:

- The study design made clear that the goal of the AHELO feasibility study was not to publish data on learning outcomes for participating HEIs, but rather to provide a proof of concept that it would be possible to develop such measures. The emphasis of the overall assessment design, analysis plan and international reports would thus be on:
  - identifying the methodological and technical questions raised by an international AHELO;
  - addressing these issues during implementation and in the analysis of results; and
  - providing conclusions on the outcomes of the feasibility study as well as guidance for future longer-term development of an AHELO.
- The target population was set as students about to finish a Bachelor-type degree in the Bologna degree-structure, or its equivalent for countries outside of the Bologna area.
- Both paper-and-pencil or electronic delivery were initially envisaged in the Call for Tenders. However the final design opted for electronic delivery recognising that this would likely be the preferred mode of delivery for an eventual AHELO follow-up main survey. As a result, the feasibility study would benefit from exploring practical implications of electronic delivery.

Participating countries

The experts groups in 2007 discussed at length the challenges of capturing higher education outcomes in a way in which cultural and linguistic differences were taken into account and agreed that the feasibility study should cover several quite different countries and include at least three languages (see Chapter 2).

At the same time, given that the AHELO feasibility study aims at providing a proof of concept, it was proposed to involve a maximum of five countries and ten HEIs within each country for each

© OECD 2012

strand of work. In order to ensure sufficient international variation in each strand, the following criteria were taken into account when inviting countries to participate in the different strands of work:

- *geographic origin*: ideally with countries from the five regions of the world (Africa, America, Asia, Europe and Oceania)[2];
- *language*: ideally with at least one English-speaking country, one Asian language, one Latin language and one other European language; and
- *culture:* ideally with a mix of Latin, Nordic, European, Anglo-Saxon and Asian cultures.

The final list of participating countries (see Box 3.2) seemed to be satisfactorily spread across the different assessment strands. Indeed, the distribution of countries' participation across the three strands of work ensured a reasonably balanced picture in terms of geographic, linguistic and cultural backgrounds, although countries from North America or the Asia-Pacific region language and culture could have provided further insights for the Economics strand.

Box 3.2 - Countries finally participating in each skills strand

The **Generic Skills** strand was tested in nine countries: Colombia, Egypt, Finland, Korea, Kuwait, Mexico, Norway, the Slovak Republic and the United States (in Connecticut, Missouri and Pennsylvania).

The **Economics** strand was tested in seven countries: Belgium (Flanders), Egypt, Italy, Mexico, the Netherlands, the Russian Federation and the Slovak Republic.

The **Engineering** strand was tested in nine countries: Abu Dhabi, Australia, Canada (Ontario), Colombia, Egypt, Japan, Mexico, the Russia Federation and the Slovak Republic.

Constraints and implications on the survey design

The feasibility nature of the project meant that its financing would not be part of the regular OECD programme of work and budget. Therefore, the AHELO feasibility study was to be funded from the outset through grants and voluntary contributions, as is commonly the case for new cutting-edge OECD projects such as PISA or PIAAC.

Although the OECD Secretariat staff involved in the management and conduct of the study was funded from the OECD central budget for the initial steps, the bulk of the financial resources for development work and implementation needed to be raised. To secure financial support, the first step was to convene a critical mass of OECD countries to participate in, or support financially, the AHELO feasibility study. Countries were invited to participate between July and September 2008 and by December ten countries had committed to supporting the work financially, raising some EUR 1.5 million in voluntary contributions.

Implications of the global financial crisis and other constraints on the study

However, the costs of the feasibility study were estimated at just over EUR 10 million at that time and despite countries' genuine interest, it was evident that the bulk of the funding for the

© OECD 2012

international development costs would need to come from other organisations and foundations, and contacts were made with the European Commission and foundations in Japan, Europe and the United States.

Unfortunately, soon after the project was launched, the global financial crisis in September-October 2008 and the surrounding climate of economic uncertainty affected philanthropic foundations as well as the private sector, and made funding for innovative projects like AHELO more difficult to obtain. Within the context of an economic crisis, fundraising efforts were a continuing challenge throughout the feasibility study.

Along with the funding challenge, other constraints included:

- OECD rules concerning procurement of external services, whereby projects and activities cannot operate in deficit and contracts can only be signed once funds are available to cover the whole amount of the contract;
- the costs of delays, since delaying the conclusion of contracts until funding was made available led to additional costs such as staff and meetings costs at the international and national levels; and
- risk management considerations in terms of countries and institutions' engagement, public relations and philanthropic engagement.

Overall, fundraising challenges and associated delays obliged participating countries and HEIs to extend and/or revise the arrangements they had put in place for managing the work. Despite this, the feasibility study benefited from the continued commitment and support of its participating countries: only two countries withdrew (as a result of national constraints and shifting priorities), yet they maintained their initial financial contribution. Moreover, this was offset by more countries joining the study in subsequent stages, reaching 17 participants in 25 field implementations by the end of the study.

The implications of funding constraints for public relations and philanthropic engagement were more complex. Indeed, the AHELO feasibility study was subject to high media interest from the outset. Within this context, fundraising challenges and questions as to whether the study would be able to proceed had an impact on credibility, which in turn made fundraising more challenging than had been anticipated.

Closing the funding gap

In light of these fundraising challenges and management constraints, several approaches were adopted to close the funding gap.

Reducing costs

With the GNE's support, the OECD Secretariat negotiated with the two principal contractors to reduce their costs in several ways without sacrificing the goals of the study:

- reducing the number of face-to-face expert and NPM meetings, and where possible, organising them back-to-back with GNE meetings to reduce travel costs;

© OECD 2012

- foregoing the programme leadership and institutional leadership contextual instruments;
- opting to not use multiple choice question items (MCQs) to complement the constructed-response tasks (CRTs) for the assessment in the Generic Skills strand (later restored[3]);
- developing only a provisional assessment framework for the Economics and Engineering strands, and producing an instrument that did not include innovative CRT item types;
- narrowing the scope of the Engineering assessment from two subfields to civil engineering alone; and
- deferring until a second phase the design and development of all facets of implementation including: sampling, fieldwork training, coder training, facets of quality management, data preparation and validation, and analysis and reporting.

Finding alternative funding

The second approach was to find alternative funding streams and secure broader support for the AHELO feasibility study including by:

- Calling on participating countries to increase their level of financial support to the project, which all did. All provided additional contributions, thereby maintaining momentum until completion.
- Inviting non-participating OECD member countries to support the work. England, Ireland and Sweden responded positively and offered financial support to the study.
- Increasing the number of participating countries and the EDPC and the IMHE GB agreed to extend participation to Abu Dhabi, Colombia, Egypt, Kuwait and the Russian Federation. This both increased the funding base for the study and also enhanced geographic, cultural and linguistic diversity.
- Allowing the five countries who expressed interest in participating in more than one strand to do so, thereby generating additional resources for the study.
- Continuing to solicit funding from foundations throughout the course of the study.

Reviewing the project design and timeline

The third approach to fill the funding gap was to review the project design and timeline. To this end, the AHELO GNE considered four alternative work plans to accomplish the work:

- an “initial scenario” proposed to maintain the initial objectives, scope and timeframe of the project;
- a “delay scenario” under which the study would proceed only with the work already contracted and started (the adaptation and translation of the CLA instrument for the

© OECD 2012

Generic Skills strand) while work in the disciplinary strands and contextual dimension would be put on hold until funding would make it possible to proceed;

- a "conservative scenario" to minimise the risk of uncertain fundraising outcomes and scale down the work to available budget, thereby limiting the scope of the work to the Generic Skills strand alone in the short term, for which a contract was already in place and work was underway; and
- a "phasing scenario", whereby the first phase would proceed with developing assessment frameworks and instruments in all strands of work while the implementation of the work would be deferred to a second phase and subject to funding availability.

In their deliberations, the GNE members underlined the importance of keeping momentum with the work, even if for more limited activities than initially envisaged. The GNE also stressed the value in maintaining the breadth of the feasibility study – i.e. not restricting the feasibility to one strand of work – even if this meant that the implementation of the work had to be delayed. *De facto,* this ruled out the delay and conservative scenarios, and as funding was insufficient to proceed with the initial scenario, the AHELO GNE opted to phase the work in line with funding availability.

The funding gap was eventually closed and this eventually allowed the study to proceed to its second phase, and to restore the bulk of the deferred elements. This ensured that the overall scope of the study was retained, without compromising its integrity.

In the end, participating countries bore the bulk of the AHELO feasibility study costs (84%), while contributions from non participants and foundations added another 13% and the balance was funded from OECD central costs. The AHELO feasibility study received generous support from the Lumina Foundation for Education (United States), the Compagnia di San Paolo (Italy), the Calouste Gulbenkian Foundation (Portugal), Riksbankens Jubileumfond (Sweden), the Spencer Foundation (United States) and Teagle Foundation (United States). Indirect support was also provided by the William and Flora Hewlett Foundation (United States) which contributed to the international costs of the three participating US states.

Impacts of a phased approach on the AHELO Feasibility Study timeframe

The initial timeframe for the AHELO feasibility study spanned the period from November 2009 to late 2011 but was adjusted in line with the GNE's decision to adopt the "phasing" scenario. While work on the Generic Skills strand instrument had started earlier (see above), the contract with the selected contractor could only be finalised in July 2010, and only covered the first phase of the work. The contract for the second phase was finalised in February 2012 (see Box 3.3).

© OECD 2012

Box 3.3 - Contractual arrangements

As the AHELO feasibility study work was complex and highly technical, much of the work was subcontracted to a Consortium of specialised agencies each bringing specific areas of expertise to the study. Contractual arrangements therefore had to be devised between the OECD and its main contractors, as well as between the AHELO Consortium lead organisation (ACER) and its various sub-contractors.

Strict rules and constraints govern the OECD procurement policy to ensure a competitive, transparent and impartial procurement process as well as to provide the best overall value for the Organisation while guaranteeing quality and reliability. Accordingly, an international Call for Tenders was sent out in mid 2009, except for the development of the Generic Skills instrument for which a unique service supplier had already been identified. Derogation from the tendering process for this instrument was sought and obtained from the OECD Procurement Board.

The OECD therefore contracted separately with the CAE for the adaptation of some components of the Collegiate Learning Assessment (CLA) for international use in the Generic Skills strand of the AHELO feasibility study[4], and with ACER as leader of the Consortium for the other facets of the work. This contractual architecture was unusual insofar as the OECD contracted with CLA and ACER separately for the instrumentation part of the work, while it brought them together within the Consortium for subsequent activities related to fieldwork, analysis and reporting.

Two factors contributed to the further complexity of the contracting arrangements:

- The decision to phase the study due to funding constraints resulted in more contractual arrangements for subsequent phases of the work, not only between the OECD and the AHELO Consortium, but also between ACER and its various sub-contractors.
- Some countries' late and staggered participation confirmation required a significant number of contract amendments, each multiplied by the number of main contracts covering the feasibility study work.

From a practical perspective, these additional contracts and amendments proved far more complex and time-consuming than it would have been had all countries confirmed participation at the outset and had funding been available to cover the whole scope of work from the start.

From an operational standpoint, the contractual arrangements worked fairly well overall. However, the initial decision to position the Generic Skills strand independently of the other elements in the first phase of work led to distinct processes and blurred reporting lines, especially around the transition from instrumentation to fieldwork.

The Generic Skills strand development work had an early start in January 2010. Once the performance tasks were fully translated and adapted in participating countries, some pre-implementation work for the Generic Skills strand was contracted to the CAE to adapt the CLA test system platform for international use. The decision was also made to proceed with Phase 1 activities for the other strands of work, i.e. framework and instrument development and their small-scale validation. Proceeding with the work aimed to maintain momentum and allow for

© OECD 2012

participating countries in this strand of work to be ready for implementation when funding for the second phase of the feasibility study would become available.

Building on existing CLA materials and its online platform, the pre-implementation work included developing detailed work plans for testing operations, adapting the testing platform, developing proctor training and scorer training manuals. However, this pre-implementation development work was conducted ahead of the other strands. This led to some duplication of work as well as to some inefficiency. Some elements of the detailed work plans were not used for implementation and, although elements of the manuals were integrated into the materials developed for the other strands, the manuals were not distributed *per se* for use in participating countries.

Pre-implementation work also involved adapting the CLA online platform and its three interfaces, the proctor, the student and the scorer interfaces. Proctor training videos and a proctor reference manual were developed, and translated by some participating countries. However, integrating the adapted CLA online platform with the Consortium's online platform rendered some of the developed materials, such as the proctor training videos, irrelevant for the new integrated online platform.

Despite the challenges caused by the misaligned timing of the activities in the different strands of work, once most of Phase 1 activities were completed, the roll-out of Phase 2 followed the timelines as scheduled despite a tight timeframe. Indeed, the late decision to proceed with Phase 2 meant that the Consortium had to deliver testing materials, the online platforms, survey operations procedures, as well as training NPMs and LS in a tight timeframe that left little flexibility in terms of time delivery. These time constraints impacted activity timelines at the country level where NPMs were required to react quickly, without much time for activities such as consulting on and translating the contextual dimension materials as well as verifying and validating data files.

Phases of work

Phase 1 - instrumentation and initial proof of concept

The first phase of the work (See Figure 3.2) focused on providing an initial proof of concept: the feasibility of devising assessment frameworks and instruments with sufficient validity to reflect various national, linguistic, cultural and institutional contexts.

This phase consisted of adapting and/or developing provisional assessment frameworks and instruments suitable for an international context in a first step. The assessment instruments were then validated in a second step through small-scale testing in participating countries to get a sense of their cross-linguistic and cross-cultural validity.

Development of the generic skills instrument

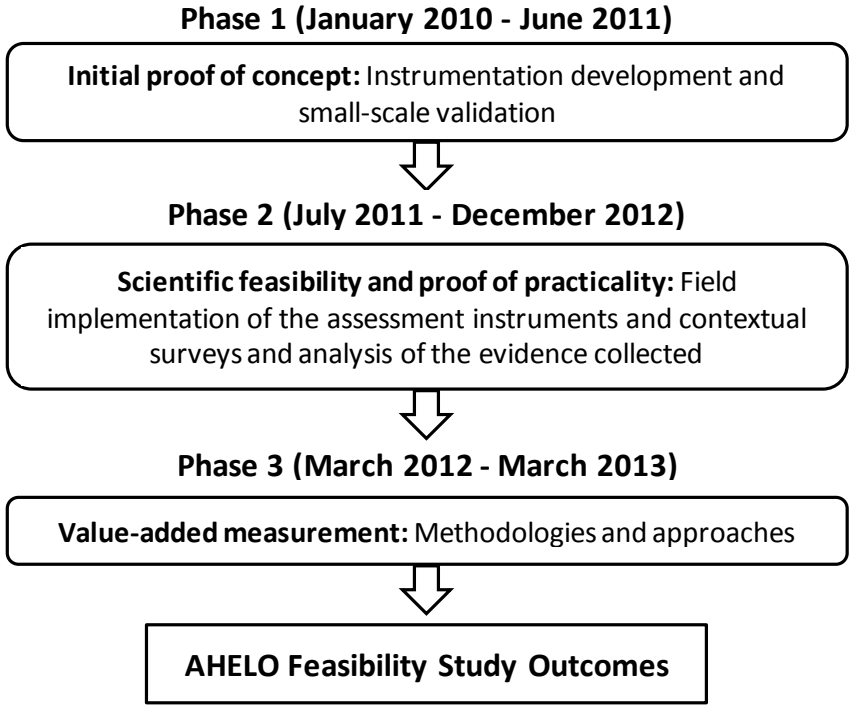
In January 2010, the CAE started adapting and "internationalising" some components of the existing US CLA instrument. In developing the generic skills assessment instrument, two CLA performance tasks were selected and adapted by participating countries in co-operation with

© OECD 2012

CAE to ensure their cross-cultural appropriateness. In addition to the two performance tasks, MCQs were added, drawing from existing pre-validated items designed to measure generic skills[5]. Participating countries were consulted and approved these items which were then adapted culturally, translated, and independently verified.

The original study design for developing the generic skills instrument relied on the adaptation of an existing instrument and did not foresee the development of an international Generic Skills framework. As the work on instrumentation progressed and MCQs were added, it became apparent that an assessment framework was needed to reach an international consensus on the generic skills to be measured[6]. Consequently, a Generic Skills Framework was developed by the AHELO Consortium (AHELO Consortium, 2012a). It describes the various conceptions of generic skills and how the construct was assessed in the AHELO feasibility study. Its development was overseen by the TAG.

Figure 3.2 - Phases of the AHELO Feasibility Study



Development of the discipline-based instruments

In July 2010, the AHELO Consortium started to develop assessment frameworks and instruments for the economics and engineering strands of work. The first task consisted of developing the assessment frameworks using the preliminary frameworks developed by the Tuning-AHELO expert groups, selected to cover a range of continents and countries, as well as different schools of thought in each discipline (Tuning Association, 2009a, 2009b).

© OECD 2012

- The development of the economics framework and assessment instrument was undertaken by Educational Testing Service (ETS) and overseen by an international economics expert group (see Annex A).
- The development of the engineering framework and assessment was undertaken by ACER, Japan's NIER, and the University of Florence and overseen by an international engineering expert group (see Annex A).

The two developed frameworks demonstrated that agreements on domain definition could be reached in two disciplinary fields, economics and engineering, and as such, provided a preliminary output of the AHELO feasibility study. They also provided useful input for test developers to design instruments to assess the performance of students who were close to obtaining their Bachelor's degree (AHELO Consortium, 2011a, 2011b).

Qualitative and quantitative validations of assessment instruments

Once expert groups and participating countries were satisfied with the materials developed, i.e. assessment frameworks and instruments, countries translated and adapted the assessment instruments to prepare for small-scale testing. Quality control during the translation process was an integral part of instrument development to ensure that the small-scale testing provided data comparable across countries (see Chapter 4).

Qualitative validation of the assessment instruments was conducted with focus groups and small numbers of students in various HEIs within each participating country.

- In the case of the Generic Skills strand, pre-testing included cognitive laboratory procedures and "think aloud" interviews conducted with student respondents to review their thought processes retrospectively and ensure that the translation and adaptation process had not altered the meaning or difficulty of the task. These interviews were conducted using a verbal probing method in which the interviewer probed for further information about the response given by the student. Although no data was collected and the validation process was not as extensive as in the disciplinary strands, it allowed identifying some cross-cultural appropriateness issues.
- In other strands, students participating in the small-scale testing also completed a questionnaire regarding their experience with the assessment instrument. In the discipline strands, faculties and ICs were also involved in the process and invited to provide feedback on the instruments. Student responses were scored using the translated scoring rubrics. The data was collected, entered by country teams, and reported to the AHELO Consortium for quantitative and qualitative validation.

The qualitative validation that followed the small-scale testing also provided aspects to be taken into account in an eventual main study. For example, feedback indicated that students' familiarity with performance tasks used for the generic skills assessment instrument varied across countries. Initial feedback also suggested that the authentic scenario tasks used for the engineering assessment instrument stimulated students' interest in the tasks.

© OECD 2012

The psychometric analysis of student performance data in the discipline strands also provided insights into the way in which individual test items functioned across different groups of students. It helped identify the need to clarify and improve the scoring rubrics areas. Results obtained from this quantitative validation were used to review and refine the assessment instruments.

Development of the contextual dimension instruments

In February 2011, the AHELO Consortium was contracted to develop the framework and survey instruments for the contextual dimension. The development work was undertaken by CHEPS at the University of Twente, ACER, and the Indiana University Centre for Postsecondary Research (CPR). It was overseen by the AHELO Technical Advisory Group (TAG).

The framework built upon earlier work undertaken by the OECD at initial stages of the study (Ewell et al., 2008, 2009). It was developed through research and consultation, and by seeking the expert opinion of a range of groups and individuals from different regions and countries[7]. Feedback from consultations was used to finalise the framework and provided a basis for survey instrument development, translation and adaptation, validation, small-scale testing and delivery (AHELO Consortium, 2011c).

Three context survey instruments were developed with the intent of identifying factors explaining observed learning outcomes of the target population: 1) a student context instrument (SCI); 2) a faculty context instrument (FCI); and 3) an institution context instrument (ICI). In addition, a range of indicators were specified for collection at the national level to provide additional context data.

Qualitative validation of the survey instruments was done through focus groups organised in Australia, Japan, the Netherlands and the United States to gather information from students and HEI personnel. Feedback from these consultations along with subsequent revisions ensured that the survey instruments reflected the framework's intent. In addition, individual countries were asked to confirm the utility of the survey instruments while verifying the translations and adaptations of country-specific terminology.

Phase 2 - practical implementation and psychometric analysis of results

The goal of the second phase was to evaluate the scientific and practical feasibility of an AHELO by focusing on the practical aspects of assessing student learning outcomes. During this phase, assessment instruments and contextual surveys were administered to diverse HEIs to explore the best ways to implicate, involve and motivate leaders, faculty and students to take part in the testing and to build an interpretive context for the learning outcomes to help identify factors leading to enhanced outcomes. The second phase involved both the implementation *per se*, including test administration and scoring of student responses, and data analysis, reporting and evaluation of the scientific and practical feasibility of an AHELO.

© OECD 2012

Implementation of the AHELO instruments to assess practical feasibility

Once the three assessment instruments and the three contextual surveys were developed, translated, adapted, pre-tested and revised, but before they were administered in each participating country, other preparations were required.

The Assessment Design and its derivative Analysis Plan (AHELO Consortium, 2010a and 2010b) were produced early in the study to guide the development and the analyses that followed. The Assessment Design identified the methods and analyses selected to assess the cross-national and cross cultural validity of the assessment and survey instruments. The derivative Analysis Plan included specifying the research questions and setting out the quantitative criteria to assess the various dimensions of the AHELO feasibility study.

The students and faculty sampling plan was developed and included the use of an unbiased probabilistic sample (AHELO Consortium, 2011d). Using probabilistic sampling guaranteed that HEIs' estimates would be comparable across institutions.

Technical Standards were developed by the Consortium to ensure a high degree of uniformity in all participating countries in achieving standardised administration of the instruments (AHELO Consortium, 2012b). The standards pertain to administration, development, implementation, analysis, reporting and review activities of the AHELO feasibility study. The standards helped ensure that all assessment activities provided data comparable across HEIs and subgroups by monitoring the quality of resources and processes, and detecting variations to design during implementation.

Online tools for student testing, data collection and scoring were developed and tested. The AHELO test system was utilised for all but one component of AHELO, namely the Generic Skills CRTs, which were administered to students and scored using the adapted CAE test system platform. The use of two different test systems required integrating their functionalities to ensure a seamless transition between the two systems for countries participating in the Generic Skills strand.

Online computer delivery also required that international and HEIs' technology and physical resources met appropriate technical standards. Both computer-based platforms had customised interfaces, required management functionality, and the capacity to be deployed and supported internationally. Testing of the online systems was conducted in participating HEIs prior to implementation ensuring that the assessment and survey instruments were deployed in all the languages of the participating countries and that technical requirements were in place at all participating HEIs. Security protocols were put in place to ensure institutional and individual confidentiality and test security.

Survey procedures and operations were developed and shared with all participating countries and training activities were conducted for test administration and scoring procedures. In November 2011 and March 2012, NPMs received training to prepare for activities such as sampling, use of the AHELO online test system, national management and procedures to implement the test instruments and contextual surveys. Training was also conducted for LSs from each country to provide them with detailed instructions on how to score CRTs, train

© OECD 2012

scorers and monitor scoring in their countries and the Sampling Manual for NPMs and ICs (AHELO Consortium, 2011d), Test Administration Manual for NPMs and ICs (AHELO Consortium, 2011e), and International Scoring Manual (AHELO Consortium, 2011f) were prepared.

The actual implementation of the test instruments and contextual surveys started at the beginning of 2012 and concluded in July 2012 with the scoring of student responses. Between February and early June, data was collected from almost 23 000 students, 4 800 faculty and more than 240 Institution Coordinators across all 17 participating countries.

To supplement the evaluation of the practical feasibility of implementing an assessment of learning outcomes on an international scale, online evaluation questionnaires were submitted to all NPMs, to a sample of ICs and to all LS. The questionnaires collected information on national contexts and asked participants to reflect on their experience and suggest improvements.

Data analysis to assess scientific feasibility

The main activities conducted once implementation was completed, included data file preparation, verification and validation, sampling analyses and weighting, scaling of student performance data, validity analyses, contextual analyses, and the production of data products and written reports.

Prior to data collection, data file structures were defined to include data from participating countries, HEIs, faculty and students, and created to enable linkage across different instruments and ensure accurate and consistent storage of information. Following data collection, data cleaning and verification procedures were put in place to validate the data collected in collaboration with the NPMs or National Centres. Use of online collection mechanisms helped ensure that the data collection was consistent and valid.

For analysis purposes, student-based weights were computed using response rates by HEI and strand, and adjusted for student non-response within an HEI or programme (as defined by the sampling plan adopted in each HEI and strand).

The student response data analyses were conducted using scaling methodology based on item response modelling (using the Rasch model) with a focus on the cross-cultural comparability of all instruments used in this study. Each individual assessment instrument item was analysed and reviewed during scaling. Item characteristic curves (ICC) were generated for every item, providing a graphical representation of item fit across the range of student abilities for each item.

In addition to item response modelling, other classical analyses were conducted to generate reliability and validity statistics, and test the efficiency of alternate scoring methods. These analyses helped identify the extent to which the assessments were successfully – or not – generalised cross-nationally, cross-culturally, cross-linguistically and cross-institutionally.

The cross-contextual validity of the test items was also explored by assessing differential item functioning (DIF). Item response theory (IRT) was used to detect variance of item parameters

© OECD 2012

across contexts. Such variance indicated that groups of students with the same ability had different probabilities of responding correctly to an item. Referred to as "item bias", this analysis indicated that the probability of successful performance on specific items was a function of group membership as well as individual ability.

The contextual data collected was used to perform general descriptive analysis in which cognitive learning outcomes were disaggregated by major student groupings and institutional characteristics. Exploratory and explanatory analyses, using multilevel modelling, were conducted to better understand the relationships between context and outcomes. The selected models not only described the relationships, but also showed how some effects varied from institution to institution, and helped identify the sources of variation.

To assess the scientific feasibility of an AHELO using the outcomes of field implementation, all information collected was considered against the technical criteria and threshold measures defined in the Technical Standards (AHELO Consortium, 2012b). The evaluation of the degree of success of international implementation was then submitted for review to the TAG to verify that technical qualities had been met.

In addition to the evaluation of scientific and practical feasibility, AHELO was designed to produce several resources to support data analysis and disseminate internationally the results from the AHELO feasibility study:

Database and codebooks

The database includes all student scores, all final and replicate weights for sampling variance computation and any context composite indices derived from the questionnaires, together with students' responses to the questionnaire and the test questions. This database will allow the OECD Secretariat, the AHELO GNE, NPMs and participating HEIs to conduct further analyses.

Compendia

The compendia were developed to include a set of tables showing statistics for every item in the questionnaires, and the relationship of background variables with performance. The tables show the percentage of students per response category and the average performance by assessment and domain for the groups of students within each category.

Technical report

The technical report summarises all technical aspects and standards of the AHELO feasibility study. It describes all data and statistical conventions and approaches applied in the study, information on test and questionnaire design, field operations, sampling, quality control mechanisms, methodologies used to analyse the data and other technical features described at a level of detail that enables researchers to understand and replicate the analyses.

Institutional report

The institutional report provides HEIs with information on their students. The report consists of the full dataset for a given institution as well as the institutional performance profile including

© OECD 2012

benchmarks of other participating HEIs in a way that does not allow the performance of individual HEIs to be identified. The report also provides guidance to the institution in terms of understanding the content in the report and interpretation of the data provided. This includes information about the reporting metric and caveats relating to comparability and use of data.

Study report

The Consortium final report to the OECD includes the methodological and technical questions raised by an international AHELO – including domain definition, conceptual assessment frameworks, validity of instruments, translation, cultural adaptation, field implementation, scoring, scaling and reliability of results, data analysis. It also documents issues that arose during implementation and in the analysis of results and offers the Consortium's conclusions on the scientific and practical outcomes of the feasibility study as well as guidance for the longer-term development of an AHELO should the initiative be taken forward.

Phase 3 - value-added methodologies and approaches

The third and last phase of the work focuses on the exploration of methodologies and approaches to capture value-added, i.e. the contribution of HEIs to students' outcomes, or 'learning gain', after taking into account the students' incoming abilities. In addition to assessing the quality of higher education by measuring learning outcomes at a single point in time, the end of the equivalent of a Bachelor's degree in this case, the AHELO feasibility study was also tasked to provide insights on whether it would be feasible to measure growth in learning.

However, measuring value-added raises a number of scientific and practical issues and imposes layers of complexity that, though theoretically well-understood, are difficult to resolve in large-scale assessments (OECD, 2008). Given the complexity of measuring learning gain, the proposed approach was to first establish the feasibility of measuring student learning outcomes at the end of the Bachelor's degree. Then, consideration would be given to the possibility and relevance of developing value-added measures in a context like AHELO.

The purpose of the value-added measurement strand was not to actually measure value-added, but rather to review and analyse possible methods to capture the learning gain that can be attributed to the HEI. The first step was to conduct a short literature review of value-added measurement approaches, building upon similar work carried out at school level by the OECD (OECD, 2008) so as to provide a general overview of value-added modelling currently used in K-12 education as well as in higher education.

In the second step, a panel of value-added experts will meet in January 2013 to discuss possible methodologies and provide guidance on the relevance and development of a value-added measurement approach for an eventual AHELO main survey and will report their findings in early 2013.

© OECD 2012

Study management and actors

The design and implementation of the AHELO feasibility study entailed collaboration among representatives of participating countries, an international Consortium contracted to implement the study, and the OECD Secretariat.

International management

As is commonly the case for large-scale international assessments undertaken by the OECD, international management activities were shared between the OECD Secretariat and a Consortium of contracted organisations with responsibility for operational issues and analysis.

The OECD Secretariat

The OECD Secretariat was responsible for the overall management of the AHELO feasibility study. This involved preparing the terms of reference of the AHELO feasibility study Call for Tenders under the guidance of the AHELO Group of National Experts (GNE), engaging consultants and contractors to implement specified activities, managing contracts with the contractors and acting as the interface between the AHELO GNE and the contractors, as well as monitoring the contractors for quality assurance purposes (OECD, 2009).

The OECD Secretariat was also responsible for building consensus among participating countries at the policy level, during the preparation of the terms of reference and at milestone points of the study, and for presenting regular progress reports to the IMHE Governing Board and the Education Policy Committee.

The OECD Secretariat also participated actively during the development of all documents and reports as well as overseeing the documentation of the project, approving all documents before they were disseminated to participating countries. This applied, in particular, to meeting documents, manuals and test materials. It is also the OECD Secretariat's role to produce the final reports from the feasibility study in collaboration with the AHELO GNE, the Technical Advisory Group (TAG), the contractors and the participating countries, on the basis of the statistical analyses and reports provided by the contractors.

The AHELO Consortium

The development of the feasibility study and the management of its implementation were the responsibility of an international Consortium[8] led by the Australian Council for Educational Research (ACER) from its appointment in mid-2010. The Consortium operated within the guidelines established by the IMHE GB and the AHELO GNE.

© OECD 2012

Box 3.4 - AHELO Consortium partners

As set out in the AHELO feasibility study terms of reference, the Consortium was formed to maximise synergies across the different strands of the feasibility study, streamline communications and generate economies of scale. The Consortium's partner organisations included:

- The Australian Council for Educational Research (ACER), an educational research organisation with increasing international reach through its work especially in the Asia-Pacific region and Europe, and through its offices in India, the Middle East, and the UK.
- The cApStAn Linguistic Quality Control Agency, widely recognised as the leader in linguistic quality control and equivalence checks and whose staff and verifiers brought extensive experience from PISA, PIAAC and other international studies to the Consortium.
- The Centre for Higher Education Policy Studies (CHEPS), a leading higher education policy centre that combines basic and applied research with education, training and consultancy activities.
- The Indiana University Center for Postsecondary Research (CPR) which has led several major studies on the student experience and houses the National Survey of Student Engagement (NSSE). This is one of the most advanced and widely adopted evidence-based institutional assessment activities. CPR is also home to the newly created National Institute for Learning Outcomes Assessment (NILOA).
- The Council for Aid to Education (CAE) which has designed a large and prominent assessment of higher education outcomes, the Collegiate Learning Assessment (CLA).
- The Educational Testing Services (ETS), recognised as one of the world's foremost educational assessment and research organisations. Given its strong international focus, ETS was well positioned to provide significant input in the areas of questionnaire development.
- The International Association for the Evaluation of Educational Achievement (IEA) Data Processing and Research Center (DPC), with more than 15 years of experience concerning data processing for large-scale surveys.
- The National Institute for Educational Policy Research (NIER), Japan's premier educational research and development agency. NIER has participated in many OECD, IEA, UNESCO and APEC projects of direct relevance to AHELO.
- SoNET systems, which has substantial experience developing and delivering large and complex software development and IT infrastructure projects, including the development of several online testing systems.
- Statistics Canada, the national statistical office of Canada which has contributed to numerous international research studies in recent years.
- The University of Florence School of Engineering, which has conducted significant work on engineering education, most recently via its leadership of the European and Global Engineering Education academic network (EUGENE) representing 76 international partners.

© OECD 2012

Within this undertaking, the Consortium was responsible for:

- Appointing a project director and staff with the skills and abilities to implement the tasks outlined in the AHELO Call for Tenders (OECD, 2009). The project director was expected to devote the vast majority of his/her time to implementing the AHELO feasibility study and responding to enquiries from and maintaining open communication with the OECD Secretariat. It was also the responsibility of the project director to establish a level of trust and understanding between and among the Consortium partner organisations, the OECD, and key representatives from participating countries.
- Establishing and supporting processes to assist the development and management of the AHELO feasibility study.
- Designing and providing necessary training for national project staff, for administration (National Project Managers [NPMs] and Institution Coordinators [ICs]) and scoring/coding procedures.
- Establishing and managing the AHELO Group of National Project Managers (NPMs), including:
  - defining the role and expected profile of NPMs;
  - setting up the intended working relationships with NPMs for consideration and agreement by the AHELO GNE at the start of the work;
  - defining the frequency and location of NPM meetings; and
  - organising and hosting NPM meetings, including some training sessions before administering the AHELO instruments.
- Establishing and managing the AHELO Technical Advisory Group (TAG) [9], including:
  - proposing potential members and a potential Chair of the TAG, for approval and appointment by the OECD Secretariat in consultation with the AHELO GNE; and
  - organising meetings of the TAG including the compensation of members.
- Liaising with the OECD Secretariat to ensure the overall success of the AHELO feasibility study, including providing advice regarding the effects on the international costs of countries that join later, withdraw, or cause delays to the project.
- Creating a dedicated AHELO website for all key documents, meeting papers and records to which participating countries and the OECD Secretariat were provided access.
- Developing and implementing quality assurance processes, including the development of procedures and a schedule for the review of data with NPMs to ensure their accuracy.

© OECD 2012

Expert groups involved

In carrying out the AHELO feasibility study, both the OECD Secretariat and the AHELO Consortium were advised and supported by a number of international expert groups that took on various roles in the steering and implementation of the study. The membership of each of these groups is presented in Annex A. These groups played a more indirect role in the steering of the AHELO feasibility study through the sharing of expertise, dialogue and providing advice.

Expert Panel for the Contextual Dimension

This group of experts was convened by the OECD to provide advice on the contents and construction of the AHELO feasibility study contextual dimension, and to develop a conceptual and analytical framework for the contextual instruments. The expert panel comprised individuals whose experience and research backgrounds centred on the effects of learning environments and pedagogy on undergraduate student learning, and the organisation and governance of national higher education systems and HEIs (Ewell et al., 2008 and 2009).

Technical Review Panel (TRP)

The Technical Review Panel (TRP) was established to review the technical aspects of proposals submitted in response to the AHELO Call for Tenders (OECD, 2009) and prepare recommendations for final approval by the AHELO GNE. The TRP included five individuals with strong policy, technical, or budget expertise in the area of large-scale international assessments. Each member was required to sign a statement indicating the absence of conflict of interest and the commitment not to work for any of the contractors or sub-contractors, in any manner, shape or form that could be related to the AHELO feasibility study for the study's duration.

Technical Advisory Group (TAG)

A Technical Advisory Group (TAG) was established to provide a mechanism through which the contractors and the AHELO GNE could draw on a wide range of international expertise and advice on the development of instruments and questionnaires as well as on operational and methodological issues, more generally. The role of the TAG was to provide advice on matters such as instrument development, translation and adaptation procedures, validation activities, scoring and verification procedures, and feasibility evaluations throughout the unrolling of the feasibility study. The TAG was also asked to identify best practices and standards for developing and implementing assessment instruments and surveys, as well as to judge whether those standards were adhered to in the best possible manner. The TAG comprised experts and individuals who were appointed for their expertise, and not to represent specific stakeholder groups.

The TAG took on increasing responsibilities over the course of the AHELO feasibility study, including serving as an expert group for the Generic Skills strand and the contextual dimension, as well as providing overall quality control for the study[10].

© OECD 2012

Other experts groups

Other expert groups were involved in the development of the feasibility study. They are mentioned in Chapter 4 and Annex A.

National and institutional coordination

National authorities of participating countries were responsible for establishing a National Centre (NC) and nominating an NPM following guidelines provided by the AHELO Consortium for each of the three strands of work. NPMs were responsible for overseeing the national implementation of the feasibility study. In some countries, the management of the implementation was the responsibility of education ministries. In others, the management was overseen by senior faculty or employees of independent educational research organisations.

National Project Managers liaised with the Consortium on all issues in their country related to the implementation of the AHELO feasibility study. They played a vital role in ensuring that the AHELO feasibility study was administered in accordance with prescribed technical standards and survey operation guidelines, and in documenting processes implemented at national level for the completion of the study's final reports. For some participating countries, the same individuals served as representatives to both the AHELO GNE and the AHELO NPM.

Within the framework and standards established by the AHELO Consortium, the NPMs were responsible in their countries for:

- managing the translation of the assessment instruments and contextual surveys into the languages to be used in their country;
- coordinating survey operations activities in participating HEIs; and
- submitting data to the Consortium, co-operating in the cleaning of data and preparing national reports.

National Project Managers were also responsible for nominating a Lead Scorer (LS) responsible for scoring all the student responses. In countries participating in more than one strand, one LS was required for each strand. LSs were supported by a national scoring team.

Each participating HEI within a country nominated an IC. The role of the IC involved liaising closely with the NPM, assisting the NPM with drawing samples of students and faculty within the HEI/department, providing institutional information, and working with test administrators (TAs) to organise the administration of assessment and context instruments at institutional level.

The TAs worked closely with ICs to administer tests to students in HEIs. TAs were expected to have experience in managing and supervising tests and exams in university settings and to be familiar with the use of computer-based deployment systems. It was also required that TAs would not have a direct personal or professional relationship with any of the students in the testing sessions they administered.

© OECD 2012

REFERENCES

AHELO Consortium (2012a), Generic Skills Assessment Framework

AHELO Consortium (2012b), *AHELO Technical Standards*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2012)16&doclanguage=en

AHELO Consortium (2011a), *Economics Assessment Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)19/ANN3/FINAL&doclanguage=en

AHELO Consortium (2011b), *Engineering Assessment Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)19/ANN5/FINAL&doclanguage=en

AHELO Consortium (2011c), *Contextual Dimension Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)21/ann1/final&doclanguage=en

AHELO Consortium (2011d), *Sampling Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN3/FINAL&doclanguage=en

AHELO Consortium (2011e), *Test Administration Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN5/FINAL&doclanguage=en

AHELO Consortium (2011f), *International Scoring Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN2/FINAL&doclanguage=en

AHELO Consortium (2010a), *AHELO Assessment Design*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2010)17&doclanguage=en

AHELO Consortium (2010b), *Analysis Plan*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2010)18&doclanguage=en

Ewell, P. T. et al. (2009), *Analytical framework for the contextual dimension of the AHELO feasibility study*, OECD, Paris.
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2009)9&docLanguage=En

© OECD 2012

Ewell, P. T. et al. (2008), "Report on the 1st meeting of experts on the AHELO contextual dimension", paper presented to the Group of National Experts, OECD, Paris, 17-18 December 2008. *http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2008)5&docLanguage=En*

OECD (2009), *Call for Tenders: Feasibility Study for an Assessment of Higher Education Learning Outcomes (AHELO)*, OECD, Paris. *www.oecd.org/edu/ahelo/callfortenders*

OECD (2008), *Measuring Improvements in Learning Outcomes,* OECD Publishing, Paris.

Tuning Association (2009a), "A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Economics", OECD Education Working Papers No. 59, OECD Publishing, Paris. *www.oecd-ilibrary.org/education/tuning-ahelo-conceptual-framework-of-expected-and-desired-learning-outcomes-in-economics_5kghtchwb3nn-en*

Tuning Association (2009b), "A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering", OECD Education Working Papers No. 60, OECD Publishing, Paris. *www.oecd-ilibrary.org/education/a-tuning-ahelo-conceptual-framework-of-expected-desired-learning-outcomes-in-engineering_5kghtchn8mbn-en*

© OECD 2012

NOTES

1 A more complete description is available in the terms of reference of the AHELO Call for Tenders (OECD, 2009).

2 In the initial stages of the AHELO feasibility study, participation in the study was restricted to OECD member countries until a number of non-member countries expressed interest. Participation in the study was then opened to a broader range of participants.

3 The GNE decision to not use or develop MCQs for the Generic Skills assessment was later reversed based on the TAG recommendation to add a subset of MCQs for validity and reliability analyses. These were kindly provided by ACER at no extra cost to the project.

4 Work contracted directly with CAE also included some pre-implementation work related to the internationalisation of the CLA online testing platform for use outside of the U.S. context.

5 In April 2011, the GNE decided to follow the AHELO TAG's recommendation to complement the performance tasks with MCQs, given their importance for scaling, psychometric equating, quality control, and reliability.

6 The AHELO TAG served as the generic skills expert group and provided general monitoring of the development work for this strand.

7 Notably the AHELO GNE, the TAG, NPMs, the SCG, the Economics Expert Group, Engineering Expert Group and the AHELO Consortium.

8 The decision to select the ACER-led Consortium was made at the third meeting of the AHELO GNE on 18-19 November 2009, after a review of proposals received in response to the AHELO Call for Tenders (OECD, 2009).

9 The management of the TAG was under the responsibility of the AHELO Consortium until February 2012 when the AHELO GNE transferred the responsibility of the TAG management to the OECD Secretariat as a way to strengthen its independence in the concluding stages of the study.

10 It became apparent after a few meetings that the sampling and psychometric experts in the initial group found it difficult to be deeply involved in the work, and new experts were thus added to the group in 2012 to provide stronger technical oversight on these critical issues at the implementation and analysis stage of the project.

© OECD 2012

CHAPTER 4

INSTRUMENT DEVELOPMENT

This chapter first provides a brief overview of the standard process used in developing instruments to measure student performance and collect contextual data in educational settings. It then sets out a description of the frameworks and assessment instruments as well as the contextual surveys developed for the AHELO feasibility study.

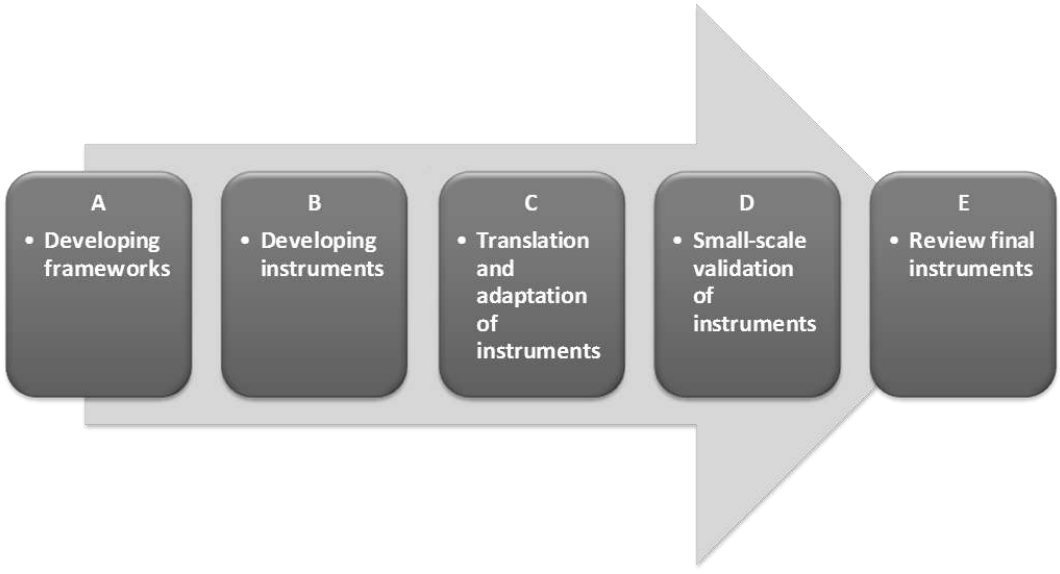
© OECD 2012

The quality of the AHELO feasibility study results relies greatly on the quality of the instruments developed and used to assess higher education students' learning outcomes. In designing the AHELO feasibility study, two types of data collection instruments were considered: i) student learning outcomes assessment instruments and ii) contextual (background) survey instruments. While the assessment instruments provided specific information on student performance on the learning outcomes for the domain tested, contextual survey instruments gathered background information from students, faculties, higher education institutions (HEIs) and National Project Managers (NPMs) to shed light and help contextualise differences observed in student performance. This chapter describes the instrumentation development process for each of the AHELO assessment instruments as well as the contextual surveys administered as part of the AHELO feasibility study.

The instrumentation process

Developing assessment instruments to measure student learning outcomes involves a wide range of activities from designing the instruments to validating the quality of the final instruments. Figure 4.1 below illustrates the five steps generally followed for developing assessment and survey instruments in cross-cultural educational settings. Each step is briefly described in the following section, along with its application within the context of the AHELO feasibility study.

Figure 4.1 - Instrument development: from design to final review



© OECD 2012

Step A: Developing assessment frameworks

The first step of the process (A in Figure 4.1) generally entails developing the assessment framework so as to establish the purpose of the assessment and outline an agreed upon definition of the domain to be tested. The instruments are then developed based on this agreed upon definition. The framework provides the link between the construct being assessed and the test outcomes. A well-designed framework includes three essential components: i) a clear definition of what is being assessed; ii) a description of the items to be used and the mix of different types of items; and iii) the basis for interpreting the results.

For the purpose of the AHELO feasibility study, full coverage of internationally-agreed domain assessment frameworks was not necessary since the goal was to provide a proof of concept that an assessment of learning outcomes could be carried out across countries. What mattered most was to demonstrate that some aspects of an assessment framework could be agreed upon in contrasted national and institutional settings and could be validly and reliably measured across diverse countries and higher education institutions (HEIs). Consequently, the decision was made to develop "provisional frameworks" not meant to be exhaustive and final, but which should contain most of those elements which the majority of experts in each domain agree are essential.

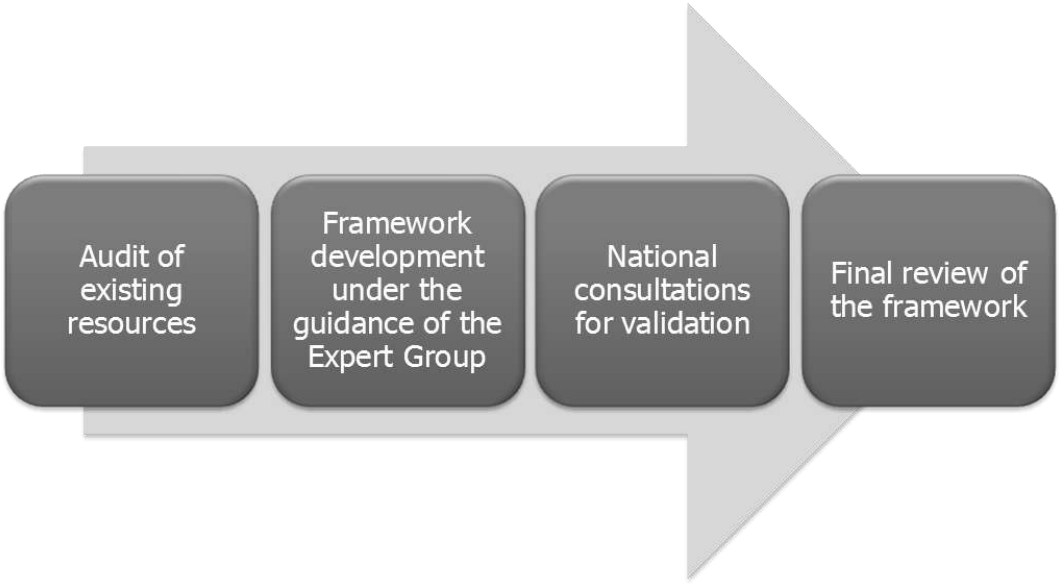
In order to develop provisional framework with an international scope, reaching international agreement with consultation and review is essential. Cross-cultural comparisons of academic performance require that different countries, and even different HEIs within countries, agree on the definition of the domain to be tested. This is one of the major difficulties with making cross-cultural comparisons.

Given the wide-range of countries participating in each strand of the AHELO feasibility study, stakeholder agreement on the different frameworks was an important step in the framework development process. Even though the frameworks developed were only to be considered provisional, framework development still required substantive input from both participating countries and subject-matter experts to ensure that the resulting instruments would be valid and would reflect the cultural context of the countries in which the AHELO feasibility study was implemented.

The four AHELO feasibility study frameworks, namely the i) Generic Skills assessment framework; ii) the Economics assessment framework; iii) the Engineering assessment framework; and iv) the Contextual Dimension framework, were developed using the process illustrated in Figure 4.2 below, albeit with fast-tracked procedures in the case of the Generic Skills assessment framework (more details are provided later in this chapter).

© OECD 2012

Figure 4.2 - Developing frameworks



The first stage of the framework development process used was to conduct audits of existing materials.

- The development of the assessment framework for the Generic Skills strand built upon existing work done for the Collegiate Learning Assessment (CLA). But given the need to develop a framework international in scope, a review of literature from different cultures was conducted to build on the existing CLA theoretical basis[1].
- For the two discipline strands, ground work was undertaken by the OECD in 2009 via the Tuning-AHELO project, which applied the Tuning methodology to reach consensus on expected learning outcomes in economics and engineering undergraduate programmes (Tuning association, 2009a, 2009b). The economics and engineering test developers then built upon these Tuning-AHELO frameworks in creating their provisional frameworks. Audits were conducted based on input from two international Expert Groups who provided suggestions for additional sources and connections within participating countries.
- For the contextual dimension, survey developers were guided by the conceptual and analytical framework established by a group of commissioned contextual dimension experts (Ewell et al., 2009), and the results of a prioritisation process of the underlying contextual variables by AHELO participating countries. In addition, an audit of existing survey resources and practices was carried out, focusing on initiatives that were international, national and institutional in nature.

© OECD 2012

The second stage of the framework development process was to review the draft versions of the frameworks that were developed using existing materials. In the discipline strands, two subject-matter Expert Groups (Economics and Engineering), drawn from participating countries and key international organisations, were established to review the draft framework specifications and content in preparation of a version for wider and targeted consultations. The Technical Advisory Group (TAG) was responsible for oversight of the development of the Generic Skills strand and Contextual Dimension frameworks. For the contextual dimension, in addition to oversight by the TAG, there were also consultations with the Stakeholders' Consultative Group (SCG).

The third stage consisted of national consultations for validation. This involved submitting the assessment frameworks to the AHELO Group of National Experts (GNE), National Project Managers (NPMs) and relevant experts in all participating countries.

Finally, the last stage was to conduct a final review of the assessment frameworks based on the results of the consultations. Further details on each of the four frameworks developed for the AHELO feasibility study are provided later in this chapter.

Step B: Developing assessment instruments

The second step of the instrument development process (B in Figure 4.1) consists of developing assessment and survey instruments reflective of the frameworks developed. The use of the table of specifications usually guides instrument developers to ensure that items created, or selected and adapted, match the learning outcomes and contribute to the purpose of the instrument.

Typically, this process starts with the creation or selection of items that match the table of specifications. Assessment and survey instruments are constructed by either developing entirely new materials created for the specific testing/survey situation, or by using existing instrument(s), or parts of it, possibly with modifications of existing items. Once the items are written, selected, or modified, subject-matter experts and assessment experts review the drafted instrument to ensure that each item is relevant to the test, and that the resulting combination of all items reflects the table of specifications and is aligned with the assessment framework. In addition to being submitted to experts, drafted items and instrument may also be submitted to a reference group representing stakeholders to ensure that the resulting assessment materials are approved by potential users of the assessment results.

In the context of the AHELO feasibility study, the design of the study involved the development of an assessment instrument for each of the three strands to be assessed (generic skills, economics and engineering) as well as three survey instruments to collect contextual information through a Student Context Instrument (SCI), a Faculty Context Instrument (FCI) and an Institution Context Instrument (ICI). In addition, a range of indicators were specified for collection at the national level to provide additional context.

It was decided early in the design process that assessment instruments would require, at most, 90 minutes testing time[2] while contextual survey instruments would require a maximum response time of 15 minutes for the student and faculty surveys.

© OECD 2012

The nature of the feasibility study work had the following implications on the instrument development process:

- Because the focus of the AHELO feasibility study was on providing proof of concept on whether it was possible to develop international measures of learning outcomes, there was no need to develop perfect and exhaustive instruments to measure higher education learning outcomes and contexts – i.e. covering all possible aspects of an ideal framework.
- The timeframe to accomplish the work was more limited than for a full-scale study and left little scope to develop entirely new test materials.
- Although developing instruments for use in different national, cultural, linguistic and institutional settings presents many challenges, using existing material developed in a specific national context within the context of the AHELO feasibility study was acceptable to prove the concept, provided that the material was adequately adapted to be implemented internationally.
- In developing instruments, it was important to ensure that what was being tested was valid in as many as possible institutional and national contexts, e.g. through a cross-national review of assessment frameworks and instruments to check for cultural and linguistic appropriateness as well as a review of their suitability in diverse institutional contexts.
- Lastly, in implementing instruments, emphasis was placed on ensuring that the resulting measures of learning outcomes would be valid, reliable and free of bias to the extent possible, e.g. through adequate processes for the translation and cultural adaptation of instruments, the sampling of students and faculties within HEIs, and the scoring of tasks, scaling, etc.

All three assessment instruments were developed with the intent to include a broad sample of items covering a range of difficulty to enable the strengths and weaknesses of populations and key subgroups to be determined with respect to the components of each assessment competency.

For each assessment instrument, two types of items were developed:

- The first item type had a constructed-response format in which students had to provide their own responses to questions.
- The second item type was a multiple-choice format in which students were asked to choose the correct answer to a question out of four possible choices.

All three assessment instruments and the three contextual surveys were administered electronically during the implementation phase of the feasibility study, with the exception of a few students in Kuwait who responded on paper[3]. More details on the delivery mechanisms are provided in Chapter 5.

© OECD 2012

Step C: Translating and adapting assessment instruments and surveys

The third step of the process (C in Figure 4.1) entails the national translation and adaptation of the assessment and survey instruments. In assessments in which more than one language is required, careful consideration to translation and adaptation of the assessment and survey instruments is needed to ensure cross-language comparability. Literature on empirical comparative research indicates that translation issues are one of the most frequent problems in cross-cultural surveys. Translation errors are much more frequent than other problems such as clearly identified discrepancies due to cultural biases or curricular differences (Harkness et al., 2003; Hambleton et al., 2005).

Guidelines and quality controls guide the translation and adaptation (localisation) process and ensure that the translated and adapted items and instruments in all the various languages are faithful to the source version of the assessment materials and equivalent to one another. To ensure the quality of the localisation process used for the AHELO feasibility study instruments and surveys, a set of procedures was structured in two distinct parts[4]:

- Quality assurance procedures included defining the localisation design, preparing translation and adaptation guidelines (including criteria for the selection of translators), resolving early on potential equivalence issues, training national teams of translators and related assistance; and providing consolidated monitoring instruments in which both team recommendations and translation/adaptation issues addressed by countries are documented, followed up and archived.
- Quality control procedures, checking whether the standards are met and proposing corrective action when not.

  Quality control procedures included thorough verification of target versions against source versions, effective reporting of residual errors and undocumented deviations, expert advice in case corrective action was needed, a final check procedure, and standardised quantitative and qualitative reports on how effectively the different target versions followed the translation and adaptation guidelines.

Detailed information on step C is included later in this chapter.

Step D: Small-scale validation of the assessment and survey instruments

The fourth step of the process (D in Figure 4.1) consists of the small-scale validation of the assessment and survey instruments. This step involves pre-testing, or pilot-testing the items developed with students who have the same characteristics as the population targeted for the assessment situation. This step provides the opportunity to assess the quality of the items and their appropriateness for the testing situation. Where pre-testing settings include a minimum of 200 respondents, analyses of the pre-test data provides a basis for selecting or revising items for the final instrument.

In addition to providing data, pre-testing is also an opportunity to collect feedback from students, as well as from teachers or faculty, on the instrument and its items, including their

© OECD 2012

insights on the instrument characteristics such as test length, difficulty, and relevance to the discipline being tested.

For the AHELO feasibility study, once the instruments were translated and adapted for each country, they were pre-tested and refined based on findings. The pre-testing of each instrument involved a small number of students per participating country and in some cases, comprised cognitive laboratory procedures, with an initial cognitive walk-through by analysts, and a set of cognitive interviews with respondents to review their thought processes retrospectively. Respondents were then probed to identify particular steps that might have been problematic.

Step E: Final review of the assessment and survey instruments

The fifth and last step (E in Figure 4.1) in the development of assessment and survey instruments is the final review. Results from small-scale validation activities, feedback collected from respondents, consultations conducted with stakeholders all contribute to the final review of the instruments.

Although the previous steps of the instrument and survey development process produced the final version of the instrument, only the results of the quantitative analyses determine the final set of items that would be used to interpret the results. Despite careful application of the rules throughout the instrument and survey development process, it is common that some items do not perform as expected and are discarded from data analyses. Consequently, they not contribute to the final assessment results.

We now examine in details how this development process played out in the Generic Skills, Economics and Engineering strands.

The AHELO feasibility study Generic Skills Assessment

As indicated earlier, the development of the Generic Skills instrumentation did not follow the usual development process where the framework is developed first and then followed by the instrument development. The development work started with the adaptation of one component of the existing US Collegiate Learning Assessment (CLA) instrument (the constructed-response section). The development work for the constructed-response section was contracted to the Council for Aid to Education (CAE), which operates the CLA. The work consisted of selecting two CLA performance tasks (PTs) to assess generic skills, and adapting them so that they would be suitable to an international context, and administered in 90 minutes (see Annex B for an illustrative performance task used in the AHELO feasibility study Generic Skills assessment instrument).

The constructed-response section was complemented by MCQs[5], drawing upon an existing cross-curricular generic skills test developed by Australian Council for Educational Research (ACER). These MCQs required an additional 30 minutes of testing for a total administration time for the Generic Skills instrument of 120 minutes. Two key issues in selecting the MCQs were considered. First, the MCQs needed to be suitable to translate into a range of languages. Second, they needed to be appropriate to use in different cultures. Overall, the items needed

© OECD 2012

to be international in perspective, capable of being appropriately translated, and accessible to higher education students. Within the limited amount of testing time available, and taking account of the few items it was possible to include, the intent was to collate items that would represent a fairly broad and balanced sample of generic thinking skills that can be reported in a single score[6].

Work on the AHELO Generic Skills Assessment Framework began in late July 2011 following confirmation of the implementation phase by the AHELO Group of National Experts (GNE). The development work was led by ACER in consultation with CAE, and with oversight by the TAG. A first draft was produced and submitted to the TAG for review. Further consultation and development strategies were conducted and the framework was finalised by mid 2012 (AHELO Consortium, 2012a).

Development of the Generic Skills assessment framework

Many countries are placing ever more emphasis on understanding how to produce graduates with general as well as specific skills. These generic skills, or competencies, are often called "core"; "key" or "workplace" (transferrable) skills. The AHELO feasibility study's Generic Skills strand explores the possibility of measuring such learning outcomes across HEIs, cultures, and languages.

For the purpose of the AHELO feasibility study, generic skills are considered the general abilities, capacities or capabilities deployed in all kinds of cognitive activity. These outcomes are seen as broad, enduring skills that, although not domain specific, are developed by frequent practice and exercise in specific domains. These skills are those that would be desired of any student regardless of subject matter or discipline, and are teachable and learnable. They are distinguished from the knowledge or skills of particular performances or domains. Such general abilities are the flexible, fluid capacities developed by and used in learning new things.

Given the nature of the generic skills concept, there are several major challenges in developing valid measures of generic skills in higher education. Firstly, different stakeholders in different countries must have the same understanding of the concept. Secondly, in developing generic skills measurements, generic skills may not be separable from the knowledge or skills of particular performances or domains. This could mean that a student's ability to display a generic skill in a particular context could, in part, be determined by familiarity with the context (Clanchy and Ballard, 1995). While the issue of testing generic skills can be highly contentious, it has been actively undertaken by testing agencies for several decades.

The original study design to assess generic skills sought to adapt an existing instrument. It did not include developing an international version of an assessment framework. As the work on instrumentation evolved and MCQs were added, it became apparent that an international consensus was needed on the generic skills to be measured. This contributed to the decision to develop the Generic Skills Assessment Framework (AHELO Consortium, 2012a).

The concept of generic skills in the AHELO feasibility study addresses the challenge of generic skills testing by targeting the common core of generic abilities. The Generic Skills Assessment Framework (AHELO Consortium, 2012a) describes conceptions of generic skills and the

© OECD 2012

construct used in the AHELO feasibility study. It explains the assessment that is used and how it relates to this construct. The framework was developed within a short timeframe and in a context in which there is little international consensus about "generic skills", and debate as to whether they exist outside of professional, cultural or disciplinary contexts. As such, the framework focuses on what appears to be the most significant and general skills on which there is most agreement.

Although the Generic Skills Assessment Framework (AHELO Consortium, 2012a) was developed at about the same time as the MCQs, and not prior to the development of the instrument as is usually the case, consultations were conducted with participating countries to determine the content of the Generic Skills instrument.

- To determine which MCQs to include in the instrument, a list of items was sent to NPMs who were asked to indicate whether each item was suitable or not for use in their country. Despite clear country differences in their selection of MCQs, convergence emerged on which items were most suitable for use.
- In the development of the Generic Skills CRTs, NPMs were also asked to provide feedback on the suitability of a number of performance tasks for use in their cultural context.

Both consultation processes indicated a certain level of international agreement on the assessment of generic skills and contributed to the final shape of the Generic Skills Assessment Framework (AHELO Consortium, 2012a).

However, due to time constraints, the framework development did not proceed with an achieved consensus of an international group of experts. Broad consultation with relevant experts from around the world would be required to make it more reflective of an international consensus on which aspects are important to assess and the degree to which the framework accounts for institutional, language and cultural differences. The framework does, however, summarise the approach used to assess generic skills in the AHELO feasibility study.

Development of the Generic Skills instrument

For the assessment of the generic skills, two types of assessment items were used:

- CRTs, also referred to as PTs, adapted by the CAE from the CLA; and
- A set of MCQs provided by ACER.

The comments above about the nature of generic skills and the characteristics of a generic skills assessment apply equally to the CRTs and the MCQs. The CRTs focus more on critical thinking than the MCQs as the generative nature of the CRTs lends itself to evaluative questions. The MCQs deal with a wider range of topics and various types of thinking than the CRTs, but they are not testing generative thinking or writing skills. Complementing each other, both the CRTs and MCQs aim at the central core of cognitive skills (AHELO Consortium, 2012a).

To better reflect the framework targeting the common core of generic abilities, the selection of items for the assessment instrument was guided by the principle to minimise the importance of

© OECD 2012

prior knowledge or familiarity with context in the assessed performance. The items selected presented students with unfamiliar or generally accessible material, and asked them to demonstrate their ability to interpret this material. This strategy aimed to minimise concerns about the potentially confounding influence of demography or context on performance as much as possible.

The selection of the two CLA performance tasks was done in February 2010 by the NPMs of participating countries at the time[7]. Selecting the two performance tasks began with an initial review by the NPMs of a subset of nine CLA performance tasks used in the U.S. with the goal of selecting two that would be most appropriate for international implementation. Choices were constrained by time and budgetary constraints, since only two PTs could be selected (instead of three initially envisaged). Out of four PTs considered suitable and valid in an international context, NPMs selected two for modification and use in the AHELO feasibility study. The subset of PTs was selected using the following criteria: (a) universality of PT theme, (b) ease of translation based on complexity of language in PT, (c) ease of scoring (based on U.S. experience with the CLA).

Following the selection of the two PTs, NPMs reviewed the contents and suggested modifications to fit within their individual country's context following agreed upon procedures. Subsequently, each country's recommended modifications were collected for review and consensus was reached on the two PTs based on the CAE and NPMs' evaluations and recommendations for final development. The CAE modified the two PTs (tasks, scoring rubrics and IT administration procedures) following agreed upon modifications and resubmitted the two PTs for country approval.

Subsequently, the two PTs were translated and adapted following an agreed upon localisation process. The localisation process included conducting cognitive workshops to obtain information on how students think when undertaking the performance tasks. In each participating country, a small sample of students was asked to "think aloud" as they carried out the tasks. Extensive research showed that "thinking aloud" provided insights into the students' thinking process and thereby verified that the thinking elicited by the performance task was the thinking sought. This process was used to verify that the performance tasks measured the same thinking across countries

The MCQs used as the second type of items for the assessment of generic skills, were selected from a catalogue of pre-existing items that had already been developed by test writers at ACER. In the limited time available to develop these items, it was not possible to develop a shared understanding of generic skills across countries, or to conduct a pilot test of the recommended items. However, participating countries were consulted on the selection of those MCQs that were then adapted culturally, translated, and independently verified.

Item selection for the MCQs was based on two factors: suitability for translation and cross-cultural appropriateness. The MCQs needed to be suitable to translate into a range of languages and appropriate for use in different cultures. These factors also drew attention to the diversity of candidate abilities which were likely to be encountered, and the subsequent need for MCQs to be of various levels of difficulty. Overall, the items needed to be perceived as

© OECD 2012

valid internationally, capable of being appropriately translated, and accessible to higher education students.

The Generic Skills assessment instrument was administered between February 2012 and June 2012 in 9 of the 17 AHELO feasibility study participating countries. A total of 98 HEIs participated in this strand, involving 10 657 students coming from different fields of education.

The AHELO feasibility study Economics Assessment

The process of development of assessment frameworks was similar for both discipline-specific strands. For the sake of clarity we are presenting these separately but this means that similar information is repeated under each strand.

The development of the framework and assessment materials for the Economics strand was undertaken by Educational Testing Service (ETS) as part of the AHELO Consortium, along with the contribution of several international consultants, while the work was overseen by an international Economics Expert Group. The framework was based on the "Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics" (Tuning Association, 2009a) and the "QAA Subject Benchmark Statement for Economics 2007"[8] (QAA, 2007), as well as with the consensus of the Economics Expert Group.

The Economics Assessment Framework defines the domain to be tested and specifies the learning outcomes for students in the target group. The Economics assessment instrument comprises both a CRT and MCQs. It is designed to be completed in 90 minutes.

The Economics assessment instrument assesses the skills and knowledge of final-year bachelor's degree students and comprises one CRT and 45 MCQs to provide additional coverage (see Annex B for illustrative items used in the AHELO feasibility study Economics assessment instrument). The items were translated and adapted by participating countries, and their qualitative validation was completed by focus groups and small numbers of students in various HEIs within each country. Analysis of the qualitative validation results guided the revision and the further development of the instrument for the main administration.

Development of the Economics assessment framework

The AHELO Economics Assessment Framework (AHELO Consortium, 2011a) was the guiding document during instrument development. Materials were developed in direct consultation with the learning outcomes and competencies specified in the framework. The framework defines the domain to be tested as follows:

> The AHELO Economics Assessment does not focus on the recall of factual knowledge, but rather focuses on "above content" skills including application of concepts, use of appropriate statistical and non-statistical tools, drawing conclusions, recommending policy, and being conversant with the "language of Economics".

The framework is based on the following five learning outcomes, all of which specify outcomes which students should be able to achieve by the end of their bachelor's degrees:

© OECD 2012

- demonstrate subject knowledge and understanding;
- demonstrate subject knowledge and its application to real world problems;
- demonstrate the ability to make effective use of relevant data and quantitative methods;
- demonstrate the ability to communicate to specialists and non-specialists; and
- demonstrate the ability to acquire independent learning skills.

The framework also stipulates that assessment of these learning outcomes should require students to use the four competencies below:

- abstraction
- analysis, deduction and induction
- quantification and design
- framing

This approach used for the Economics Framework development focused on developing a provisional framework with the following characteristics:

- broadly reflects the current thinking of experts in higher economics education;
- takes into account the characteristics of the target population – in this case, final year "first-cycle" (Bachelor's Degree) economics students;
- defines the subject domain in terms of features such as content areas, skills and processes and different schools of thought;
- is specific enough to be useful in the instrument development process, but not so limited that the opportunity for the assessment of integrated skills, conceptual understandings and "above content" learning outcomes is eliminated; and
- takes into account cultural and language differences of participating countries.

From a technical perspective, the Economics Assessment Framework fulfils the requirement of a well-designed framework. It defines the domain to be tested and specifies the expected learning outcomes for students in the target population. It also offers an overview of the instrumentation required to measure the competencies, with discussion of issues such as time, language level, item type, scoring, assessment delivery and administration, and reporting.

International development and validation of the Economics Assessment Framework suggests it is possible to define discipline-specific learning outcomes internationally. It was not certain at the beginning of the feasibility study that academics from different countries would agree on what to measure in the disciplines as well as on an assessment instrument, especially in a social science like economics. Therefore an economics strand was included in the feasibility study. Consultations and feedback collected indicated that it was easier than expected to get

© OECD 2012

economics experts to agree on what an AHELO should cover and measure. This was due, in part, to the decision to assess above content-knowledge and focus on the application and use of the concepts and "language of economics".

Developing a framework which is conceptually robust, technically referenced and capable of efficient operationalisation was crucial for the success of the Economics strand of the feasibility study. It is important that key stakeholders see their policy and educational concerns and interests reflected in the framework. Numerous cycles of consultation were built into framework development. This feedback played an important role in identifying the framework's strengths and limitations. These somewhat summative consultations were also used to validate plans regarding the framework's operationalisation.

Not all components of the five learning objectives were assessed in the Economics Assessment, in part due to the time constraint and the nature of a computer delivered assessment. For example, communicating with non-experts orally could not be assessed. Other learning objectives were applicable to CRTs rather than MCQs. In any case, the interpretation of the results from the administration of the instruments needs to reflect clearly what has been or not tested.

Overall, the Economics Assessment Framework took account of curricula diversity, defined the domain being measured, offered a clear conceptual structuring of the domain, and provided details of how the domain is to be operationalised for measurement purposes. The Economics Assessment Framework was endorsed by the Economics Expert Group.

Development of the Economics instrument

The instrument development process used for the Economics strand involved the following activities:

- Identifying and revising relevant items from existing instruments used for similar populations to those targeted in the feasibility study.
- Mapping all potential items against the provisional framework to determine if the items have sufficient cultural generality for inclusion in the feasibility study.
- Working with the Economics Expert Group so that items that contain field-specific terminology (such as "demand function" and "utility value") be translated accurately to be clearly understood and valid for students in different countries.
- Working with the Economics Expert Group to develop a "mini instrument", consisting of approximately 20 MCQs and one CRT, which was not meant to cover the entire domain but constructed in such a way that all the tasks included are considered appropriate to measure learning outcomes in economics.
- Linking all the resulting items to the developed provisional assessment framework.

The allotted administration time for the Economics assessment instrument was 90 minutes. It was intended that the assessment instrument include a broad sample of items covering levels

© OECD 2012

of difficulty that enabled the strengths and weaknesses of populations and key subgroups to be determined with respect to the components of economics competency.

Two types of assessment items were developed.

- The first type of item was a constructed-response format.

  Constructed-response tasks were developed through an evidence-centred design process to assess higher order integrative skills as well as communicative competencies. They assessed all of the five learning outcomes identified in the framework. Each CRT was designed to take students 30 minutes to complete, with the assessment instrument including one of these tasks.

- The second type of item had a multiple-choice format.

  These were designed to assess the same learning outcomes as the CRT but in a different manner. They were included to provide a fast and efficient way to collect data on students' economics knowledge, understanding and skills and to complement the CRTs in terms of domain coverage. To respond to MCQs, students needed to select one correct response out of four possibilities. In total, 50 MCQs were designed for students to complete in 60 minutes.

As CRTs measuring higher order skills need to be scored reliably across cultural and linguistic groups, careful consideration was given to developing scoring rubrics for such items. A key aspect in the evidence-centred design process, used to develop the CRTs, began with a notion of what successful performance on a task might look like. Scoring guides were developed in English while the tasks were being developed.

Test developers and the Economics Expert Group considered how best to weigh the mix of item types given the testing time constraints and the desired technical characteristics of the assessment instrument. They considered that a 90-minute long assessment instrument would ensure appropriate levels of reliability and sufficient content coverage. Although the instrument configuration with the highest estimated reliability included only MCQs, such an approach would not allow for the kind of content and skills coverage required by the framework.

To validate the items used for the Economics assessment instrument, the following activities were conducted:

- Verifying that all items reflected the consensus view of a group of well-regarded international experts to ensure appropriate content representation as well as diverse cultural and national perspectives, experiences and priorities.
- Conducting and recording a small number of cognitive labs and "think aloud" interviews with United States university seniors, using drafts of English versions of the items.

© OECD 2012

- Holding focus groups using a "mini test"' with small groups of students in a convenience sample of HEIs identified by NPMs and members of the Economics Expert Group in each of the target languages. Students were asked to complete the items and then asked to complete a small number of evaluative questions about each item.
- Compiling a summary of validity and usability information for each item in a final report.

Focus groups were held to assist in developing the assessment instrument in Phase 1. The feedback from the focus groups was used to revise the assessment instrument for testing of students in Phase 2. The focus groups also gathered data on candidates' perceptions of the assessment instrument.

Focus groups on the Economics assessment took place in all of the countries participating in the Economics strand of the feasibility study. In each country, focus groups were conducted at between 5 and 10 HEIs with volunteer participants. During the focus groups, participants worked on a printed version of the assessment instrument which includes a number of tasks. Based on the data collected from participants, the assessment instrument was reviewed by the translator and assessment developers. During the focus groups, participants were presented with either Form A or Form B of the assessment instrument. Each form consisted of one long CRT with multiple sub-tasks and twenty-five MCQs to be completed in one hour. One-half of the students in each focus group were given Form A, and the other half Form B.

Focus groups have contributed to identifying the following issues as they relate to CRTs:

- Given the need for CRTs to assess the above-content learning of final-year economics students worldwide, problems related to cultural specificities were written in such a way so as to not disadvantage students due to cultural factors.
- Due to variations in economics programmes in different countries, it was important to identify areas of essential commonality in different programmes so as not to favour certain candidates over others. The emphasis on above-content assessment, however, means that the CRTs were developed with a view towards assessing a student's economics competencies and capacities rather than strictly assessing content knowledge.
- As the CRTs present authentic economics contexts, copyright issues could pose a problem. It was important to ensure that materials used in the CRTs could be used without restrictions, and in some cases, to gain permission to use data or other published information, especially when these derived from official documents or reports.

Focus groups also contributed to identifying the following issues related to MCQs:

- Due to the nature of economics and the interdependency of content, some items could potentially map to multiple framework competencies. Every item, however,

© OECD 2012

was drafted to measure primarily a single component of economic knowledge and skill as specified in the framework.

- Economic principles are often divided into two major components: macroeconomics and microeconomics. Careful consideration was given to balancing items so that neither one of these components was emphasised over the other.

After revision and review, the final design of the instrument used for the Economics strand contained two CRTs and 48 MCQs. The MCQs were divided into four modules, each with 12 items and students were presented with all of these in different rotations. Each student taking the Economics assessment instrument was presented with one of two CRTs, and was assigned randomly a test that contained all 48 MCQs with one of four rotations.

The Economics assessment instrument was administered between February 2012 and June 2012 in 7 of the 17 AHELO feasibility study participating countries. A total of 60 HEIs participated in this strand, involving just over 6 000 students.

The AHELO feasibility study Engineering Assessment

Development of the framework and assessment materials for the Engineering strand was undertaken by the AHELO Consortium – specifically ACER, Japan's National Institute for Educational Policy Research (NIER), and the University of Florence. Several international consultants contributed to the development. Framework and assessment instrument development were overseen by an international Engineering Expert Group.

The Engineering Assessment Framework[9] (AHELO Consortium, 2011b) defines the domain to be tested and specifies the expected learning outcomes for students in the target group. The Engineering assessment instrument comprises both CRTs and MCQs, and is designed to be completed in 90 minutes.

The Engineering assessment instrument assesses the skills and knowledge of final-year bachelor's degree students and comprises test units developed around a key engineering problem coverage (see Annex B for illustrative items used in the AHELO feasibility study Engineering assessment instrument). The test units include a range of MCQs and CRTs. The items were translated and adapted by participating countries and qualitative validation was completed with focus groups and small numbers of students in a range of HEIs within each country. Analysis of the qualitative validation results guided the revision and the further development of the instrument.

Development of the Engineering assessment framework

The Engineering Assessment Framework (AHELO Consortium, 2011b) was the guiding document during instrument development. Materials were developed in direct consultation with the key competencies, which are set out in more detail in the framework itself.

The Engineering Assessment Framework was built on the AHELO-Tuning document (Tuning Association, 2009b), the AHELO Engineering assessment workshop held at ACER in Melbourne in January 2010, the Tertiary Engineering Capability Assessment (TECA) document (Coates and

© OECD 2012

Radloff, 2008), and broader AHELO technical materials. It draws on the processes and practices adopted in the PISA literacy surveys (e.g. OECD, 2005). Subsequent drafts incorporated review comments from Consortium and Engineering Expert Group members.

The Engineering Assessment Framework defines the domain to be tested, specifically: final-year Bachelor's degree students competency is the demonstrated capacity to solve problems by applying basic engineering and scientific principles, engineering processes and generic skills. It includes the willingness to examine such problems in order to improve the quality of life, address social needs, and improve the competitiveness and commercial success of society.

An assessment instrument must tap into the different aspects of a test taker's proficiencies. Engineering competency entails applying relevant skills and knowledge to solve problems of interest to an engineer. Recognising that engineering problems occur in a wide range of situations, a representative sample of contexts was selected to elicit the constituent components of engineering competency. The contexts in which students need to demonstrate their skills and knowledge included both those specific to civil engineering and those more generally applicable across a number of fields of engineering.

From a technical perspective, the Engineering Assessment Framework (AHELO Consortium, 2011b) developed for the AHELO feasibility study fulfils the requirement of a well-designed framework. It defines the domain to be tested and specifies the expected learning outcomes for students in the target population. The framework was endorsed by the Engineering Expert Group reflecting a consensus on the important learning outcomes. The framework took account of curricula developments, defined the domain being measured, offered a clear conceptual structuring of the domain. The framework also provided details of how the domain was to be operationalised for measurement purposes with discussions of issues such as time, language level, item type, scoring, assessment delivery and administration, and reporting.

However, particular issues of relevance to the review and final development of the Engineering Assessment Framework that were identified include:

- The differences among participating countries and HEIs in their use of concepts relevant to the AHELO feasibility study definition of an engineering programme assessment, and the implications of these differences to operationalise the framework through item development;
- The relationship between basic engineering and branch-specific competencies, including:
  - the extent to which a common set of basic competencies applies across all branches and what role that plays in assessing students in their final year of study in civil engineering; and
  - how basic competencies can be assumed: as attained by candidates much earlier in their study and not assessed; as likely to have been met and assessed by a very small proportion of items; or as assessed through a relatively large proportion of achievement items.

© OECD 2012

- The structure of representing competencies across the different branches, including:
  - whether the structure of knowledge and understanding of outcomes suggested in the AHELO Tuning framework is the best way to represent the outcomes; and
  - whether different scores (or groups of scores) should be weighted equally within each branch, and whether they are expected to exist as different scales within branches or contribute to a single scale.
- The role of generic engineering skills, including:
  - how the generic engineering skills described in the AHELO feasibility study Tuning framework relate to other (non-engineering) generic skills measured in the AHELO feasibility study[10]; and
  - what proportion of the assessment should be allocated to the assessment of generic engineering skills.

Development of the Engineering instrument

Following the development of the provisional assessment framework for the Engineering strand, an assessment instrument was developed to test one branch of engineering: civil engineering. Instrument development involved:

- identifying relevant items from existing instruments used for similar populations to those targeted in the AHELO feasibility study;
- developing new items for the instrument that will cover different item types;
- developing scoring guides in consultation with the Engineering Expert Group and NPMs to ensure that CRTs measure higher order skills and can be scored reliably across cultural and linguistic groups; and
- mapping all potential items against the established framework, in consultation with the Engineering Expert Group.

The Engineering assessment was expected to be completed by students in 90 minutes. The intent was for the assessment instrument to include a broad sample of items covering various levels of difficulty that enabled the strengths and weaknesses of populations and key subgroups to be determined with respect to the components of engineering competency.

Two types of assessment items were developed. The first type of item was a constructed-response format. CRTs were designed to comprehensively assess four key competencies defined in the Engineering Assessment Framework: i) engineering generic skills; ii) engineering analysis; iii) engineering design; and iv) engineering practice.

Constructed-response items typically required students to do more than provide a simple numerical or short-text response to a question. In the engineering context these items may, for example, have required students to complete short engineering designs (typically in their specialty branch), describe analytic processes or evaluate and make use of complex data to make recommendations or suggest solutions to engineering problems. The CRTs introduced an

© OECD 2012

authentic engineering scenario structure, design, situation or problem in a specific context and presented students with a set of items related to that context. The scenario was introduced through a range of stimuli including photographs, diagrams, tables and charts. Students then responded to a number of items requiring both short answers and longer responses.

The second type of item was a multiple-choice format. MCQs were designed to assess the fifth key competency defined in the Engineering Assessment Framework – basic engineering science. They were included to collect data quickly and efficiently on students' engineering knowledge, understanding and skills, as well as to complement the CRTs. They cover a wide range of basic engineering knowledge, along with specific above-content competencies. The MCQs were also included as a means of verifying the robustness of competencies assessed by the CRTs. Applying basic engineering and scientific principles requires proficient understanding, results from MCQs should indicate whether students have in fact developed the fundamentals that underlie competencies required to analyse and synthesise solutions to complex engineering problems.

Students needed to select one correct response out of four possibilities for the MCQs. In total, 40 multiple-choice items were developed, grouped in 4 sets of 10.

The item development team made use, wherever possible, of test items and source material submitted by participating countries. Including materials from participating countries helped ensure that, from a cognitive and conceptual point of view, test items reflected diverse modes of thought, and various cultural and national perspectives, experiences and priorities, particularly with a view to considering the possibility of developing an instrument that could be used in a broad range of countries in the future.

Once the assessment instrument was developed, it was submitted to the process of 'panelling'. The purpose of the panelling exercise was to ensure that items performed their intended function and were unambiguous.

The following questions were used during panelling sessions for the review of the assessment materials:

a) Content validity
   i. How does the material relate to the assessment specifications?
   ii. Do the questions test the assessment framework?
   iii. Do the questions relate to the essence of the stimulus or do they focus on trivial side issues?
   iv. How will this material stand up to public scrutiny (including project stakeholders and the wider community)?

b) Clarity and context
   i. Is it coherent? Unambiguous? Clear?
   ii. Is the material interesting? Is it worthwhile? Of some importance?
   iii. Is it self-contained? Or does it assume other prior knowledge, and if so is this appropriate?
   iv. Is the reading load as low as possible?

© OECD 2012

v. Are there any 'tricks' in the question that should be removed?
vi. When a unit includes more than one item, are there dependencies between the items? For example: Does one item give a clue to the next one? Would a different order of items within a unit make a difference? If a response to one item is incorrect, does this affect possible responses for other items in the unit?

c) Format

i. Is the proposed format the most suitable one for the item?
ii. Is the key (the correct answer to a multiple-choice item) indisputably correct?
iii. Are the distracters (the incorrect options to a multiple choice question) plausible but indisputably incorrect?

d) Test takers

i. How will the test-takers perceive this material? To answer this, panel members must imagine the cognitive, cultural and response format demands of the items from the point of view of test-takers.
ii. Is it at the right level, both in terms of the expected ability level, age or school year-level(s) of the test-takers?
iii. Does the material breach any ethical, cultural or other sensibilities?
iv. Is it likely to be biased, i.e. is it likely to be easier or harder for certain subgroups in the assessment population for reasons other than differences in the ability being measured?
v. Is it clear what would constitute an answer to the question? That is, will test-takers know exactly what they are being asked to produce (as distinguished from knowing how)?

e) Scoring

i. Is the proposed scoring consistent with the underlying ability described by the assessment domain? Would test-takers possessing more of the underlying ability always score better than test-takers with less?
ii. Is the chosen marking scheme the most suitable one for the purpose?
iii. Are there other answers that have not been taken into account in the marking guide (e.g. those that may not fall within the 'correct' answer category description, but seem to be correct)?
iv. Are there different approaches to arriving at the same answer? Do these different approaches represent equivalent or different levels of proficiency?
v. Should partial credit be given if part of the answer is achieved?
vi. Are the scoring criteria practicable for the markers, or are they excessively cumbersome? Are the different levels of performance clearly distinguishable?

© OECD 2012

After panelling, the items were modified in response to the panel recommendations and finally reviewed by the development team and Engineering Expert Group.

To establish the validity of items and decide on a suitable configuration of items that best meets the over-arching goals of AHELO, the following activities were conducted:

- Mapping items to the conceptual framework.
- Ensuring that items reflect diverse modes of thought, and various cultural and national perspectives, experiences and priorities.
- Conducting a small number of cognitive labs and 'think aloud' interviews with Australian university seniors, using drafts of English versions of the items.
- Pre-testing the 'mini test' in-country with small groups of students (focus groups) in a convenience sample of HEIs identified by NPMs and by the members of the Engineering Expert Group in each of the target languages. Students were asked to complete the items and then asked to complete a small number of evaluative questions about each item.
- Compiling a summary of validity and usability information for each item in a final report.

Four focus groups were conducted with final year undergraduate civil engineering students. In each of the four focus groups, students were asked to complete three CRTs within 60 minutes. A further 60 minutes was then dedicated to discussion and feedback. This process provided a chance to obtain some initial data from current students.

Following data collection from focus groups, the scoring guides were reviewed in light of authentic student responses in the different languages. This review allowed the test developers to:

- check and when necessary refine the descriptions of student achievement described in the scoring guides in light of actual student responses;
- refine the guides to accommodate any previously unanticipated valid responses; and
- supplement the guides with student response examples that are indicative of the different substantive categories described in the guides.

Feedback from the focus groups suggested that the authentic scenario tasks that were developed stimulated students' interest in the tasks (see Boxes 4.1 and 4.2). The aim of the focus groups was to gauge the way in which students actually engaged with the assessment instrument and thus to inform assessment development. In addition to providing useful feedback during discussion, each participating student was asked to fill in a questionnaire which asked certain questions regarding the items.

The students participating in these focus groups provided an overwhelmingly positive response in the feedback questionnaire and during the group discussions. They found the CRTs to be authentic, practical, clear, interesting and engaging.

© OECD 2012

Focus groups also contributed to identifying the following issues related to CRTs:

- Given the need for CRTs to assess the above-content learning of final-year civil engineering students worldwide, problems related to cultural specificities were written in such a way that students from certain countries cannot be disadvantaged due to cultural factors.
- Due to variations in civil engineering programmes in different countries, it was important to identify areas of essential commonality in different programmes so as not to favour certain candidates over others. The emphasis on above-content assessment, however, means that the CRTs were developed with a view towards assessing a student's engineering competencies and capacities, rather than their content knowledge.
- There was concern that students in some countries would find certain contexts difficult due to familiarity issues, and that this would affect their results. Allocated testing time was another issue. After the focus groups and expert feedback, it was deemed that 20 minutes was not sufficient for a student to complete each CRT. Many students struggled to complete the CRTs in the designated time during the initial focus groups. It was decided to increase the time allocated to 30 minutes per CRT.
- As the CRTs present authentic engineering contexts, copyright issues could pose a problem. It was important to ensure that materials used in the CRTs could be used without restrictions. It might be necessary to gain permission to use data and images, especially when these are derived from official documents or reports.

Focus groups also contributed to identifying the following issues related to MCQs:

- A large proportion of the MCQs selected for inclusion in the clusters needed to be revised so that they were acceptable internationally. It was important to maintain acultural contexts and content in the MCQs, as with the CRTs. If the content of an MCQ was deemed to be country-specific, it was removed from the pool of items. For instance, one item assessed student's knowledge of the specific engineering codes in Japan. An item such as this is not within the scope of the AHELO project.
- There was a wide range of difficulties in the MCQs for possible inclusion. It was decided to spread the difficulties across the clusters. Some items were deemed to be too difficult, however, as they did not focus on basic engineering sciences but contained a more specialised focus for a specific type of engineering student.
- Furthermore, some items would have taken a student too much time to complete, which was not practical due to the time constraints of the AHELO assessment.
- Several MCQs which focused on non-technical aspects of civil engineering were also removed from the pool for two reasons. First, these items mapped to areas of the framework that were covered by the CRTs. Second, these items were rather easy and it was deemed that any individual with high-level critical thinking skills would have selected the correct answer.

© OECD 2012

Box 4.1 - Australia - Students' feedback on the AHELO assessment instruments

In the first phase of the AHELO feasibility study, ten Australian universities conducted focus groups with final year civil engineering students to provide feedback on the assessment items that were drafted for the project. There were two types of items examined by students: a constructed response task (which requires students to respond to a range of stimuli in order to present solutions to authentic problems) and multiple choice questions (relating to basic engineering science). Students were asked to complete the items, individually fill out a survey detailing their opinions about the items and then engage in a discussion with their fellow students about the assessment.

Overall students from Australia who participated in the focus groups saw the assessments as having content relevant to their coursework and being in a format that was new and challenging.

Typical feedback from students relating to the constructed response task included:

> "It's a realistic problem which made me think and understand that the knowledge I learned from university is being applied in the real world";
>
> "Interesting question which challenges people to think. A real situation for real application was interesting"; and
>
> "It is a real project - I may meet the same problem in my future career. It was challenging for me".

For the multiple choice questions feedback included:

> "It was a very comprehensive summary of most things related to civil and structural engineering. Personally good to revise";
>
> "Interesting and challenging questions. Have to think critically and apply the skills learnt in past four years. Very relevant to my program"; and
>
> "The task covered a broad range of knowledge".

NPM for Australia

The Engineering assessment instrument was administered between February 2012 and June 2012 in 9 of the 17 AHELO feasibility study participating countries. A total of 70 HEIs participated in this strand, involving just over 6 000 students.

The AHELO feasibility study Contextual Surveys

The contextual framework and survey instruments were designed by experts from the Centre for Higher Education Policy Studies (CHEPS) in the Netherlands and the Center for Postsecondary Research (CPR) in the United States, based on work carried out by the OECD, the AHELO Group of National Experts and consultants.

The development of the framework was undertaken through research and consultation, and by seeking the expert opinion of various groups and individuals from all over the world. Feedback

© OECD 2012

from consultations was used to finalise the framework and provided a basis for instrument development, validation, small-scale testing and delivery.

Three context survey instruments were developed to identify factors that could shed light on observed learning outcomes of the target population: i) a Student Context Instrument (SCI); ii) a Faculty Context Instrument (FCI); and iii) an Institution Context Instrument (ICI). In addition, various indicators were specified to be collected at the national level in the National Context Instrument (NCI) to provide additional context data.

Development of the Contextual dimension framework

The Contextual Dimension Assessment Framework (AHELO Consortium, 2011c) was developed and validated which, for the purposes of the AHELO feasibility study, reflected an international consensus on the important contexts that shape higher education learning outcomes. The development work built on work already completed by the contracted contextual dimension experts (Ewell et al., 2009).

In broad terms, the framework for the context survey instruments set the conceptual focus and scope for this aspect of the work, provided the technical foundations for managing collection, analysis and reporting, and enhanced the efficiency of the assessment.

As with the frameworks developed for the assessment instruments, a number of guidelines and processes were observed in developing the framework for the context surveys. The framework was designed with the following principles:

- To reflect the views of the participating countries as well as the thinking of leading experts in the field. Input from countries, AHELO's Technical Advisory Group (TAG), the Economics and Engineering Expert Groups and AHELO's Stakeholder Consultative Group (SCG) has been critical. Individuals and groups have worked together with the design team to reach consensus on how best to measure the constructs.
- To take into account the characteristics of the target population. The AHELO feasibility study is being conducted in a wide range of countries. Countries vary culturally, in how higher education fits into society and in many other ways. These country and cultural differences have been taken into account by the design team, countries and experts.
- To define the constructs to be measured in the contextual surveys, help operationalise the constructs, specify item characteristics and design parameters, and provide a template for mapping context items.
- To reflect the relative importance of constructs and scales so that question inclusion can be decided and communicated to the various stakeholders.

The development work on the contextual dimension began with a series of background reviews. This integrated series of reviews detailed relevant substantive, pragmatic and technical considerations relevant to development.

© OECD 2012

Using existing work as a guide, an extensive review of relevant research was conducted. Methodological papers were considered so as to highlight potentially important technical characteristics of the surveys. This research helped further establish the relevance of the concepts to be addressed, to chart technical characteristics of surveys in use, and to clarify and position the development approach. The review was international in scope.

Simultaneously, an audit of existing context survey resources and practices was conducted. This focused on initiatives that are international, national and institutional in nature. These three levels of analysis are important as many contemporary collections are trans-national rather than national in focus, and many of the largest collections are cross-institutional in nature.

A database of relevant and available survey items was compiled. These were categorised according to pertinent substantive, technical and practical criteria. An attempt was also made to source psychometric information on selected items. The item inventory was intended to be a working document that would facilitate the item development process.

The review considered what existing documentation was available in participating countries and HEIs with a view to minimising data collection, and reducing the burden on participating countries. This required reviewing the confidentiality, currency and reliability of existing information. This audit was undertaken via NPMs, in parallel to the research review, and through direct liaison with selected HEIs and research institutes.

The results of the reviews were analysed and condensed into a synthesis of existing research and practice. This was based directly on existing work undertaken by both the U-Map project (Van Vught, 2009) and Ewell et al. (2008, 2009), providing validation and extension where necessary. Documentation of the background reviews was an intrinsically important activity, but the main operational purpose of doing this was to inform the formation of a parsimonious but sufficiently rich framework.

The framework helped maximise collection efficiency while at the same time ensuring conceptual alignment and technical rigor. A single high-level framework was developed to situate the constructs and variables to be measured, assist with item generation and management, and facilitate statistical analysis and reporting.

The structure proposed by Ewell et al. (2009) aligned well with the hierarchical input-process-output framework. Over the years, this basic but robust framework has been used across a large number of diverse education systems and contexts (Bottani and Tuijnman, 1994; Astin, 1978 and 1985; Ewell and Jones, 1996; Jaeger, 1978; Nuttall, 1994). The most established contemporary international manifestation is the OECD's indicators of education systems (INES) framework (OECD, 2004). To build links with other analytical and empirical work, the INES framework used the general organising principle. The INES framework organises policy and educational issues to be considered in the AHELO feasibility study using two dimensions:

- the level of the country higher education to which the resulting indicators relate (systemic, provider, instructor, learner); and
- whether they relate to outputs or outcomes, policy-amenable determinants of these outcomes, or antecedents or constraints.

© OECD 2012

This structure was populated with variables nominated by Ewell et al. (2009), the GNE and additional background reviews. As illustrated by Ewell et al. (2008), a wide variety of context variables could be addressed by the AHELO feasibility study. Given the need for parsimony, the following priorities were used to guide the selection:

- Is the construct of policy relevance and interest? Is the issue of timely interest? Is there a public interest in the particular issue? Will the information obtained be of value in understanding student performance and other outcomes and taking steps to improve them? Is the issue related to a construct in which trends over time are of importance?
- Is the factor appropriate for international assessment? Is the issue one that can be addressed cross-culturally? That is, will it have a similar meaning across countries? Will comparisons relate to important policy issues? Is addressing the issue in the AHELO feasibility study adding value beyond what might be accomplished with an institutional evaluation?
- Within the context of the AHELO feasibility study design, is it technically feasible to address the issue? Are the variables used to address the issues reliable and valid? Do we know that from prior variable usage, or from the research literature? Is the time and cost associated with the variables used to address the issue reasonable? How do we evaluate whether items are working properly? Are there less expensive ways to collect the information?

The INES structure provided a sufficiently broad foundation locating the concepts of interest to the AHELO feasibility study contextual dimension. Following the process detailed in the 2003 PISA Technical Report (OECD, 2005) and the European Commission sponsored U-Map project (Van Vught et al., 2008; Van Vught, 2008 and 2009), additional dimensions were added to this framework to both reflect the multidimensional nature of contemporary conceptions of higher education quality and enable specification of substantive concerns at a relevant level of detail. The resulting framework was designed to identify specific links between various factors and design analyses that explore the direct or indirect relationships between these.

The Contextual Dimension Assessment Framework specifies the rationales for developing surveys as part of a contextual dimension of the AHELO feasibility study:

- First, to help assess the feasibility of capturing contextual variables across different countries and institutional contexts in valid and reliable ways.
- Second, to help test psychometric analyses to identify relevant contextual variables for longer-term development and demonstrate the analytical potential of the AHELO feasibility study for institutional improvement.
- Third, to help manage sampling, data collection and quality control and, finally, to inform bias analyses undertaken as part of the objective assessments.

From a technical perspective, the Contextual Dimension Assessment Framework fulfils the requirement of a well-designed framework. The framework situates the constructs and

© OECD 2012

variables to be measured, assists with item generation and management, and facilitates statistical analysis and reporting. Broad consultation took place to seek the expert opinion of a range of groups and individuals (notably the AHELO GNE, TAG, SCG, Economics Expert Group, Engineering Expert Group, and the AHELO Consortium). NPMs and members of the SCG were able to provide comments on specific indicators via an online survey. The resulting framework is reflective of an international consensus about important learning processes and outcomes.

Development of the Contextual dimension surveys

With the framework in place, three surveys were developed to be conceptually and psychometrically linked: i) an SCI; ii) an FCI; and iii) an ICI. In addition, a range of indicators was specified for collection at the national level to provide additional context.

Additional information at the institutional level was also required to assist with management, sampling, fieldwork, analysis and reporting. NPMs and Institution Coordinators (ICs) also collected the information that was absolutely necessary, from a technical and practical perspective, to implement the feasibility study.

The approach used to develop and validate the context items align with those used for the assessment instruments. Its defining features were that it would be highly consultative, phased and conceptually based. In recent years, the approach has been tested and refined in several national and international studies. It is detailed here in terms of defining design specifications, mapping and drafting items, qualitative testing, quantitative testing, and final review. The same approach was used for all instruments regardless of variation in the target population.

Survey instrument development was guided by a number of general design considerations to enhance the power of measurement and ease of administration. These aligned with the standards set for international data collections (Schmidt and Cogan, 1996; Mullis, Martin and Stemler, 1999; Siniscalco and Auriat, 2005), characteristics of large-scale existing context assessments (for instance: NSSE, 2008; AUSSE, 2009) and linked with other survey design specifications recorded during the background reviews.

In summary, the surveys developed for the contextual dimension, were designed to:

- measure the target constructs;
- have high levels of face, content and construct validity;
- provide reliable and precise measurement of target constructs;
- be efficient to administer, analyse and report;
- align with and enhance existing instruments and practices; and
- provide a basis for ongoing research and development.

Item specifications were set to enhance the quality of measurement and minimise response interference effects. Item specifications were defined from studies that have sought to determine the characteristics of effective survey items (Kuh, 2002; Laing et al., 1988;

© OECD 2012

Yamaguchi, 1997; Converse and Presser, 1989; Andrich, 1993; Bradburn and Sudman, 1988). Such specifications included, for instance, that:

- items should be as succinct as possible;
- grammar and spelling should be accurate;
- the cognitive demand of each item should be less than the cognitive demand of the phenomenon being measured by that item;
- only a single phenomenon should be tested in each item;
- each item should have distinct and independent content;
- items should be distributed across the range of the variable;
- use of negative phrasing should be minimised or avoided;
- the information should be known to the respondent;
- items should refer to recent activities;
- items should be reviewed for possible bias to environmental or individual contexts;
- items should be phrased clearly and unambiguously;
- item design should encourage serious and thoughtful response; and
- items should not threaten, embarrass or violate the privacy of the respondent or encourage them to respond in socially desirable ways.

The surveys were designed to be completed within 15 minutes.

The work to operationalise the multidimensional framework for the purposes of measurement involved identifying the desired level and unit of analysis for each construct in the framework, and then mapping the constructs to the different surveys, potentially in a many-to-many fashion. This exercise linked the framework with the instruments and charted the data structures and algorithms required for item sampling.

On this basis, initial item drafting and review began. For this, the item inventory compiled earlier which contained the most advanced items in use in relevant existing instruments. Every effort was made to link with widely used available resources, including through establishing licensing arrangements that may be required.

The items were drafted with reference to the underpinning assessment framework, a process that was managed continuously throughout the development. This mapping process provided information about how each survey item fit in the framework and how the combination of survey items in the surveys and existing resources achieved the overall measurement goals described in the framework.

Items were developed for use in online versions. The survey items and materials were first developed in English. The translation of the surveys into the languages used in participating

© OECD 2012

countries was the responsibility of participating countries. However, central management of the surveys and their development was needed to ensure that cross-national comparability of instruments is achieved in the translation process.

Feedback on the draft surveys was sought from a range of stakeholders. This feedback helped refine and shape items to align them with the assessment framework, educational contexts, existing quality assurance activities and research foundations.

It was essential that appropriate language be used in the instruments. An appropriate reading level was set, as common semantic conventions.

A series of focus groups were held to capture student and faculty insights into the range and characteristics of the items. The small-scale trial tests collected data to undertake an initial psychometric examination of the context survey items, and provided further information that would help refine items and scales.

The survey instruments were distributed to a heterogeneous sample of current learners. The faculty-focused instruments were distributed to a sample of current faculty with teaching activities. For efficiency, these tests were conducted online. Data was entered, verified and compiled into files for analysis. Around 50 staff and 200 students were asked to take part in this field test of the student and faculty survey instruments.

At the same time, NPMs were allowed to remove any items from localised versions which they felt were entirely inappropriate for the local context, or add items which had been approved by the AHELO Consortium. This occurred principally in the instruments for Institutional Coordinators.

The Expert Groups from all strands provided input into the process of item development for the survey instruments. This included both thinking about how key reporting issues can best be addressed, and what particular wording in the questions was most likely to produce sound items with good psychometric qualities, i.e. valid and reliable.

A range of psychometric analyses were conducted to explore the characteristics of learners' interactions with the student survey items, the behaviour of the items, and relationships between items and target constructs. Technical reviews and developments were undertaken to bring together the various validation activities, cross check the measurement properties of the instruments, and develop a range of resources for managing and analysing the items and instrument.

The final set of items was proofed and cross-checked against project objectives and survey instrument specifications. The items were reviewed in terms of the generic measurement criteria specified at the start of the development. The item mapping initiated at the start and managed through the development process was verified.

A codebook was developed to manage the operationalisation of the items, assist with item sampling, underpin data file production, and guide analysis and reporting.

The contextual survey instruments were administered between February 2012 and June 2012 in all of the 17 AHELO feasibility study participating countries. Contextual data was collected

© OECD 2012

from 22 977 students, 4 807 faculty in 248 HEIs, providing critical insights into environmental and demographic factors relevant to learning outcomes.

Box 4.2 - Mexico - Students' feedback on the AHELO assessment instruments

Students participating in AHELO showed their appreciation for this test. They answered a questionnaire related to the contextual dimension; additionally, various HEIs conducted informal motivational surveys to learn about the students' motivation and performance on the test.

For the Generic Skills strand, most students answered that they were unfamiliar with constructed-response tests, but they considered them engaging and challenging. For some of them, this kind of test is quite unusual and, they would like them to be used during the learning process. They also wanted to know the results of their performance on the test.

For the Engineering strand, students expressed that the test was interesting, especially the open questions section in which they applied their knowledge to solve problems. Some of them observed that the test included some topics that they did not study in their course curriculum but they believed that their knowledge helped them to solve it anyway.

For the Economic strand, surveys were conducted prior to the test during the focus groups. In these groups, students said they liked the test although they would rather have something more qualitative. They noticed that many items were about knowledge from the first semesters so they could not recall them easily.

NPMs for Mexico

Localisation of assessment and survey instruments

Two major aspects needed to be assessed when considering cross-cultural and cross-linguistic generalisability of instruments: i) how well participants met the established technical standards for detailed translation, adaptation and verification (localisation) processes; and ii) psychometric analyses results providing evidence of data comparability across contexts[11].

The section below focuses on how the procedures used for localisation of assessment and survey instruments were undertaken in the study. The following steps were followed:

- A preliminary examination of test materials was conducted by a panel (test developers, verifiers, domain specialists) to anticipate potential adaptation issues, ambiguities, cultural issues or item translatability problems. The information gathered at the earliest stages was integrated into a centralised record of the translation and adaptation history of each item.
- The AHELO feasibility study Translation and Adaptation Guidelines (AHELO Consortium, 2011d) were developed to assist national translation teams. This document includes general as well as item-by-item guidelines for individual instruments and surveys. It provides a list of mandatory, desirable or rejected adaptations that should be reflected in the centralised monitoring tool for translation, adaptation and verification. The item-by-item guidelines draw the

© OECD 2012

translators' attention to possible terminology problems, translation traps, literal matches (e.g. between stimuli and items), patterns in response options, etc. The guidelines were produced jointly by test developers, domain experts and linguists.

- A file management system was used to coordinate file transfer between participating countries. Instructions on how to use the SoNET environment to develop national versions of the instruments and surveys were provided to translators.
- Double translation and reconciliation was used to translate and adapt instruments. Participating countries were asked to prepare two independent translations of test instruments and survey items into their target language(s). A reconciler merged the two independent translations and prepared a final national version. The translation and adaptation of the constructed-response items scoring rubrics followed a slightly different approach. It was suggested that countries produce a single translation of the coding guides and have this translation vetted by their local domain specialists. The translation team then checked for thorough equivalence of the translated scoring guides to ensure that scorers in each country shared the same understanding.

All source instruments were prepared in English. For each field implementation, a number of distinct instruments required translation, including constructed-response tasks, multiple-choice items and student, faculty and institution context surveys. The majority of countries tested students in one language, with the exception of Egypt and Kuwait which offered students the option to take the test in either Arabic or English (see Box 4.3)

Box 4.3 – Kuwait - Assessing students in a multicultural society

As an important multi-cultural center of the Gulf and the larger Middle East, the State of Kuwait operates on a multi-language platform in all sectors of society both publically and privately. Bi- and even tri-lingual competencies of Kuwaiti society are becoming the norm rather than the exception, which is made apparent by the range of international academic curricula that are operational from K-12, undergraduate as well as graduate education and beyond. The State of Kuwait thus chose to provide its educational institutions with the option of assessing its students' competencies in a language that best reflected its curriculum, its students' cultural and academic profiles, and in accordance with the pre-dominant academic language that is operational at each respective institution/college.

In the case of students majoring in Quranic studies who took the written test over and above the computerized platform: the National Committee believed that students in more traditional majors were less comfortable with the computerized technology and word processing when applied to the Arabic language. The committee believed that the assessment of student's generic skills should not be hampered by the student's inability or discomfort with the method of assessment and the platform used for assessment. The committee thus decided that providing those students with the option of the handwritten test eliminated a variable that could potentially discriminate students who are not sufficiently competent in Arabic word-processing.

NPM for Kuwait

© OECD 2012

The number of translation, adaptation and verification activities required for the feasibility study was substantial. In all 120 translation and adaptation activities took place. All of these processes were subsequently verified, resulting in 240 activities overall. Detailed information about the processes used to localise instruments used in AHELO can be found in the AHELO Country Profiles Translation and Adaptation Process (AHELO Consortium, 2012b).

In AHELO, as in many major international assessment studies, the localisation process followed mostly the decentralised model in which National Centres were each responsible for localising assessment materials for use in their respective systems while the AHELO Consortium guided and assisted the systems throughout the process. In specific cases, however, an alternative approach was used due to the lack of available time to implement the decentralised model. To accommodate the large volume of work and tight timelines, three types of workflows were used to implement localisation processes. These types are indicated by the letters A, B and C in Table 4.1.

© OECD 2012

Table 4.1 - Country localisation workflows

Country or system	**Language of testing**	Generic Skills MCQ	Generic Skills CRT	Economics MCQ	Economics CRT	Engineering MCQ	Engineering CRT	Student Contextual Instrument	Faculty Contextual Instrument	Contextual Institution (Generic Skills)	Contextual Institution (Economics)	Contextual Institution (Engineering)
Abu Dhabi	English (Abu Dhabi)					A	A	B	B			B
Australia	English (Australia)					A	A	B	B			B
Belgium (Fl.)	Dutch (Flanders)			A	A			B	B		B	
Canada (Ontario)	English (Canada)					A	A	B	B			B
Colombia	Spanish (Colombia)	B	C			A	A	B	B	B		B
Egypt	Arabic (Egypt)	B	C			A	A	B	B	B	B	B
Egypt	English (Egypt)	B	C	A	A	A	A					
Finland	Finnish	B	C					B	B	B		
Italy	Italian			A	A			B	B		B	
Japan	Japanese					A	A	B	B			B
Korea	Korean	B	C					B	B	B		
Kuwait	Arabic (Kuwait)	B	C					B	B	B		
Kuwait	English (Kuwait)	B	C					B	B	B		
Mexico	Spanish (Mexico)	B	C	A	A	A	A	B	B	B	B	B
Netherlands	Dutch			A	A			B	B		B	
Norway	Norwegian	B	C					B	B	B		
Russian Federation	Russian			A	A	A	A	B	B		B	B
Slovak Republic	Slovak	B	C	A	A	A	A	B	B	B	B	B
USA (CT, MO, PA)	English (USA)	B	C					B	B	B		

Workflows A and C were both decentralised translation models. Workflow A was managed by the AHELO Consortium and Workflow C by CAE. While differences between the two approaches existed, they had similar characteristics. First, centralised tools and documentation was

© OECD 2012

prepared for National Centres detailing how the adaptation and translation processes should take place. Second, National Centres received training in implementing the translation approaches.

A dual translation model was used in each country, with two translators working independently to provide a full translation of the relevant instrument into the testing language. Country teams then determined the best reconciled translation, taking account of the strengths of each version. This was then verified by the AHELO Consortium or an agency working with CAE. The localisation process provided an opportunity for some countries to cooperate on this activity (see Box 4.4).

Box 4.4 - Country cooperation on translation

Netherlands and Belgium (Flanders)

The Netherlands and Belgium (Flanders) both participated in the Economics strand. The Flemish team joined Phase 1 at a rather late stage. This created an unfortunate delay in their progress. This delay could only be handled thanks to a generous offer of cooperation from the Dutch team. Both countries use Dutch as their national language. They even have a linguistic union and this guarantees that the same grammatical and spelling rules are applied. In such a situation it was evident that a close cooperation in the translation of the questionnaires and handouts could be an advantage to both teams.

The Dutch team made their translation of the initial set of questions for the economics test available to the Flemish team. When this translation was submitted to a Flemish expert team to assess the extent to which it could be applied to their country, it turned out that there were considerable cultural differences between the two countries on how some of the questions were phrased. The wording of some questions, though perfectly understandable, sounded rather strange and artificial to Flemings. They needed to be adapted to avoid negative reaction from students. In some cases the translation of technical terms, though correct, was assessed as inadequate for use by Flemish students. These translations were adapted or supplemented with the original English terminology. The same procedure was applied to the final version of the economics test in Phase 2.

The advantage of such cooperation between countries using the same language is evident. The tasks in the translation procedure can be split between the teams and this lowers the work load. But they must be aware that, due to cultural and linguistic differences, a round of adapting the translated instruments to the national culture is a must.

NPM for Flanders

Colombia and Mexico

Colombia and Mexico were the only two Spanish speaking countries that participated in the study. Both countries´ leading teams made an agreement to share translations of different materials. These include guidelines, workshops and test materials of the two common strands they participated in: Generic Skills and Engineering. The possibility to cross check Spanish translations of test materials enabled a higher validity of the comparison of results between both systems.

NPM for Colombia

© OECD 2012

The purpose of verification was to ensure linguistic correctness and cross-system equivalence of the different language versions of AHELO instruments. For the Economics and Engineering strands, one of two verifiers was a domain specialist. This was the first time in an international study that a domain specialist had been involved in the verification process. This added element ensured that the translation of domain-specific terms was reviewed by an expert in the testing domain.

In countries which tested in English, translation was not required but adaptation was nevertheless used to ensure that the instrument was entirely appropriate to the country context in which it would be used. The changes made were then verified.

Workflow B was used for Generic Skills multiple-choice items and for the contextual surveys. Due to the late decision to include these elements in AHELO, there was insufficient time for Workflows A or C to be followed. In Workflow B, instruments were centrally translated by BranTra, part of the AHELO Consortium and then provided to National Centres for review and adaptation prior to verification.

National Centres appreciated the use of centralised translation mainly because the translation cost was borne by the AHELO Consortium, reducing the burden on country resources. At the same time, this approach was much faster than the decentralised approaches used in Workflows A and B. A drawback, however, was that National Centres felt a lesser sense of "ownership" over their system versions of the instruments.

Most NPMs provided positive **feedback** on the translation, adaptation and verification of test and context materials. The majority of NPMs found the process to be professional and straightforward, standards high and support and feedback from cApStAn (in relevant strands) excellent. The use of Adaptation Translation and Verification (ATAV)[12] spreadsheets was welcomed by NPMs, with one commenting that "ATAV is a system that gives all parties an overview of the translation process and it also makes it possible to 'argue' a point or a position".

Several countries did nevertheless offer suggestions on desired improvements for the future:

- In the Generic Skills strand, the translation process for the performance tasks materials was reported to be challenging and more time-consuming than first estimated by some countries. For instance, the texts were much longer than expected at the time of estimating costs in order to secure funding for ATAV activities.
- In the disciplinary strands, countries pointed out the importance of involving domain experts early in the translation process to support NPMs when translating the instruments. Those experts should be familiar with the relevant field in the country and could provide not only technical feedback, but also feedback on idiomatic expressions and wording.
- Several countries recommended improvements in the schedule of the item translation and adaptation process. Countries recommended that the translation and adaptation process should only begin once the test instrument is fully understood.

© OECD 2012

For instance, it was suggested to review test materials including questions, benchmark papers, and samples of students' answers before completing translation and adaptation activities.

- There was also feedback from one NPM who indicated that it would be beneficial to have "less focus on word to word equivalence, and more focus on substantive equivalence". One NPM in the Generic Skills strand, for example, asserted "to be sure we need [the] expertise of [a] translation and adaptation expert but it is not enough".
- With respect to training and ongoing support, several NPMs recommended follow-up support, for example, through an online chat.
- Finally, a number of NPMs recommended that, in any future AHELO main survey, the key manuals used by ICs, TAs and LSs undergo a formal ATAV process.

Overall, the localisation of the study assessment instruments and contextual surveys to be delivered to students, faculty and ICs in the language of instruction at participating institutions throughout seventeen countries involved a number of detailed, time consuming and complex activities.

The objective of localisation was to obtain the highest possible level of equivalence across countries as well as languages. As such, it was imperative to ensure that each assessment item tests the same skills and brings into play the same cognitive processes as the original version, while being culturally appropriate within the target country.

© OECD 2012

REFERENCES

AHELO Consortium (2012a), *Generic Skills Assessment Framework*

AHELO Consortium (2012b), *Country Profiles Translation and Adaptation Process*

AHELO Consortium (2011a), *Economics Assessment Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)19/ANN3/FINAL&doclanguage=en

AHELO Consortium (2011b), *Engineering Assessment Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)19/ANN5/FINAL&doclanguage=en

AHELO Consortium (2011c), *Contextual Dimension Framework*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=edu/imhe/ahelo/gne(2011)21/ann1/final&doclanguage=en

AHELO Consortium (2011d), *Translation and Adaptation Manual/Guidelines*

Andrich, D. L. (1993), "A Hyperbolic Cosine Latent Trait Model for Unfolding Dichotomous Single-simultaneous Responses", *Applied Psychological Measurement*, Vol. 17(3).

Astin, A. W. (1978), *Four Critical Years: Effects of College on Beliefs, Attitudes, and Knowledge*, Jossey-Bass, San Francisco.

Astin, A. (1985), *Achieving Educational Excellence: A Critical Analysis of Priorities and Practices in Higher Education*, Jossey-Bass, San Francisco.

Bottani, N. and A. Tuijnman (1994), "The Design of Indicator Systems", In A. C. Tuijnman and T.N. Postlethwaite (Eds.) (1994), *Monitoring the Standards of Education: Papers in Honour of John P. Keeves*, Pergamon, Oxford.

Bradburn, N. M. and S. Sudman (1988), *Polls and Surveys: Understanding What They Tell Us*, Jossey-Bass, San Francisco.

Clanchy, J. and B. Ballard (1995), "Generic Skills in the Context of Higher Education", *Higher Education Research and Development, Vol. 14(2), Routledge, London.*

Coates, H. and A. Radloff (2008), *Tertiary Engineering Capability Assessment: Concept Design*, Australian Council for Educational Research (ACER), Camberwell.

Converse, J. M. and S. Presser (1989), *Survey Questions: Handcrafting the Standardized Questionnaire (Quantitative Applications in the Social Sciences)*, Sage Publications Inc., Newbury Park.

Ewell P. T. et al. (2009), "Analytical Framework for the Contextual Dimension in the AHELO feasibility study", paper presented to the AHELO Group of National Experts, OECD, Paris, 27-28 April 2009.
http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=EDU/IMHE/AHELO/GNE(2009)9&doclanguage=en

© OECD 2012

Ewell, P. T. et al. (2008), *Adult Learning in Focus: National and State-by-State Data*, National Centre for Higher Education Management Systems (NCHEMS), Boulder.

Ewell, P. T. and D. P. Jones (1996), *Indicators of "Good Practice" in undergraduate education: A Handbook for Development and Implementation*, National Centre for Higher Education Management Systems (NCHEMS), Boulder.

Hambleton, R. K. et al. (2005), *Adapting Educational and Psychological Tests for Cross-cultural Assessment*, Lawrence Erlbaum Associates Publishers, Mahwah.

Jaeger, R. (1978), "About Educational Indicators", *Review of Research in Education*, Vol. 6, pp. 276-315. Wadsworth Publishing, Belmont.

Kuh, G. D. (2002), *The College Student Report (4th ed)*, National Survey of Student Engagement, Indiana University, Bloomington.

Laing, J., R. Sawyer, and J. Noble (1988), "Accuracy of Self-Reported Activities and Accomplishments of College-bound Students", *Journal of College Student Development*, Vol. 29(4), Alexandria.

Harkness, J. A. et al. (2003), "Questionnaire Design in Comparative Research", in *Cross-cultural Survey Methods*, John Wiley & Sons, Hoboken, pp. 19-34.

Martin, M., I. Mullis and S. Stemler (1999), "TIMSS Questionnaire Development", *TIMSS 1999 Technical Report*, International Study Center Boston College, Chestnut Hill.

Nuttall, D. L. (1994), "Choosing Indicators", in Riley, K. A. and Nuttall, D. L. (Eds.) *Measuring Quality: Education Indicators - United Kingdom and International Perspectives*, Falmer, London.

OECD (2005), *PISA Technical Report*, OECD Publishing, Paris.

OECD (2004), *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications*, OECD Publishing, Paris.

QAA (2007), *Benchmark Statement for Economics, Quality Assurance Agency for Higher Education* http://www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/Economics.pdf

Schmidt, W. H. and L. S. Cogan (1996), "Development of the TIMSS Context Questionnaires" in M. O. Martin and D. L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*, Boston College, Chestnut Hill.

Siniscalco, M. T. and N. Auriat (2005), *Quantitative Research Methods in Educational Planning: Questionnaire Design*, International Institute for Educational Planning/UNESCO, Paris.

© OECD 2012

Tuning Association (2009a), "A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Economics", OECD Education Working Papers No. 59, OECD Publishing, Paris. *www.oecd-ilibrary.org/education/tuning-ahelo-conceptual-framework-of-expected-and-desired-learning-outcomes-in-economics_5kghtchwb3nn-en*

Tuning Association (2009b), "A Tuning-AHELO Conceptual Framework of Expected/Desired Learning Outcomes in Engineering", OECD Education Working Papers No. 60, OECD Publishing, Paris. *www.oecd-ilibrary.org/education/a-tuning-ahelo-conceptual-framework-of-expected-desired-learning-outcomes-in-engineering_5kghtchn8mbn-en*

Van Vught, F. (2008), "Mission Diversity and Reputation in Higher Education", *Higher Education Policy*, Vol. 21(2), pp. 151-174, CHEPS, Eschede.

Van Vught, F. (Ed.) (2009), *Mapping the Higher Education Landscape: Towards a European Classification of Higher Education*, Springer, Dordrecht.

Van Vught, F. et al. (2008), *Mapping Diversity: Developing a European Classification of Higher Education*, CHEPS, Eschede.

Yamaguchi, J. (1997), "Positive Versus Negative Wording", *Rasch Measurement Transactions*, Vol. 11(2).

© OECD 2012

NOTES

1 Initially, the feasibility study design did not envisage the development of a framework for the Generic Skills strand since the decision was made to adapt an existing instrument developed in the United States and, consequently, its underlying theoretical underpinning. However, during the unfolding of the feasibility study the TAG and some stakeholders felt a need for a Generic Skills reference framework while discussing the validity of the Generic Skills Instrument in an international setting. The decision was thus made by the AHELO Consortium to also develop a Generic Skills framework.

2 The later addition of multiple-choice items to the Generic Skills Instrument required extending the administration time for this instrument to 120 minutes.

3 The only exceptions are for a few students in Quranic studies in Kuwait who responded to the test on paper as a result of limited exposure to computer use in their programme.

4 This is covered in more details later in the chapter.

5 The decision to complement the Generic Skills CRTs with MCQs was based on the TAG's recommendation, made in April 2011, to provide additional data for quality control, equating, DIF analyses, and scaling of the assessment instrument results.

6 Due to scheduling, time and resource constraints, MCQs were not subjected to qualitative review by students or faculty.

7 Finland, Korea, Kuwait, Mexico, Norway and USA (CT, MO and PA). Three countries joined the Generic Skills strand after the selection of performance tasks was made: Colombia, Egypt and the Slovak Republic.

8 Subject benchmark statements provide a means for the academic community to describe the nature and characteristics of programmes in a specific subject or subject area. They also represent general expectations about standards for the award of qualifications at a given level in terms of the attributes and capabilities that those possessing qualifications should have demonstrated (QAA, 2007).

9 Among the broad range of engineering fields, the field of civil engineering (hereafter, "engineering") was selected for the AHELO feasibility study to test whether it was possible to produce an assessment in a scientific professional discipline.

© OECD 2012

10 The TAG suggested comparing student performance results on generic skills discipline items, such as those identified in the Engineering framework with those generic skills assessed within the Generic Skills strand, taking advantage of the four countries administering both strands to conduct cross-testing with the same body of students in each country. However, the recommendation could not be pursued due to time constraints.

11 Psychometric analysis results will be presented in Volume 2.

12 The AHELO Translation, Adaptation and Verification (ATAV) workbook was used to track detailed transaction-level records of any changes made to the source version of the document for management, analysis and audit purposes (see AHELO Consortium, 2012d for more details).

© OECD 2012

CHAPTER 5

IMPLEMENTATION

This chapter goes through the various aspects of the second phase of the feasibility study, the implementation. It looks at the process of selection of participants (including sampling), assessment delivery, response rates and scoring.

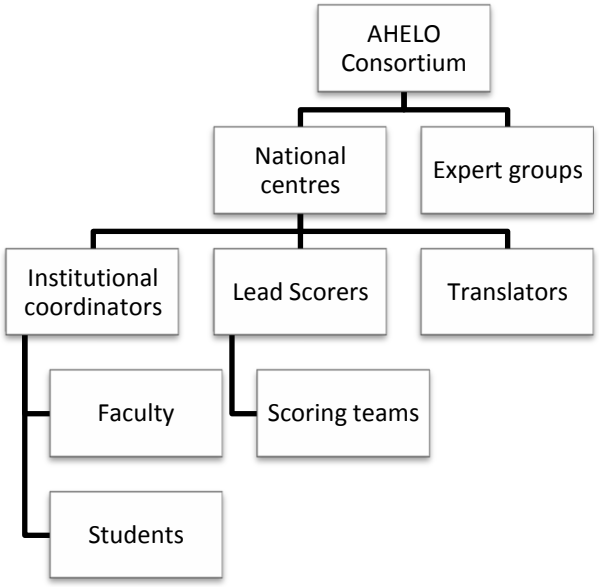
© OECD 2012

Management of AHELO implementations in countries and institutions

National higher education systems and volunteer higher education institutions (HEIs) played a key role in assessing practical feasibility of the AHELO study. When the terms of reference for the AHELO feasibility study were published in 2009 (OECD, 2009), 10 countries indicated their intention to participate in 11 field implementations: four in the Generic Skills strand, four in the Economics strand and three in the Engineering strand. By the time student testing sessions started in 2012, 17 countries had joined the study and the number of field implementations (i.e. strand and country combinations) had reached a total of 25.[1]

AHELO participating countries first had to set up national infrastructure for both instrumentation and implementation phases of the study. The role of the NPM or National Centre was to act as an intermediary between the OECD, government agencies, the AHELO Consortium, institutions and other stakeholders, and to coordinate all AHELO activities within that country (See Figure 5.1). Within countries, the AHELO National Centre and roles of National Project Managers (NPMs), Institution Co-ordinators (ICs) and Test Administrators (TAs) were most directly relevant to survey operations, and more specifically, test administration.

Figure 5.1 - AHELO feasibility study communications structure



Whether or not countries chose to establish their National Centre as part of a government structure or as part of a university research group, they were required to identify a NPM who would assume the coordinating role for the implementation of the study in their country (see Box 5.1). All communications between the AHELO Consortium and those involved in implementing the study at the national level were channelled through the National Centres at country level.[2] ICs were responsible for all communications with students, and usually with faculty, although in some cases National Centres communicated directly with faculty. Lead

© OECD 2012

Scorers were responsible for managing the system scoring team and liaised closely with National Centres.

Box 5.1 - Key players within countries

National Project manager (NPM)

The NPM was responsible for implementing the AHELO feasibility study at the national level, and ensuring that all required tasks were carried out on schedule and in accordance with the prescribed technical standards and operations guidelines, and for documenting processes implemented at national level. NPMs were also responsible for attending international conference calls and meetings, recruiting different groups of experts, coordinating country level general meetings with institutions and other stakeholders, including policy level discussions.

Institution Coordinator (IC)

The AHELO feasibility study implementation procedures required an IC to be appointed at each HEI participating in the study to ensure a successful assessment and standardized implementation of the assessment. In addition to liaising closely with the NPM, the IC's role involves assisting the NPM in the drawing of student and faculty samples, providing institutional information, and working with TAs to organise test administration and context instruments. ICs were responsible for maintaining test security and were required to report all deviations from the survey procedures to their NPMs (AHELO Institution Coordinator Manual, AHELO Consortium, 2011a). ICs were responsible for selecting and training TAs to manage testing sessions.

Most of the ICs held relatively senior roles in their HEIs and this influenced the support they received from their HEI. Mandatory training for ICs – including training on the use of the test system – needed to be arranged about 1-2 month prior to testing. At some HEIs a large number of people were involved in implementing AHELO, with teams of up to 23 individuals in some HEIs, including administrators, IT specialists, TAs and graduate students. In other HEIs, ICs managed all activities themselves, which they found quite burdensome.

Test Administrators (TAs)

TAs also played an important role in making sure that test administration was carried out smoothly and uniformly. TAs worked closely with ICs to administer tests to students within HEIs and they were responsible for managing and supervising the test administration (AHELO Test Administration Manual, AHELO Consortium, 2011b).

The number of TAs involved in each HEI was determined by the number of participating students and testing rooms to be used, based on one TA per test room and an approximate ratio of one TA to every 40 students. TAs should not have a direct personal or professional relationship with any student in the assessment sessions that he or she administers.

Countries reported that appointed TAs' academic backgrounds ranged from fellow professors and teaching colleagues to graduate and PhD students. In some countries, members from the NC team served as TAs, as well as PhD students, administrative staff or IT-personnel from the HEI (Brese and Daniel, 2012).

© OECD 2012

Communications were an important aspect of the practical implementation of the AHELO feasibility study. Several strategies were used to provide as much support as possible to National Centres as they implemented the study in their countries, including the AHELO Exchange (an online wiki) which was managed by the AHELO Consortium.

Training courses and instruction manuals also played a key role. International manuals and guidelines describing explicit procedures for each distinct AHELO activity were developed for National Centres, but some were specifically created for Institution Coordinators (ICs), Test Administrators (TAs) or Lead Scorers (LSs) and drew upon methodologies developed for previous other large-scale international educational assessment initiatives (e.g. PISA) or higher education policy analysis.

The timetable for field implementation was relatively short (see Table 5.1). The AHELO Consortium provided NPMs with a National Management Manual (AHELO Consortium, 2011c) in August 2011 that outlined the schedule of key activities for field implementation help them plan their country's activities. In the initial design of the AHELO feasibility study, testing sessions were to be administered in line with countries' academic calendars[3] but due to the decision made in March 2010 to phase the study (see Chapter 3), the window for test administration was limited to the five-month period between January and May 2012.

Table 5.1 - Key AHELO feasibility study implementation activities and timelines

Implementation activity	Scheduled timelines	Completion date
A. Instrument localisation and institution recruitment	From August to November 2011	January 2012
B. Finalising online versions of instruments	From November to December 2011	January 2012
C. Recruiting and training ICs, TAs and scorers	Two months prior to testing (training)	Well in advance of testing
D. Sampling students and faculty	One month prior to testing	From one or two weeks prior to testing and until 24 hrs prior to testing
E. Testing	From January to May 2012	2 June 2012
F. Scoring	Following testing and prior to 30 June 2012	2 July 2012
G. Data collection (FCI and ICI)	From January to May 2012	17 June 2012

Overall, scheduled timelines were adhered to by most countries but the limited timeframe posed three challenges for NPMs. *First*, delays occurred in finalising student populations at institutions – for one country in the Engineering strand, for example, populations were not finalised in April 2012, just a few weeks before the deadline to complete country implementation. *Second*, it was more difficult for some countries to schedule testing with many students on internships, taking examinations, or finalising major research projects. *Third*, with

© OECD 2012

different academic calendars, the definition of student populations had to become any students in their final-year of a bachelor-level degree.

Late-joining countries also struggled to meet international timelines. One country officially joined the AHELO feasibility study in February 2012, and the decisions of two other countries to participate in an additional strand were not made until April 2012. In the latter two cases, the AHELO Consortium was able to work with the two countries on some ground work ahead of the official commitments being made. The late decision to join the study also generated uncertainty for one country that led a number of institutions that had previously indicated their willingness to participate, to withdraw, and the process of adapting the instruments to local context could not be fully completed.[4]

Despite variations in countries' experiences with managing the AHELO feasibility study, both with establishing their National Centres and the timing at which they joined the study, comments from NPMs suggest that the organisational arrangements functioned well (Brese and Daniel, 2012).

Selection of institutions and securing institutional engagement

Ensuring participation of selected institutions in an international study can be challenging. For this study, NPMs were asked to organise ten institutions per strand to participate voluntarily while trying to get a mix of participating institutions that would reflect the diversity of the higher education system.[5] Notwithstanding the requirement for countries to reflect diversity in their selection of HEIs, the participating HEIs were not, and should not be, considered representative of regional or national populations.

Almost all NPMs reported that invited HEIs in their countries showed great interest in the study and in one country more than 30 HEIs expressed interest in participating. National Centres developed a range of institutional engagement strategies to recruit institutions and sustain their interest and participation in the AHELO feasibility study. For example, several systems created national websites to promote AHELO to institutions.

In international studies it is not unusual to see selected institutions withdraw at the time testing starts. However in the AHELO feasibility study, for 19 out of the 25 field implementations (19/25), the number of participating institutions was as expected and in total, 248 institutions took part[6] (see Table 5.2).

© OECD 2012

Table 5.2 - Number of participating institutions per field implementation

Strand	Country or system	Total participating HEIs
Generic Skills	Colombia	15
	Egypt	10
	Finland	12
	Korea	9
	Kuwait	6
	Mexico	13
	Norway	5
	Slovak Republic	16
	United States (CT, MO, PA)	11
Economics	Belgium (Fl.)	2
	Egypt	10
	Italy	10
	Mexico	10
	Netherlands	2
	Russian Federation	17
	Slovak Republic	8
Engineering	Abu Dhabi	3
	Australia	8
	Canada (Ontario)	9
	Colombia	27
	Egypt	10
	Japan	12
	Mexico	10
	Russian Federation	10
	Slovak Republic	3
Total		**248**

For six field implementations, the number of participating institutions was lower than expected and NPMs pointed to project delays as the implementation phase was postponed, resulting in budgetary uncertainties for the HEIs. Another reason for institutions' withdrawal was that the timing of implementation coincided with the busiest time of the institutions' academic calendar.

NPMs and ICs explicitly suggested that HEIs be funded directly or supported in funding student incentives to engage them in the study. Better and more public relations work to promote AHELO in the media was seen as worthwhile to encourage HEIs to participate. The most frequent recommendation to ensure participation of HEIs was to provide feedback to HEIs and

© OECD 2012

students on their own results as well as in comparison with other institutions in the country and on an international level (see Box 5.2).

Box 5.2 - Italy - Engagement of institutions

The interest raised among Italian Universities by the economics strand of AHELO was remarkable. Even though only ten institutions were selected for the final test administration, there were initially 32 universities wishing to take part, out of the 52 institutions which deliver courses in economics. The high rate of institutional participation turned into the involvement of more than 1 000 students in the country. Such figures, coupled by a successful management of the procedures, shed the light on the key features that led to the outcomes achieved: an extensive activity carried out by the NPM and her national office, aimed at promoting the importance of the project among academics and other relevant stakeholders; the opportunity, presented to the universities, to adopt the results performed in the test as a means of incentive to stimulate students' participation, awarding them extra credits for their university career according to the score achieved, compared with the average score reported at institutional level.

However, the possibility, initially allowed by ACER to use, as an incentive, the students' scores proved not to be feasible after the test: only few and insufficient data were made available to the National Centre and, in turn, to the Institution Coordinators. Unfortunately, this led to a loss of credibility of the National Centre in front of the ICs and, similarly, of the latter in front of their students.

NPM for Italy

Sampling of students and faculty

Even though the selection of participating HEIs in the AHELO feasibility study, by design, did not use a sampling approach, probability sampling is nevertheless an important element of the feasibility study for the selection of student and faculty respondents. This plays a critical role in ensuring the comparability and validity of the student assessment and faculty survey results. The main challenge for the study was that sampling has not been widely used in the higher education context, due to the perceived difficulty of identifying or obtaining population lists.

The AHELO feasibility study sampling design therefore provided for the selection of probability samples of student and faculty respondents within each HEI, whereby individual participants are randomly selected from the list of in-scope potential participants for each population of interest in participating HEIs.

All sampling activities rely on the existence of a complete, exhaustive and current sampling frame (or target population). NPMs[7] were asked to provide the AHELO Consortium with a sampling frame for students and for faculty. In broad terms these frames consisted of unit-record information for a range of variables specified in the AHELO Sampling Manual (AHELO Consortium, 2011d). These frames were then validated by the AHELO Consortium, samples were drawn as appropriate, and login information was provided to the NPM or NC prior to data collection.

© OECD 2012

Sampling was primarily the responsibility of National Project Managers (NPMs), with support from the AHELO Consortium and using the AHELO Sampling Manual describing the steps required to acquire a suitable selection of students and faculty using a single-stage random sampling design approach (AHELO Consortium, 2011d). It was developed in general and flexible terms to accommodate for a range of country- and institution-specific situations and variations in the general design were possible, provided they were approved by the AHELO Consortium to guarantee that estimates would remain comparable across HEIs even though the sampling strategy employed within each might differ.

Student sampling

The student target population includes all students at the end of a three- or four-year undergraduate degree in a participating HEI, i.e. if they are expected to graduate or obtain their qualification if they successfully complete all the classes in which they are enrolled.[8] Students on internships, or in a foreign country during the data collection period and students studying via an external/distance mode were excluded. In addition, students with special needs (e.g. visual impairment, reduced mobility) are to be exempted – although in-scope.

In addition:

- For the Generic Skills strand, the target population comprised all students at the end of a three- or four-year undergraduate degree in one of the programmes offered by their HEI.
- For the Economics strand, the target population comprised all students at the end of a three- or four-year undergraduate degree offered by the Economics department, or in a multidisciplinary programme who significantly majored in Economics (i.e. with over 2/3 of their credits earned in the Economics discipline).
- For the Engineering strand, the target population comprised all students at the end of a three- or four-year undergraduate degree in civil engineering of the Engineering department, or in a multidisciplinary programme who significantly majored in civil engineering (i.e. with over 2/3 of their credits earned in civil engineering).

The next step was for ICs in each participating HEI to select probability samples of students and faculty from in-scope lists, although in some countries the samples were drawn by the NPM and then returned to the ICs. For the purposes of the feasibility study, a systematic equal probability sample of 200 students was to be drawn from the implicitly stratified lists of in-scope students. These sample sizes were set to allow for some amount of non-response. In HEIs (Generic Skills strand) or programmes (Economics and Engineering strands) where fewer than 200 in-scope students could be found, a census of all in-scope students was required instead. Alternative sampling designs could be implemented whenever there was no comprehensive list of students available to the IC (or if available lists did not provide the programmes in which students are enrolled) and if time and resources did not allow the IC to construct such a list.

Unit record information for all or most data elements of the sampling frame were provided by 72% of the 249 HEIs that went into fieldwork. However another 5% of HEIs provided only

© OECD 2012

summary statistics for some data elements, 16% provided only a total count of the number of students in the population and 7% provided no information on the population at all.

The patterns of sampling frames' provision were shaped by country rather than by strand or HEI.[9] There were various reasons why National Centres did not provide a frame in the prescribed format. In some cases, privacy considerations prevented access to such information or its aggregation, or such information was available but HEIs would not supply it due to ethics considerations, which might have been overcome had there been more time.

For those HEIs that provided information on the frame the survey population ranged from 10 to 8 967 across participating HEIs, with a median size of 201 students. In practice however, managing exclusions and exemptions created a considerable amount of confusion and complexity:

- In many HEIs where most or all students were on internships in their last semester this could have resulted in negligible or zero student target populations. NPMs and ICs therefore chose to deviate from prescribed protocols and to include such students in their student target population.
- Some ICs also struggled with the definition of disciplines and offered various interpretations of "economics" or "civil engineering", with implications on identifying student exclusions.[10]
- Some programmes within HEIs did not have clear academic years while some HEIs catering to students from disadvantaged backgrounds had more flexible learning paths and did not rigorously enforce the order of units studied or "years" of study.
- Some "irregular students" only needed to complete 1 or 2 courses to finish their degree.

In the end, few countries identified exclusions in the sampling frames, and the numbers were low in cases where this was done.[11] Given the few exclusions, the survey population closely resembled the target population. Generally, as expected, the populations for the Generic Skills strand spanned the entire HEI and were larger than for economics or engineering. Not surprisingly, the testing strand was the major determinant in whether a census or random selection procedure was used. Sampling was employed for around 90% of HEIs[12] in the Generic Skills strand, around 15% in Economics (which often have cross-faculty or larger programmes), and only around 3% in the Engineering strand (see Box 5.3).

In practice, the AHELO Consortium received frames and samples much later than planned in the majority of cases, which severely compressed the time available to check sampling and provide logins prior to testing. Most countries were able to supply composite frames and samples which incorporated all participating HEIs, but in some cases, the frames and samples were provided separately for each HEI, thereby multiplying the work required for the Consortium to check samples and return logins.

© OECD 2012

Box 5.3 - A Census Approach in Abu Dhabi

Given the small size of the sector, particularly representing the final-year cohort of students in a Bachelor of Civil Engineering program, Abu Dhabi decided to engage in a census to obtain complete system coverage (for that programme) rather than consider sampling. It was quite challenging to catch up with two years of AHELO activities and meet the tight deadlines of the project, but the tremendous support of the government facilitated the manageability of mobilizing all stakeholders. Also, the careful selection of highly experienced ICs, who worked closely with the students, TAs, and the National Centre (Abu Dhabi Education Council), resulted in the high student participation rates.

Abu Dhabi NPM

Faculty sampling

The faculty target population includes all faculty members whose primary appointment is in an academic or interdisciplinary unit and whose responsibilities include teaching undergraduates. Part-time, replacement or invited faculty, faculty on parental or sabbatical leave, faculty teaching courses only to students outside of their respective department, research academics or senior executives without current teaching roles, honorary staff, and administrative staff were excluded. Faculty teaching courses within their respective departments but also assuming a significant amount of administrative work (e.g. reduced teaching load during their tenure as Department Head) were to be exempted – although in-scope.

- For the Generic Skills strand, the target population comprised all faculty who, during the academic year covering the collection period, were responsible for one or many undergraduate classes offered to the students enrolled in one of the programmes of their own department.
- For the Economics strand, the target population comprised all faculty members of the Economics department who, during the academic year covering the collection period, were responsible for one or many undergraduate classes offered to the students enrolled in the in-scope Economics programmes of the Economics department.
- For the Engineering strand, the target population comprised all faculty members of the Engineering department who, during the academic year covering the collection period, were responsible for one or many undergraduate classes offered to the students enrolled in civil engineering of the Engineering department.

For the purposes of the feasibility study, a systematic equal probability sample of 40 faculty was to be drawn from the implicitly stratified lists of in-scope faculty. However, in many cases, HEIs experienced difficulties in defining in-scope faculty. It was not easy to determine from institutional records the fraction of staff, whether or not they held teaching responsibilities and the cohort of students they primarily taught. Exclusions and exemptions were also used somewhat inconsistently across countries.

© OECD 2012

No faculty populations were provided at all for 53 HEIs and only 119 HEIs provided unit-record data for all or most elements of the faculty sampling frame as per the technical standard requirements.[13] Of these 119 HEIs, 48 carried out a faculty census and 71 applied random sampling. Random sampling was more frequent in the Generic Skills strand, while the Economics and Engineering strands relied more on censuses given the smaller population sizes of faculties at department level.

For the 53 HEIs where no faculty data was provided, the AHELO Consortium decided to provide faculty logins nevertheless, in the hope that at least some responses would be received. This led to 550 more faculty responses than would otherwise have been obtained. For these cases, NPMs were requested to match these responses with individuals in the population at a later date, but the retrospective matching of these responses with individuals in the population proved difficult to achieve, and it is thus not possible to ascertain response rates in these cases.

In the majority of cases faculty sampling was completed only one or two weeks before data collection, and even closer in the most difficult cases. This caused significant pressure to check samples, allocate login details and return this information to NPMs at very short notice.

Timing was also an issue with regard to the chosen data collection period which, according to several ICs, coincided with the end of the semester and collided with exams. In such cases, faculty were reportedly concerned with exam issues adding an additional burden to participate in an external survey.

Electronic delivery

Student testing, data collection and scoring were all undertaken online. The SoNET assessment system was adapted and used for the Economics Assessment, Engineering Assessment, Generic Skills multiple-choice questions (MCQs), Student Context Instrument, Faculty Context Instrument and Institution Context Instrument. The one component for which the test system was not used was the Generic Skills constructed-response tasks (CRTs) that were both administered to students and scored on an adapted version of the CLA platform.

The adapted SoNET test system contained three modules: one for testing, one for administration and one for scoring. The testing module was used for both student testing and contextual data collection from students, faculty and HEIs. Working from an existing platform, the test system was customised over a period of several months, using the English source version of the assessment instruments and contextual surveys to set up user functionality and appearance. The development process involved load testing of servers and input of dummy data in all national versions to ensure that data was recorded correctly.

The CLA platform was adapted for online administering and scoring of the Generic Skills strand constructed-response tasks. The adapted CLA platform contained three interfaces: the proctor (invigilator), the student and the scorer interfaces. While adaptations necessitated removing some features of the original test system that were irrelevant to an international context, other functionalities were added, such as the facility for NPMs to monitor and support all testing activity at HEIs within a country. Similarly, the scorer interface included added features for Lead

© OECD 2012

Scorers (LS) to use during scoring to approve or "re-queue" entered scores. User Acceptance Testing (UAT) was used to verify the functionality of the adapted platform interfaces before localisation work began. Back and forth testing between the AHELO test system and the CLA adapted platform was also conducted to ensure seamless transition between the two test systems.

The two test systems were integrated into the AHELO test system.[14] The AHELO test system allowed for a seamless transition between the administration of the assessment instrument and the student contextual survey.

Study delivery mechanism and preparation

Once the translation, adaptation and verification of assessment instruments and contextual surveys had been completed for each country, all testing materials were uploaded to the AHELO test system[15] and NPMs were given an opportunity to review them. NPMs were also asked to ensure that the AHELO test system was tested at each participating HEI (see Box 5.4). This was particularly important as the technology used varies greatly within HEIs from different countries around the world and this check identified a range of issues that were resolved prior to student testing. However, as indicated below, some issues were not identified during this system test.

In one country, the Ministry of Education purchased 67 notebooks that were transported from one HEI to another (see Box 5.5). This approach meant that test administration was centralised and there was no need to rely on, or to pre-test, institutional computers.

Box 5.4 - IT strategy and preparation for AHELO tests in Egypt

An effective IT strategy was developed and implemented to carry out the concurrent AHELO online tests in 19 Egyptian Universities:

- Establishing database of all Universities' IT facilities (number of computer labs, specification of computers, internet service specifications and qualification of technical teams).
- Selecting only universities that fulfilled the required IT facilities.
- Field visits by a national technical team for verification of IT specifications and facilities of the selected universities using checklists, and for training of TAs and IT staff.
- Three check cycles of the whole AHELO online test system using the assigned computer labs.
- Ensuring Internet Service by coordination with the Egyptian Universities Network which works on through two different sources of international internet provider (2 Gigabit/sec speed) and with University Data Centres (68 Mb/sec). At the time of test sessions, they cooperated to dedicate the bandwidth only to the assigned labs in each university to ensure high availability of bandwidth, redundant of services, data security, and vulnerability monitoring during the test sessions. The test sessions' timing, starting 2 pm Cairo time (at the end of working hours) was chosen to ensure a relatively smaller overload of the internet networks all over the country.

© OECD 2012

- Making all universities totally secured through a series of security hardware appliances and software applications as well as restricted all applications except the internet explorer and all sites except the AHELO Test URL.
- Establishing a Help Desk Team for continuous reviewing and fine tuning of the technical stability of the selected universities, and following up problems with the Rectors of Universities and ICs.

Ibrahim Shehatta, Egyptian AHELO NPM

The AHELO test system was used by students, faculty and ICs between January and July 2012. Peak usage occurred in April and May 2012, with almost 10 000 users accessing the AHELO test system during that period. Testing involved almost 23 000 students at 248 HEIs in 17 countries. All countries reported to have complied with the required test session timing, i.e. 120 minutes overall for the two discipline-based strands, and 150 minutes for the Generic Skills strand, including the administration of the student contextual survey and the introduction and closing of the test sessions.

The duration of testing periods varied from one field implementation to another. The number of test sessions held at each participating HEI also varied significantly, according to various factors including the availability of test rooms, number of computers in test rooms, size of student cohorts and the timing of the test in relation to other institutional assessments. Some HEIs could test all students in one session, whereas others needed to organise more than 20 test sessions.

Around half of the participating countries reported to have offered several test sessions per HEI, at different times and days, in order to increase student participation. The number of test sessions in HEIs was reported to have been between one and 14 sessions, where the number of students per test session varied between one and 75. Generally it was reported that bigger testing groups and rooms were more challenging with regard to administration and coordination, whereas smaller testing groups (i.e. up to 20 students) worked best. Few countries reported that test sessions needed to be cancelled or postponed due to the availability of too few students.

Box 5.5 - The use of notebooks to limit IT preparation within institutions in the Slovak Republic

Slovakia joined the AHELO project later than the majority of other participating countries so the time for preparation of HEIs was shorter. Trying to ensure the smoothest possible operation of the testing phase, the Ministry in cooperation with 3 universities working as NPMs in single strands decided to provide NPMs with “mobile computer labs” to carry out the tests. The labs contained 30 notebooks for the Economics and Engineering strands and 50 notebooks for the Generic skills strand. So it was necessary to organize several sessions carrying out tests at single HEIs.

The process was as follows: the NPM came with his team to an HEI participating in AHELO with all the notebooks set up in a uniform way for all sessions. The HEI just had to provide a room for testing and one connection point to the Internet. All notebooks were connected to a private network and tested students could access only the tools inevitably needed for the test. All other

© OECD 2012

means were filtered out or unplugged.

This setup proved very efficient and effective. Compared with a situation where the computers of the HEIs would be used, much fewer IT technicians and administrators had to be involved, especially as the HEIs have, in general, slightly different IT security policies. So doing things in this way it was possible to a large extent to limit the necessary discussions with IT administrators at single HEIs regarding the setting of the IT environment for testing.

Slovak Republic NPM

The majority of testing sessions ran smoothly as most HEIs had been able to deal with potential problems prior to the start of testing activities. In the few cases where problems did occur, some of the difficulties were due to institutional internet connections (insufficient bandwidth to allow for all students to test simultaneously) or security settings (firewalls, pop-ups, etc.) on individual computers. When students were unable to log on to the test, a number of issues were diagnosed and overcome locally. In general, almost all ICs reported that they were heavily involved and relied on IT staff from their HEI.

The AHELO Consortium investigated every report of problems and the AHELO test system worked relatively smoothly throughout fieldwork, except on two occasions. The first occasion occurred early in the testing cycle, when a small amount of data was recorded only temporarily in the test system. This was due to instability in third-party JSON library software applications which were unstable with Internet Explorer 6/7. The issue was identified rapidly and fully resolved within 12 hours although all of the data could not be recuperated, which resulted in one HEI losing over 90% of its data.

The second occasion occurred on 26 April when all students in the Economics Strand in one country unsuccessfully attempted to enter the test at the same time. The testing session had to be cancelled and rescheduled later in May 2012. The technical problem experienced might have been due to the large number of students attempting to test simultaneously. This possible server overload was avoided in the following large-scale testing sessions (with over 1 000 students) by staggering start times across participating HEIs to avoid server overload.

Close attention to protocols, by which logins were distributed to NCs and HEIs, ensured that high levels of test security were maintained. In common with all local, national and international testing, however, it was impossible for NCs or the AHELO Consortium to control the access of each individual to test materials or to monitor testing in each testing room. In HEIs with a number of test sessions, it was difficult for even ICs to monitor all testing in their own HEIs. As such, there is no certainty that sampled students used the test login details allocated to them and there is equally no certainty that test procedures were followed precisely.

Plagiarism, cheating and collusion between students are known to occur in all forms of testing and are always a possibility despite every possible precaution. In the AHELO feasibility study, one report of collusion between students at one HEI was received from one country in the Economics Strand, where scorers noted that students provided the same exact responses to some questions.

© OECD 2012

In two countries, the test sessions were video-recorded. This enabled the AHELO Consortium to see the test conditions in the test room. While it is clearly not feasible for international staff to watch every test session taking place around the world, video-recording test sessions may provide a useful record that testing protocols were adhered to in the case of any suspected collusion.

Response rates and incentives

A key challenge for the fieldwork was engaging and motivating sampled students and faculty to ensure they participated in the survey (faculty) or attended for the test session (students). The Technical Standards in relation to participation and response thresholds provide that:

- NPMs and ICs must develop appropriate mechanisms to promote participation, in line with any defined international methods developed to recruit students and faculty into the study. NPMs must consult with the AHELO Consortium regarding the appropriateness of any student participation incentives, and must record any incentives offered by participating HEIs in the national context instruments.
- The target student and faculty response rate is set at 75% of the sampled groups in each participating HEI.

Student response rates

Student response rates varied quite dramatically across countries and strands (see Table 5.3), including one HEI that did not secure any student response. The number of student responses per HEI ranged from three[16] to 273. The distribution of participating HEIs was roughly bimodal: around a quarter of HEIs secured between 20 and 50 student responses and another quarter between 125 and 175.

© OECD 2012

Table 5.3 - Student response rates by strand and country (%)

Strand	**Country**	**Minimum (%)**	**Median (%)**	**Maximum (%)**
Generic Skills	Colombia	91.5	95.0	99.0
	Egypt	43.5	75.0	94.0
	Finland	3.5	11.8	31.5
	Korea	37.7	52.3	62.3
	Kuwait	17.8	30.7	39.5
	Mexico	32.0	78.5	94.7
	Norway	4.7	8.3	10.0
	Slovak Republic	16.9	55.3	96.3
	United States (CT, MO, PA)	4.0	29.5	66.8
	Total	**3.5**	**52.7**	**99.0**
Economics	Belgium (Fl.)	19.1	46.6	74.2
	Egypt	38.2	53.8	78.5
	Italy	7.5	39.6	79.3
	Mexico	46.2	83.0	100.0
	Netherlands	3.8	4.1	4.4
	Russian Federation	54.5	88.3	100.0
	Slovak Republic	32.5	78.8	92.9
	Total	**3.8**	**74.1**	**100.0**
Engineering	Abu Dhabi	77.2	96.6	100.0
	Australia	12.3	16.8	98.1
	Canada (Ontario)		58.8	79.2
	Colombia	79.2	97.9	100.0
	Egypt	60.5	87.0	93.8
	Japan	13.9	81.1	95.0
	Mexico	70.0	86.1	97.8
	Russian Federation	80.0	91.5	100.0
	Slovak Republic	50.5	68.6	78.3
	Total	**12.3**	**89.2**	**100.0**

A close examination of response rates achieved in different countries and strands reveals that four countries faced consistently low response rates across their HEIs while five other countries achieved consistently high response rates suggesting possible country-specific systemic or motivational issues. The other 16 countries achieved a broader range of response rates across their HEIs, suggesting that institutional strategies might have played a greater role relative to systemic or cultural factors in those contexts.

Observed response rates were considerably higher in the Engineering strand and the Economics strand than the Generic Skills strand. These patterns might be explained by cohort effects. Cohorts in the Engineering strand tended to be smaller than in other strands and it may be easier to motivate small groups of students who know each other than larger groups, particularly where selection is randomised across a large HEI, as was more commonly the case

© OECD 2012

in the Generic Skills strand. Perhaps more importantly, the type of student selection method is not neutral. Across all strands and countries, the average response rate achieved was 89% for HEIs who used a census, 68% for HEIs in which a non-random selection method was used and 51% for HEIs in which random sampling was used. Securing sufficient student responses proved to be a real challenge in some countries, as illustrated by the experience of two Nordic countries (see Box 5.6).

Box 5.6 - Student response rates in Norway and Finland

A key challenge in Norway and Finland was low student response rates, at 7.3% and 13.8% respectively. Institutions worked hard to engage students, through information, encouragement and generous incentives to participate, but with little effect.

The survey was promoted through student web pages, student newspapers, lecturers and repeated e-mails. The sampling approach made it hard to target specific groups of eligible students. Due to the long testing session, and need to sign up in advance, institutions offered generous incentives including iPad draws and shopping or cinema vouchers (worth 30-35 EUR). Despite this diversity of approaches, no clear way of securing high turnout is apparent.

There are several likely explanations for these countries' low participation rates. *First*, Nordic institutions have little authority to 'demand' or pressurize students to take part in a test that is not part of their degree: a study like AHELO is dependent on student cooperation to be successful. As the AHELO test is time-consuming (at least 2 hours to complete), students were unlikely to prioritize this in their last bachelor's semester, when they have many demands on their time. Furthermore, most students in Norway and Finland write a thesis in this last semester and are rarely on campus; this posed an extra obstacle to participation. *Second*, the test having to be done in an IT lab, under exam-like conditions may have made participation less attractive. *Finally*, institutional and student feedback suggests students saw little value to themselves in taking part, in the absence of immediate or individual feedback on performance.

NPMs for Norway and Finland

Response rates and possible non-response bias

Although the AHELO Technical Standards had called for response rates of 75% or above to minimise the negative impact of non-response biases, there were no agreed or established response rate criteria for determining when data should be included in various kinds of analyses or reports (AHELO Consortium, 2011d, 2012).

For countries in which a sampling frame was provided with relevant data elements, where high response rates were achieved, the sample yield matched population expectations very well, but it showed less fidelity[17] to the gender distribution in the population in the country where response rates were lower (see Figure 5.2). A further comparison can be made between the gender and age distributions of students (and the distribution per field of study for the Generic Skills strand) in the frame population and in the sample (see Table 5.4). Overall, the higher response rates, the more representative student responses are of the target population.

© OECD 2012

Figure 5.2 - Comparison of male students in population and sample in 2 countries and their HEIs (%)

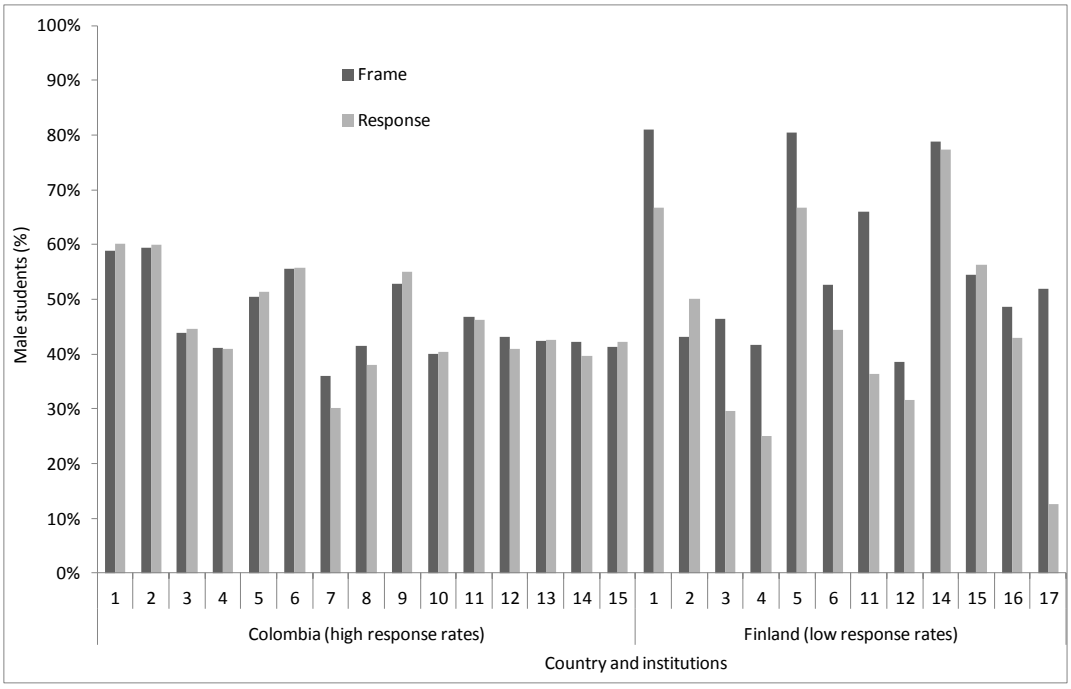


Table 5.4 - Comparison of population and response data (standard deviation)

Strand	Country	Gender (s.d.)	Field (s.d.)	Age (s.d.)
Generic Skills	Colombia	2.1	6.4	3.3
	Finland	12.9	0.9	13.4
	Korea	4.4	3.1	12.2
	Mexico	7.4	1	16.6
	Norway	7.1	5.8	11.6
	Slovak Republic	5.5	14.2	4.1
	United States (CT, MO, PA)	16.3	5.5	6.7
Economics	Belgium (Fl.)	1.7		2.5
	Mexico	4.5		35.5
	Russian Federation	3.1		2.4
	Slovak Republic	5.8		1.4
Engineering	Australia	9.4		11.9
	Colombia	1.4		3.8
	Mexico	2.8		
	Russian Federation	2		1.5
	Slovak Republic	5.6		1.3

© OECD 2012

As the above analyses suggest, response rates are important to ensure the fidelity of the estimates drawn from a sample of students to the population they represent. In the absence of agreed response rate criteria in the higher education context, the AHELO Consortium recommended the following guidelines for future practice, whereby:

- Data would be included in institutional analyses on the condition that response rates are above 25%[18], regardless of frame quality or selection method. This would prevent bias from non-response having a significant impact on scores. However they would not be included in cross-institutional analyses. Experience from the feasibility study suggests that 86.3% of participating HEIs met this criterion.
- Data would be included in cross-national analyses on the condition that *i)* a unit-record sampling frame has been provided and a census or random selection procedure has been used and *ii)* response rates are above 50%[19]. Experience from the feasibility study indicates that only 50.4% of participating HEIs would be admissible based on these conditions considered jointly. The response rate threshold would be more constraining for the Generic Skills strand, in which the nature of the sampling design made it more difficult to secure high response rates (see Table 5.5).

Table 5.5 - Proportion of HEIs attaining the 50% response rate threshold, by strand (%)

	Response rate of 50% or more
Generic Skills	51.0
Economics	76.3
Engineering	88.0

Student engagement and effectiveness of incentives

From the outset, engaging students to participate in the AHELO feasibility study was seen as the most crucial practical challenge to be overcome. Participating countries and HEIs therefore adopted a range of different strategies to ensure high participation rates.

In many cases, HEIs relied on a multi-faceted strategy of student recruitment involving communication and marketing as well as incentives. Students were often sent letters and emails of invitation, received phone calls and text messages, were invited to seminars on AHELO or received information on AHELO during class, and many HEIs also used brochures, flyers and posters for promotion.

In a few countries, the opportunity to participate in the AHELO feasibility study was sufficient to motivate students (see Box 5.7). In those countries, non-financial incentives like certificates for participation were used. NPMs and ICs reported that students felt an honour to participate in an international study (Brese and Daniel, 2012). For example, one IC explains that: “The only incentive for the students was the importance and privilege of participating in an initiative (feasibility test) that emerged from the OECD, where other members and non members of the international organization, would also participate”.

© OECD 2012

In general, however, participation of students on an entirely voluntary basis seemed to be problematic. In places where the intrinsic reward of participating in the AHELO feasibility study was deemed unlikely to be sufficient, a variety of incentives were used. It needs to be stressed in this respect that the concept of "incentive" varies across contexts from tangible material incentives to more abstract motivations. In certain instances the NPM/IC indicated that no incentive had been offered despite AHELO being pitched as a "compulsory" exercise. Some HEIs made participation in the AHELO feasibility study compulsory and as might be expected, these approaches were successful in securing high response rates.

Box 5.7 - Student engagement strategies

Slovak Republic - Pride to represent their university, curiosity and gift vouchers

In the Slovak Republic, the motivation tools to engage students to participate in AHELO feasibility study were used at two levels.

The first was at the level of the Ministry. The Ministry declared the implementation of AHELO feasibility study in Slovakia as a central development project and invited HEIs to take part. The Ministry provided funding for the participating HEIs: a fixed amount of EUR 5 000 per participating institution and EUR 50 for each student who actually participated in the test.

At the second level, the HEIs motivated students firstly by emphasizing that participation would represent their HEI. This tool was, in some cases, strengthened by awarding students with certificates from the rector. The good reputation of the OECD in Slovakia was also helpful here. In addition, most HEIs also chose different direct material motivation tools such as vouchers (mostly for purchasing books). In some cases, the students received USB keys, high quality planners with a pen or T-shirts. The value of these material motivation tools was usually around EUR 20.

The students however reported that curiosity of their own achievements was their main motivation for taking part in AHELO.

Slovak Republic NPM

Colombia - Test administered as part of compulsory national exam

The way Colombia implemented the test administration was unique among the different participating countries: both the Generic Skills and the Engineering tests were administered as part of the national end-of-tertiary-education exam, called SABER PRO. This examination has been in place for 9 years now. It aims to produce learning outcomes measures to inform policy and decision making for the improvement of education quality. Since 2009, it is compulsory for students to take the examination during their last year of studies but no passing score is set and Higher Education institutions are free to use their students' results as they please. The exam consists of tests on Critical Reading, Quantitative Reasoning, Written Communication and Citizenship, that are to be taken by all students, and more specific tests like Scientific Thinking, Research in Social Sciences, Agricultural Production or Project Management, which are compulsory for some programs and optional for others.

From among the registered students for the SABER PRO exam, 4 000 students were selected to participate in the AHELO application. They all had to sit the AHELO tests on Saturday 2 June 2012, and completed the SABER PRO the next day. The coupling of AHELO with SABER PRO allowed both achieving an excellent attendance rate for all of the participating institutions; and also taking

© OECD 2012

advantage, to some extent, of the existing managing structure which is deployed during the regular SABER PRO session. Fears of technical problems arising due to the high concurrence in the web platform proved to be baseless as there were no meaningful difficulties, and the application was successfully completed. To be able to deploy this strategy it was required to have access to individual results so they could be returned to students together with those of the SABER PRO.

Colombia NPM

Australia - The importance of institutional strategies

Participation rates in Australian institutions were not as promising as were originally hoped. However, one particular university in this country employed a number of key practices to its implementation of AHELO that successfully engaged students – achieving close to full participation. The approach taken by the University of Western Sydney offers an insight into an ideal implementation model, especially for countries and institutions where motivating students to participate may prove difficult. The approach involved strong and dispersed leadership within the institution, insightful planning, and the merging of the assessment into a dialogue about learning outcomes between students and staff.

The implementation of AHELO at this institution was planned in advance, beginning in November the year prior to testing – four months before testing – following an information session facilitated by the National Project Manager. Leadership in the project came at three levels within the institution – from the Chancellery, where a member of the Pro-Vice Chancellor's office was directly involved, from the Head of the School of Engineering and from a senior faculty member who had ultimate responsibility for implementation. During planning for the testing, the team within the institution identified a core unit (a project/thesis unit) in its final year civil engineering program in which the theme of learning outcomes and graduate capabilities was appropriate. The unit was planned in such a way that AHELO became an integral part of the work during the semester. Students were asked to undertake the assessment as part of the unit of study and following the assessment, were involved in discussions to reflect on their assessment experience and the relationship between their coursework, the skills they expected to employ in the workforce following graduation and professional responsibility related to assurance of educational and practice standards.

Australia NPM

Overall, incentives were provided by 59% of HEIs that responded to the Institution Context Instrument (Table 5.6). Incentives were, in relative terms, slightly more frequent in the Generic Skills strand – possibly reflecting the greater challenges of student engagement at institutional rather than departmental level. Incentives varied widely. Feedback from ICs suggests that cash or vouchers were the most common incentives used, followed by a certificate of participation, participation in a prize draw or a gift, food or drinks, or academic bonuses in the form of increased grades (see Table 5.7). Many HEIs used a combination of incentives.

© OECD 2012

Table 5.6 - Institutional provision of incentives by strand and system

Strand	Country	Incentive	No incentive
Generic Skills	Colombia	1	13
	Egypt	3	7
	Finland	12	0
	Korea	9	0
	Kuwait	5	1
	Mexico	3	10
	Norway	6	0
	Slovak Republic	13	2
	United States (CT, MO, PA)	11	0
Economics	Belgium (Fl.)	1	1
	Egypt	5	5
	Italy	5	2
	Mexico	4	6
	Netherlands	3	1
	Russian Federation	12	5
	Slovak Republic	5	3
Engineering	Abu Dhabi	3	0
	Australia	6	2
	Canada (Ontario)	9	0
	Colombia	2	21
	Egypt	5	5
	Japan	12	0
	Mexico	2	8
	Russian Federation	5	5
	Slovak Republic	3	0

Table 5.7 – Types of incentives mentioned by ICs

Type	Number of HEIs
Cash or voucher	69
Certificate	31
Prize draw	27
Gift	22
Food or Drink	18
Increased grade	16

Cash amounts and the value of vouchers varied considerably but on average were approximately EUR 20-40 per student, while iPads were the prizes most commonly offered in

© OECD 2012

prize draws. Gifts varied immensely, from cinema tickets to pens and USB keys to movie tickets and even slippers with a university logo. However, feedback from NPMs and ICs suggests that none of these material incentives was reported as being useful to (significantly) increase student participation rates (Brese and Daniel, 2012). A number of HEIs also provided students with food and drink either before or after participating in the test and some gave assistance with travel costs.

Some ICs reported giving students some kind of academic bonus for participating in AHELO. These included bonuses available to all participants, such as an additional 5% on the capstone grades for students. Other benefits included priority in receiving a graduation certificate, preferential conditions for admission to graduate study and a note of recommendation for the thesis examiner. These strategies may work in some contexts (e.g. census) but are problematic when random sampling is used as students not included in the sample are also not eligible for the benefits which accrue from participation.

A handful of HEIs also committed to provide specific bonuses for students who performed best in the AHELO assessment, following consultation with the AHELO Consortium during an NPM meeting. For example, one extra percentage point to the degree for students who obtained a score above the average of their peers. Although NPMs and ICs reported that these strategies worked in favour of student participation, motivating students through rewards for performance proved particularly problematic, *first*, because the design of the test was unsuitable for precise reporting at the individual student level (see Chapter 2) and *second* because although the AHELO Consortium had reluctantly agreed to deliver lists of top performers to two countries during an NPM meeting, it subsequently did not deliver the promised lists in time for the graduation rosters (see Box 5.2).[20]

Besides communication activities and incentives, experience from the AHELO feasibility study also suggests that another important factor in securing high response rates is institutional culture (see Box 5.8).

Box 5.8 - United States - Importance of an institutional culture of assessment

Across the 11 US universities participating in the Generic Skills Strand of AHELO, approximately 31% of the sampled students completed the assessment – not too bad in comparison with other types of assessment or some other nations, but far below our expectations and the international goal of 75% participation. Several observations and insights might be gained, however, because of the wide variation in student participation across institutions, from a high of over 60% at two institutions to a low of – well, embarrassingly low – in others.

Most institutions made considerable effort to promote institutional awareness of AHELO and to encourage student interest, including campus posters and press releases, individual, leader-signed letters of invitation, flexible test scheduling, and repeated email and phone encouragement. These appear to have some effect, or at least have underlying relationship to students' receptivity to assessment, since institutions with aggressive promotional activities tended to achieve higher student participation. And one Institutional Coordinator did a follow-up survey that showed that students were aware of and responded to these motivational efforts.

© OECD 2012

All institutions also provided some direct incentives to students who completed the assessment, generally waiving graduation fees, providing a USD 50 gift card, or entry into a lottery for a more valuable reward. Five institutions in one state agreed on providing the same student incentive, but achieved very different results. One institution raised the ante, offering multiple incentives and gift cards of up to USD 100 to those students requiring more incentive near the end of the testing period. Such direct incentives and rewards helped, but were not in themselves sufficient to raise participation rates to the levels sought.

So what did work? What factors appear to be associated with the ability of institutions to achieve reasonable student participation in an AHELO-type assessment? What appears to affect student participation at U.S. institutions is a well-established institutional "culture of assessment," a set of practices and expectations through which academic leaders and faculty regularly use assessment to examine and improve their teaching and learning, where students expect to demonstrate their skills and knowledge beyond classroom settings and grades. Where established, institutions were able use this culture of assessment to achieve higher overall student (and faculty) participation in AHELO. And if verified by more thorough analysis once the institutional data sets are returned, such findings could be significant for any future development and participation in AHELO.

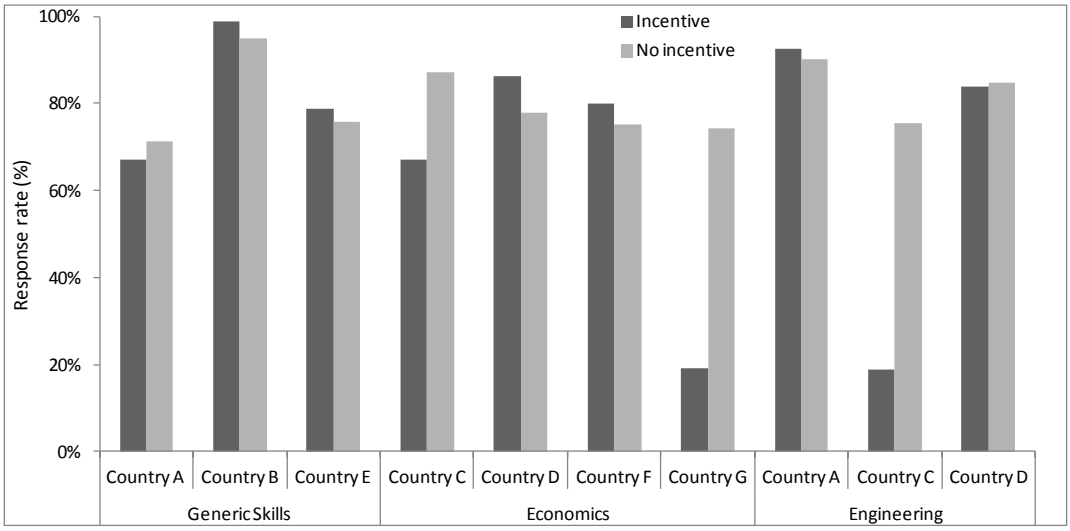
Charles Lenth, SHEEO, NPM for participating US States

Effectiveness of incentives

Incentives have the potential to affect various forms of student participation. As suggested by the substantial research into this matter, the relationships between incentives and student engagement are invariably complex, indirect and contextualised. For countries where response rates could be calculated and there was a variation in the use of incentives between HEIs, there is no clear pattern between incentives and response rates (see Figure 5.3). This underlines that institutional strategies, rather than the use of incentives *per se*, are of critical importance in determining student response rates.

© OECD 2012

Figure 5.3 – Institution response rate by incentive (%)

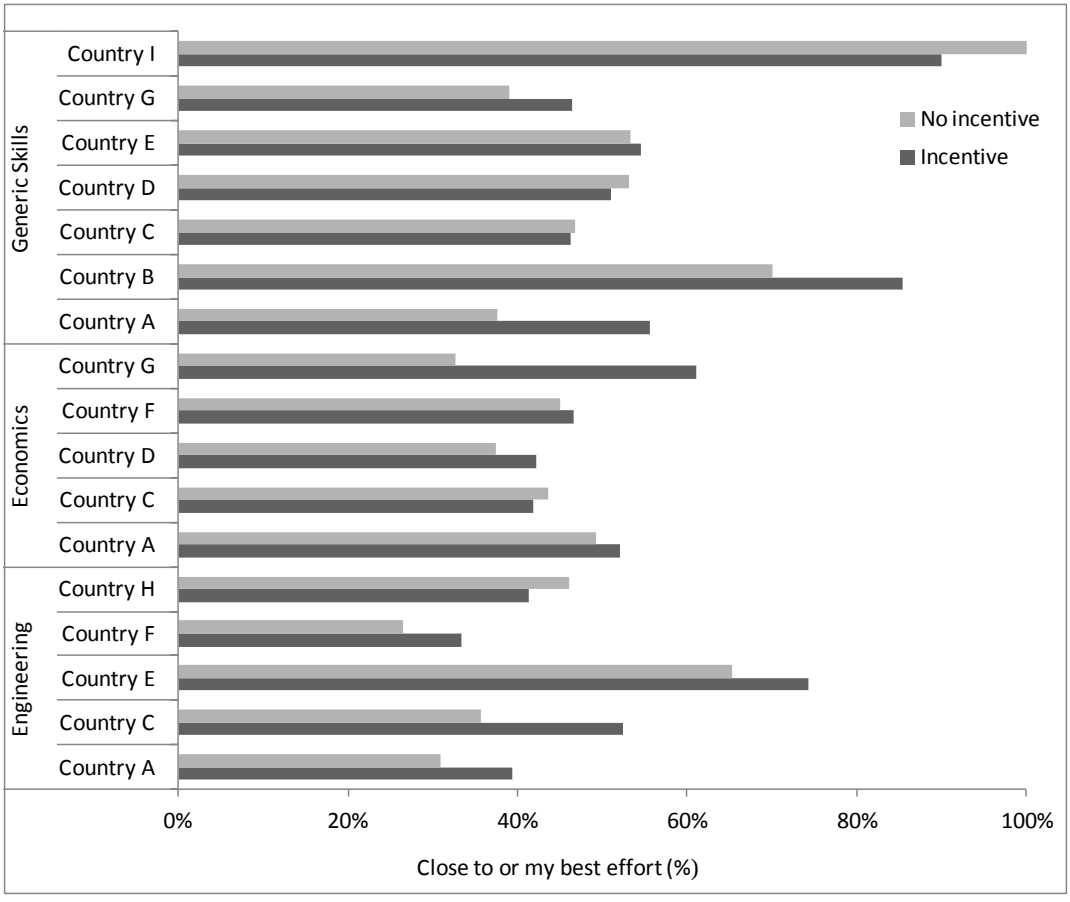


Incentives may influence more nuanced aspects of students' engagement in a non-compulsory, low-stakes test. In particular, it is useful to examine whether students motivated by an extrinsic incentive might engage less in the test. This could be manifest by moving rapidly through the test, higher levels of item non-response, or limiting the amount of effort they use to respond. In these respects, data from the AHELO feasibility study suggests that the use of incentives was not related to the time students put into the test and did not have any conclusive impact on levels of item non-response but did have a positive impact on the degree of effort that students reported putting into the test in a majority of cases (see Figure 5.4).

There is thus very little evidence to suggest that the use of incentives had any significant impact on the levels of students' substantive engagement with the AHELO assessments – except for a handful of cases where students reported putting less effort into the test. As a result, further analysis would be required at national and institutional level to assess the effectiveness of incentives in the local context.

© OECD 2012

Figure 5.4 - Effort put into taking the test by incentive (%)



Note: Only includes those field implementations where different institutional strategies were used.

Faculty responses

The average faculty response rate varied according to the selection method used. The average response rate was 65% where a census was used, 64% where random selection was used and around 83% (as much as this can be determined) where a non-random procedure was used. The number of faculty responses per HEI ranged from 1 to 55, with 19 being the median response. This is well below the target sample size of 30 faculty responses per HEI envisaged in the sampling design.

As with students, faculty response rates were highest in the Engineering strand in which student cohort sizes – and likely faculty numbers – were the smallest. It is interesting to note that in four cases faculty response rates were above 100%. This illustrates the inaccuracies which arise in using non-random sampling methods.

© OECD 2012

Nonetheless, in 12 of the 23 field implementations for which faculty information was available, the faculty response rates were better than student response rates. This finding is in line with the feedback received from NPMs who reported that it was much easier to get faculty than students to participate in the AHELO feasibility study, suggesting a greater level of interest from faculty in obtaining information on student learning outcomes. NPMs also considered that "the short length of the questionnaire clearly helped" and that success "indicates that doing a short test that can be administered via a link sent by email could be a way of generating more response".

Scoring student responses

All assessment instruments in each of the three strands included both multi-choice questions (MCQs) and constructed response tasks (CRTs). While MCQs were scored automatically, CRTs required human scoring. Scoring quality is critical to the reliability of the interpretation of CRT results and demands close adherence to standard scoring procedures, even more so in studies involving more than one language, culture or country.

Scorer technical standards and training

All student responses to CRTs were scored within the country, by country teams of scorers. In preparation for scoring and scorer training, the AHELO Consortium provided participating countries with the International Scoring Manual (AHELO Consortium, 2011e) covering scoring procedures for all CRTs in the three testing strands. More specific and secure information on scoring for each strand of work was provided in the Generic Skills scoring guide; the Economics scoring guide; and the Engineering scoring guide.

During international scoring training, test developers[21] worked closely with LSs to go through the scoring rubrics. During the October 2011 training, LSs were introduced to the approach to scoring used in the AHELO feasibility study and worked through scoring rubrics developed in each strand. In March 2012, LSs practiced scoring sample student responses. The scoring rubrics were finalised after the international training in March 2012, then translated by the AHELO Consortium and provided to national scoring teams. Examples of scoring rubrics are included in Annex B with the illustrative assessment items.

Each NPM appointed a Lead Scorer (LS) responsible for recruiting, managing and training the team of scorers and monitoring scoring procedures. Many LSs had been involved in the AHELO feasibility study for some time or were "nationally recognised experts" in their fields. The majority of LSs had prior international experience, some with the scoring of the OECD PISA Survey. The size of country scoring teams depended on an anticipated number of student responses and varied from two to 14. Scorers were faculty and graduate students. In several cases, scoring teams included one scorer from each participating HEI. Additional training for the Generic Skills strand was conducted with country teams individually to help LSs prepare for scorer training, but more importantly, to help them develop skills needed to be effective in guiding a group of scorers towards scoring agreement. Furthermore, training was based on the country's set of benchmark papers. Countries were asked to select a range of possible student responses, pre-score those responses and provide rationales for the scores allocated.

© OECD 2012

Scoring experience

Scoring started in each country once testing was complete. The majority of scoring was completed by 22 June 2012, but extensions were given to a number of countries until 30 June 2012 (and in two cases until 2 July 2012) so that outstanding scoring could be completed (see Box 5.9). All strands combined, time spent on scoring student responses to CRTs varied by country. Scoring took between two days and two weeks depending on the number of responses to be scored, the number of scorers available and the number of hours spent scoring per day. Scoring CRTs in the Generic Skills strand required more time than the two other strands due to the nature of extended responses and the requirement that all student responses be double scored. The AHELO Consortium offered content support with regard to scoring procedures as well as technical support related to the online scoring systems.

Box 5.9 - Colombia - Tight timelines

Colombia's sampled population for AHELO was high, 4 034 students were selected to take the test (3 000 for Generic Skills and 1 034 for Engineering). Adding to this the fact that testing took place in early June created a very tight schedule for AHELO.

Several factors led to the request for a time extension. 1) National scoring team training could only start once testing administration was finished. 2) Taking into account double scoring, the amount of scoring involved was very high, nearly 6 000 PTs and CRTs for Generic Skills and around 1 250 for Engineering. Meeting the initial deadline was impossible because of these factors.

The National Scoring Team worked long hours in order to achieve the largest amount of work possible in the shortest span of time and the Engineering team actually managed to finish on time. On the other hand, due to the volume of scoring involved, the Generic Skills team had to ask for an extension in order to complete scoring ensuring reliable procedures and results.

Colombia NPM

LSs monitored the performance of their scoring teams while scoring was taking place. Monitoring was conducted differently in each scoring system. While all student responses in the Generic Skills strand were double scored, 20% of student responses were randomly re-distributed for double scoring in the Economics and Engineering strands. LSs in both scoring systems could review any scores given by any scorers and provide corrections, if necessary.

For the Generic Skills instrument, a total of 97 scorers were involved across the nine participating countries, and scorers in each country varied from five to 17 (see Box 5.10). Relative to the other strands, a larger number of scorers were involved in the Generic Skills strand due to the demands of scoring extended pieces of writing, double scoring each response, and auditing every fifth or so response.

In the Economics strand, the number of scorers varied amongst countries from one to eleven. One country only had one scorer, who created two different logins to the scoring system and scored all student responses twice. This was a breach of the international scoring guidelines and the inter-scorer reliability statistics did not include data from this scorer. The number of

© OECD 2012

student responses scored by any single scorer ranged from one to around 1 500, the latter being for the country with one scorer.

In the Engineering strand, the number of scorers varied amongst countries from three to 13. In one country with a large number of responses, first scores were provided by three scorers and double scoring was conducted by just one scorer. In another country, there was a more even distribution of scoring across scorers. Two English speaking countries took the opportunity of the scoring activity to collaborate on scoring calibration.

Box 5.10 - AHELO Scoring

Finland - Scoring management in a small country

As the total number of students taking the test is Finland was low (330 students) it was decided to have a minimum number of scorers: two scorers per performance task (altogether four plus a lead scorer). This proved to be challenging when two scorers had given different scores for the same student. Then a third scorer would have been needed as the test system did not allow the Lead Scorer to make the third opinion. However, the best way to solve this problem would be to let Lead Scorer make the final decision which within this context was not the case. The second problem faced by Finland was the fact that all the scorers except one were scoring in a different city than where the coordinating institution was located. This made it difficult for the Lead Scorer to coordinate scoring activities.

NPM for Finland

Australia and Canada - International collaboration on scoring calibration

In an effort to monitor consistency of scoring across their countries, Australia and Canada engaged in a small self-funded side project in which a sample of responses from the Engineering assessment was cross-marked. In this case the collaboration was made relatively simple by virtue of both sharing the same language (none of the other participating countries in the Engineering strand were English-speaking countries).

Within the AHELO online scoring facility, each country allocated a random selection of items to be marked by a scorer from the other country. Once this sample was scored, the Lead Scorer from the country in which the scripts originated, re-scored the items to monitor consistency and ensure final scores were allocated by the country of origin. This small cross-marking exercise proved to be worthwhile in helping to validate the scoring practices being employed within the small teams undertaking the task in each country. It also helped to further strengthen the international dialogue developing about assessment of learning outcomes by building relationships between educators in civil engineering and facilitating conversations about how best practice might be achieved.

NPMs for Australia and Canada (Ontario)

Institution Context Instrument

In the AHELO feasibility study, the HEIs are both the main units of analysis and the reporting level. Therefore, it is essential that results be representative of the institution, and this requires

© OECD 2012

probabilistic sample of students and faculty members within each participating institution. However, to compare across institutions, additional contextual information is essential in reflecting institutional diversity. Learning outcomes data mean little if they are not situated within national, institutional and educational contexts. Data from contextual surveys allowed for greater nuances in the interpretation of results by identifying contextual variables that can best explain similarities and differences in student performance.

ICs were asked to complete the Institution Context Instrument (ICI) to provide institutional-level information regarding the contexts in which students study. The ICI comprised 40 items and was administered online. In total, 243 of the 248 institutions participating in the feasibility study completed the ICI. Participating institutions varied significantly in size, with the largest institution having 250 000 full-time equivalent students. The number of staff was similarly varied with a median of 51 total full-time teaching and research staff. Institutions which participated in the Engineering strand tended to be largest, and those which participated in the Generic Skills strand the smallest. Results indicate considerable variations in institutional characteristics (see Table 5.8.).

Table 5.8 - HEIs characteristics by strand (n=243)

	Generic Skills	Economics	Engineering
Full-time equivalent students (median n)	7893.0	10398.0	13300.0
Full-time teaching and research staff (median n)	443.0	758.0	622.0
Proportion budget public (median %)	62.9	69.5	50.0
Curriculum blend of broad/specialised programmes (median %)	72.4	84.5	78.7
Equal balance of teaching and research (median %)	67.7	81.0	71.6
Doctorate qualification offered (median %)	72.2	80.5	83.1
Students entering via non-traditional pathways (median %)	14.4	5.9	13.0

According to ICI responses, some institutions receive no public funding while some receive all their funding from public sources. Of participating institutions, 78% reported that their curriculum started out broad and became increasingly specialised, a characteristic which was particularly prominent among institutions which participated in the Economics strand.

Most institutions reported that their focus was equally on teaching and research. However, 29% of institutions participating in the Generic Skills strand described themselves as mainly teaching-oriented. Most institutions offered doctorate level education but, once again, this was least common among institutions participating in the Generic Skills strand.

© OECD 2012

REFERENCES

AHELO Consortium (2011a), *Institution Coordinator Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN4/FINAL&doclanguage=en

AHELO Consortium (2011b), *Test Administration Manual*, accessed 13 November from:
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN5/FINAL&doclanguage=en

AHELO Consortium (2011c), *National Management Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN6/FINAL&doclanguage=en

AHELO Consortium (2011d), *Sampling Manual*, accessed 13 November from:
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN3/FINAL&doclanguage=en

AHELO Consortium (2011e), *International Scoring Manual*
http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/IMHE/AHELO/GNE(2011)21/ANN2/FINAL&doclanguage=en

Brese, F. and T. Daniel (2012), *OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study: Report on Quality Adherence*, IEA Data Processing and Research Center, Hamburg.

OECD (2009), *Call for Tenders: Feasibility Study for an Assessment of Higher Education Learning Outcomes.*
www.oecd.org/edu/ahelo/callfortenders

© OECD 2012

NOTES

1 During the course of the study, one country withdrew its participation mostly due to funding considerations, although this decision was also partly due to the very small number of institutions offering a civil engineering program.

2 However, this communication structure did not work as effectively in the Generic Skills strand for which two different organisations were contracted for the first phase of the work (instrumentation part) and the second phase of the AHELO feasibility study (implementation). This has led to some inconsistencies in the communication structure whereby the contractor responsible for instrument development and the Consortium leading organisation frequently ran parallel communications structures that were not related to one another, and direct communications took place between the subcontractor in charge of the Generic Skills instrument development and NPMs/ICs.

3 The academic year of HEIs in many countries begins in September and ends in June, although this is not the case in all participating countries. In some countries, for example, the academic year runs from March to November, from April to March or from March to February.

4 More specifically, it was not possible to conduct focus groups to validate the assessment instrument.

5 NPMs were asked to consider a combination of criteria that make sense in their countries such as: institutional status; institutional setting; funding characteristics and institution profiles and tuition-fees dependency; size; geographical location and level of dispersion; any specificity of the institution that has been underlined or recognised nationally; and any specificity of student populations served by the institution.

6 While 249 HEIs participated in fieldwork, one institution had no students participating. As a result, student data was only gathered from 248 HEIs.

7 Depending on country-specific management structures, this was the responsibility of the NPM or National Centre.

8 In practice, this was not an issue for most participating countries in which the academic year finishes in June or July and for which this definition was directly relevant. It was more problematic, however, in countries in which academic year was such that students enrolled during the testing period (January to June 2012) and were therefore not in their final semester.

© OECD 2012

9 In the Generic Skills strand, two countries did not provide a full sampling frame as per the AHELO Technical Standards (AHELO Consortium, 2012). This was also the case of two countries in the Economics strand and four countries in the Engineering strand.

10 This suggests that there could be quite a few opportunities for over- or under-coverage of the target student populations in the survey, at least for the economics and engineering ones. This is an issue that would require further analysis and guidance to NPMs in the event of an AHELO main survey.

11 Specifically, exclusions were specified for 12 HEIs overall, although the AHELO Consortium had the impression that in various instances, countries had excluded elements prior to sending their population file for approval, (for example by asking students to volunteer to participate in AHELO and then sampling from volunteers rather than starting with institutional records).

12 These percentages only take into account countries and HEIs that provided a sampling frame.

13 Another 16 HEIs did not participate in the faculty survey.

14 For the rest of this section, the use of "AHELO test system" refers to both the SoNet and the CAE adapted test systems.

15 The AHELO test system also accommodated the testing of multiple languages within a country. In particular, Egypt and Kuwait required having the option to test students in English, in addition to the country language. US English was provided as the alternative language to Egyptian or Kuwaiti Arabic.

16 This low response at one HEI is artificially low as a result of a failure with the test system, which did not record all data and resulted in loss of data for 33 of the 36 students who took the AHELO test. It is therefore inaccurate to report on their participation rate.

17 In summary terms, the standard deviation between population and sample was 2.1 percentage points in the high response rate country, and 12.9 in the low response rate country.

18 Further replication studies of estimation biases and reliabilities would need to be conducted to verify this threshold.

19 Further replication studies of estimation biases and reliabilities would need to be conducted to verify this threshold.

© OECD 2012

20 Arguably, this extra work was out of the scope of the AHELO feasibility study and the Consortium workload was already stretched. But the fact that NPMs in these two countries committed to these bonuses on the basis of their discussions with the AHELO Consortium caused a great deal of frustration and loss of credibility at national level.

21 Test developers involved during LS training were CAE for the Generic Skills CRTs, ETS for Economics, and ACER and NIER for Engineering.

© OECD 2012

CHAPTER 6

LESSONS LEARNT ON DESIGN AND IMPLEMENTATION

This chapter goes through the main lessons which emerged from the processes described in the previous chapters: the purpose of an AHELO, the design and management of the study as well as instrument development and implementation.

© OECD 2012

The AHELO feasibility study was designed to establish whether an international survey could be developed to support better learning outcomes in higher education. Its objective was to demonstrate the operational feasibility of applying internationally comparative large-scale student testing in the context of Higher Education Institutions. Along the way, AHELO has also sharpened awareness of the importance of learning outcomes as a core part of the higher education sector's mission. Indeed, over the years since the initial proposal for AHELO was put forward at the Athens Ministerial Meeting in 2006, learning outcomes have moved from the edge of the discussions about university performance to centre stage, and the debates surrounding AHELO (see Chapter 2) have, in their own way, contributed to this movement.

Furthermore, the process of developing AHELO has provided not only raised awareness within the sector of the importance of measuring learning outcomes in higher education, but also brought to the forefront many of the complexities and challenges of doing so, not least by stimulating valuable professional dialogue and debate about the learning process and learning outcomes that institutions are – or should be – striving to achieve. This chapter reviews some of the emerging lessons from the AHELO feasibility study, which has been a valuable learning experience for all those involved. As the TAG Chair observed in 2010, sometimes it is 'the kinds of things we learn by bumping into them' which are the most valuable insights to arise from a feasibility study such as this (Ewell et al., 2010). These emerging lessons will also be useful in setting the direction of future development towards a full AHELO survey and may also provide insights for others seeking to measure higher education learning outcomes in other contexts and settings.

It is worth recalling that, from the outset, the feasibility study was envisaged as a research or "proof of concept" exercise designed to shed light on issues of feasibility, and not as a field trial for a fully-fledged main survey. The findings from the feasibility study and these emerging lessons should be seen within this spirit of exploration and learning by doing. Furthermore, while the feasibility study has served its purpose well, it is also clear that a great deal of **further work would be needed to develop a full evaluation** of higher education outcomes at an international level. To a large extent, this reflects and confirms the expectations of the experts involved in the initial development stages of the feasibility study.

Purpose of AHELO – to support improvement in learning outcomes

A key lesson to emerge from the feasibility study is the **importance of clearly establishing the purpose of AHELO**. The design of the feasibility study embodied a specific purpose, namely to provide Higher Education Institutions with a voluntary international comparative assessment of the overall performance of their students, which they can use to support and foster improvement in learning outcomes. But, as illustrated in Chapter 2, there remained considerable concerns among stakeholders about how AHELO would, or could, be used for accountability purposes, whether by linking to financing arrangements or through the development of league tables. Over the course of the feasibility study AHELO appears to have generated a range of expectations and also some anxieties about what it could be intended to do, notwithstanding the extensive communication efforts employed. With the benefit of hindsight, more could have been done to **communicate more clearly what the purpose of AHELO was *and* what it was not.**

© OECD 2012

Looking ahead to possible future development, it would be important to re-emphasise that AHELO is intended as a tool for Higher Education Institutions to support improvement in learning outcomes and, given the concerns expressed by some stakeholders, it may be necessary to **examine whether some form of safeguards might be needed to ensure that AHELO remains true to purpose and is not misused.** Such reassurance would also probably help to encourage more Higher Education Institutions to participate.

Feasibility study design

Learning outcomes are produced through a higher education system that is complex and inter-connected, with many different stakeholders. Furthermore each country has its own unique arrangements, history and context and most are seeing greater diversity emerging as the student body has expanded and society's expectations of higher education have increased. Higher Education Institutions are largely autonomous and are increasingly diverse - some institutions offer an increasingly wide range of study options while others are more specialised in meeting specific needs. In some countries differentiation is supported by policy choices (e.g. separate arrangements for Universities and Polytechnics) while in others, differentiation emerges as a result of institutions' choices and priorities.

Diversity

The design of the feasibility study deliberately sought to confront the challenges associated with diversity and, as more countries joined, the final set of field implementations encompassed even more diversity than originally planned. Overall, this **diversity proved to be a source of added richness** to the feasibility study in a number of ways, not least by stimulating close scrutiny of differences in expectations about learning outcomes across countries and across institutional settings. Efforts to bring together diverse experts to define learning outcomes and to develop the instruments paid off although it remains to be seen from the data analysis whether the process of adapting instruments to local contexts and languages went far enough.

Higher Education Institutions as the unit of analysis

The feasibility study chose the Higher Education Institution as the unit of analysis. While this proved to be a reasonable approach, it did lead to a number of unexpected challenges. Although the volunteering institutions participating in the feasibility study seemed to be highly motivated (and other institutions wanted to participate), it would be important to clarify what aspects and which data would be most useful and attractive to institutions. A **follow-up survey of participating institutions**, once they have received their institution reports from the consortium, would shed light on this and would inform future development. The views of institutions and how they would expect to be able to use the results to inform teaching and learning are likely to differ according to which strand they were in. Encouragingly, there was also a high degree of faculty interest, especially for Engineering and Economics. **However, the feasibility study was not designed to probe deeply into how institutions and faculty would use AHELO and a lesson for further development would be to give this aspect greater attention.**

© OECD 2012

Engaging key stakeholders and experts

Another general lesson concerns **the importance of seeking input from key stakeholders** and the study has clearly affirmed the value of consultative processes. Design of the feasibility study was significantly shaped not only by a wide cross-section of experts but also by detailed and constructive advice from Education International, initially, and other members of the Stakeholders Consultative Group. These contributions brought invaluable perspectives and any future development would only be enhanced by even greater involvement of all stakeholder communities as well as drawing input and ideas from a broader range of participating countries.

Assessment strands

A key feature of the feasibility study design was the decision to adopt three separate strands of assessment for Generic Skills, Economics and Engineering, supported by common contextual dimension surveys. This was recognised at the time as an artificial distinction that would yield additional insights into the relative merits and drawbacks of different domains and was not intended to serve as an "either/or" test of two alternative approaches. Nevertheless, in many ways, the discipline strands have proved more straightforward to implement than the Generic Skills strand. However, **drawing any conclusions about the relative merits of testing generic skills *versus* discipline skills *per se* is complicated** by the conscious decision to adapt an existing instrument for testing generic skills in the feasibility study, while recognising that an instrument would need to be designed from scratch for a future AHELO Survey (see Chapter 3).

It is worth noting that generic skills are now attracting more attention in the public discourse than was the case when AHELO was first being considered (OECD, 2012), which suggests that measuring generic skills effectively deserves even more attention in the future. Indeed, **choices about the relative importance of developing the discipline strands *versus* generic skills strands in future should** not hinge on the performance of the specific instruments used in the feasibility study, but **reflect instead what are the learning outcome objectives for different Higher Education Institutions and how institutions would want to use the results for improvement in teaching and learning.**

Management of the feasibility study

Management of the AHELO feasibility study presented a number of challenges, notwithstanding the OECD's experience with developing large scale surveys, such as PISA, TALIS and the OECD Survey of Adult Skills (PIAAC). In many respects those surveys are more straightforward than AHELO, not least because they are sample-based surveys with well-defined target populations and delivery channels.[1] In contrast, AHELO involves engaging with largely autonomous Higher Education Institutions and persuading them to participate. Furthermore, the learning outcomes that AHELO seeks to measure are both more diverse and more sophisticated than the foundation skills of literacy and numeracy that are at the heart of most international large-scale assessment surveys.

© OECD 2012

Funding

Uncertainties about funding were a key factor significantly impinging on the effective management of the feasibility study. While it is impossible to know whether the funding model's initial heavy reliance on grants from foundations might have been successful in a more propitious economic climate, a large investment of time and effort in attempting to raise funds from foundations bore relatively little fruit, while the funding difficulties the project faced made the whole study more challenging to manage. It led to some compromises on the scope of the study, wider (and therefore more complex) country participation and significant delays and adjustments to the study's timeline (see Chapter 3).

In the end, participating countries shouldered the bulk of the study's international costs, as well as their direct costs incurred within their country. And although the momentum was maintained, in accordance with the wishes of the AHELO Group of National Experts, the funding gap was eventually closed and most of the scope reductions were eventually restored, overall management of the study was much more complex and time-consuming as a result, while uncertainty and shifting timelines also made project management more complicated for participating countries and for the contractors.

A key lesson from this experience is **the importance of securing sufficient funding for future development at an early stage,** so as to permit development to proceed smoothly and according to plan. This **also requires early commitment from a critical mass of countries to warrant moving forward** with further development, not least because it is otherwise difficult to develop a meaningful international assessment. It also seems unrealistic, at least in the current economic climate, to expect to secure further significant funding from foundations. Thus, any further development of AHELO will need to be largely funded by countries for the foreseeable future. In the longer term, a model involving some funding from participating Higher Education Institutions might be conceivable, but this would probably discourage institutions from participating and in effect, scuttle the whole venture if instituted before a critical mass of participating institutions had been reached.

Contracting arrangements

Another management lesson from the experience concerns the contracting arrangements. As noted above, the shifting timelines and funding uncertainties resulted in less than ideal conditions for negotiating contracts with different parties, and in particular, with the AHELO Consortium, led by ACER, and with CAE. Managing these contracts also proved very time-consuming for the OECD Secretariat, which might have been avoided had the contracting arrangements been more straight-forward. One area where the **contracts would have benefited from greater clarity was on communication channels**, as the importance of clear lines of communication between everyone involved, including those in participating countries, became increasingly evident as time went by, and particularly in the implementation phase.

A key feature of the study was the establishment of the AHELO Consortium (see Chapter 3). This approach was intended to bring significant synergies, additional expertise and benefits to the study and it was done in full recognition that some of the Consortium partners might have longer-term conflicting commercial interests. While, in many ways, this approach worked

© OECD 2012

extremely well, it did rely on effective collaboration between the different partners within the Consortium. In the event, the relationship within the Consortium between ACER as leader and the CAE as a sub-contractor was not always smooth. With the benefit of hindsight, this situation was probably not made easier by the fact that OECD had also contracted separately with CAE for the adaptation of the CLA instrument, leading to some blurred reporting lines, especially around the transition from instrumentation to fieldwork. More generally, this underlines **the importance of ensuring that all contracting arrangements are as clear and straight-forward as possible and that the most comprehensive tendering processes are used (even for sole supplier situations).**

Timelines

Another lesson on management of the study is to **allow sufficient time to prepare the extensive documentation and undertake the training** that is needed to orchestrate and support such a complex international assessment across many different actors and roles. A significant investment in developing and testing implementation systems both internationally and within countries was also required, particularly as this was the world's first major attempt at an entirely on-line international assessment. The timelines for this preparatory work were often very tight, in part due to shifting project timelines, and some countries had difficulty in meeting all the deadlines especially where completely new arrangements had to be set up (see below).

Instrument development

As noted above, the decision to adopt three separate strands of assessment for Generic Skills, Economics and Engineering, was recognised at the time as an artificial distinction designed to yield additional insights into the relative merits and drawbacks of different domains. A major challenge for the feasibility study was to demonstrate that an assessment framework could be agreed upon across diverse country and institution settings and a key focus in the instrument development stage was adapting to different national, cultural and linguistic settings. This was a challenging process **but also one that led to rich exchanges of experience and important feedback that strengthened the entire instrument development process**. It also meant that **good quality control and quality assurance procedures were even more important** than for within-country testing.

A separate step was to establish, through field implementation, whether the specific instruments that were developed to assess students and tested in the field would yield reliable and valid results. This latter question will be addressed by psychometric analysis to be presented in Volume 2 but it is important to note that that **this analysis will only pertain to the specific instruments that were tested and cannot be generalised.** In other words, they will not be able to predict the validity or reliability of other instruments that could have been developed from the same assessment framework. It is also worth recalling in this context that because the feasibility study was designed to provide "proof of concept", it was generally accepted that it was not necessary to develop perfect and exhaustive instruments (see Chapter 4).

© OECD 2012

Assessment frameworks

A decision was made early on to adapt the Collegiate Learning Assessment instrument for the Generic Skills strand. However, as the process of adapting this instrument to an international context evolved and multi-choice questions were added to the assessment, it became evident that a Generic Skills Assessment Framework was needed to guide the work. However, because this framework was developed late in the development process, there was insufficient time to work through all the aspects in depth and it was not possible for an international group of experts to reach full consensus. While the relationships between generic skills and economic and social outcomes have become generally accepted, there is little international consensus about exactly what generic skills are and, in the context of higher education, their connection with professional, cultural or disciplinary contexts. With the benefit of hindsight, while adapting an existing instrument allowed some time to be gained, the downside is that it may have inhibited the process of establishing a robust international framework on Generic Skills to underpin the assessment. In part as a result of this, some aspects of the Generic Skills strand seemed to remain a matter of contention among experts throughout the feasibility study.

Two key lessons can be drawn from this experience. The first is the importance and value of **establishing international consensus on the assessment framework as an essential upstream part of the instrument development process.** The second is that **what might have seemed at the time as a reasonable short-cut may not have been the most cost-effective approach after all**. This underlines the importance of developing completely new tailor-made instruments for any future AHELO. This approach also avoids complex intellectual property issues that can arise when adapting an existing instrument.

In contrast, the development of assessment frameworks for the Economics and Engineering strands proceeded smoothly and showed that it is possible to define discipline specific learning outcomes internationally. While this was expected for the Engineering strand, there were some doubts at the outset whether it would be possible for a social science such as Economics. **In fact, it proved easier than expected to get agreement amongst the Economics experts, on what AHELO should cover and measure**.

Matrix sampling

Some more general lessons can be drawn about other aspects of the instrument design for the assessment strands. A key element that emerged from the implementation phase was the institutions' desire to provide students with feedback on their performance, in part to motivate them to participate. However the use of matrix sampling (i.e. randomly assigning different sections of the assessment to students) precludes this. **One thing this experience suggests is that the trade-off between the advantages of matrix sampling and being able to provide students with feedback should be reconsidered.** Having a small common core of multi-choice question items that could be administered to all students irrespective of item rotations might be one way to manage this trade-off. However, the time required to complete the tests (120 minutes in the disciplinary strands and 150 minutes for the generic skills strand) is already long and may have had some impact on student response rates, although further analysis of the AHELO feasibility study data would be needed to shed more light on this.

© OECD 2012

Multiple-choice questions and constructed response tasks

All the assessment instruments combined multiple-choice questions and constructed response tasks. A further lesson is to **consider more deeply the relative appeal of these two different approaches** for students (see for example, the feedback from Mexico's students in Chapter 4) and for faculty and the impact this might have both on the willingness to engage in AHELO and on the quality of responses (including effort). Finding the best combination of these two approaches also needs to be linked more clearly to what the institutions want to get out of the exercise and how faculty would want to use the results.

Contextual dimension

A well-designed contextual dimension is critical to any future AHELO. It is not only important for instrument development and calibration and the scaling of scores, but it is also essential to interpret results in context and to be able to extract formative lessons from the results. However, further analysis of the AHELO feasibility study contextual data would be needed to identify which contextual dimensions and variables add most value to a formative use of the AHELO data. Getting the contextual dimension right also needs to involve careful consideration of the purpose and uses intended for a future AHELO and extensive consultations with stakeholders.

Instrument localisation

Instrumentation localisation processes are critical to AHELO as an international assessment. Experience from the feasibility study suggests that the processes followed worked well overall. Nonetheless, several lessons can be drawn. *First,* it is important to **allow sufficient time for high quality translation** as it turned out to be a more challenging and time-consuming task than initially expected. *Second,* it is important to **get the sequencing right**, in particular reviewing all the test materials fully before embarking on translation and adaptation activities. *Third*, it would be useful to **involve domain experts early** in the translation process.

It is also worth noting that the Adaptation Translation and Verification (ATAV) system (see Chapter 4) proved to be a very useful support system. For the future, it could be valuable to **put all the key manuals to be used within countries through a more formalized ATAV process** as well. It would also be informative to undertake some **ex-post reviews of the quality of national translations** and adaptations for formative purposes at country level, e.g. through an in-depth examination of the AHELO data (country-DIF) to identify possible translation errors at country level.

Field implementation

Overall, field implementation proceeded very well, despite the challenges arising from the uncertainties of the project and shifting timelines, although it will be useful to review national and institutional experiences in more detail to identify best practices which could be used as models for any future AHELO. However, a number of lessons can already be drawn out of the feasibility study experience.

© OECD 2012

Timelines

A first lesson was that the timelines were generally too compressed (as noted above) and would need to **be extended for national and institutional activities in any future AHELO.** More generally, the compromise of undertaking all field implementation within a relatively narrow period meant that the timing of activities was sometimes inappropriate in relation to academic calendars (holidays, exam periods, students on internships…). The complications that this timing imposed on field implementation suggests that a future AHELO development should **endeavour to time the assessment to fit optimally within each country's academic year.** Ideally, confirming an institution's participation in an AHELO Survey before the start of the academic year would allow it to be more effectively integrated with other relevant activities and demands on students and faculty.

Capacity, readiness and quality control

For many countries participating in the feasibility study meant establishing entirely new structures and processes to be able to carry out the field implementation, while participating in AHELO was an entirely new process for participating institutions. This presented many capacity challenges and the fieldwork benefited from a deep commitment to making it work among participating countries. In the context of the feasibility study, it was well understood that everyone was "learning by doing" and support arrangements provided by the Consortium (manuals, training, phone assistance, etc.) played a crucial and effective role in helping build the capacity to undertake the field implementation. Nonetheless, there were some procedures that were not fully complied with in some countries and institutions and strict adherence to prescribed processes was somewhat less important in the feasibility study due to the formative nature of the exercise.

For the future, **participation readiness criteria could be drawn up, to ensure that participating countries and institutions can identify whether they have all the systems and mechanisms in place to be able to follow prescribed processes**. Indeed, if countries or institutions are unable to meet the readiness criteria (e.g. with having the data systems in place to make it possible to draw a sampling frame), then it would probably make more sense for them to invest in building that capacity rather than participating in an AHELO assessment which would produce unreliable results.

A related lesson is the importance of **establishing a strong external quality control function** to monitor the work of contractors as well as participants' adherence to prescribed processes. External quality assurance processes are important to ensure that the study is implemented in a standardized way across countries and institutions, which is a pre-requisite for the cross-national and cross-institutional comparability of the final results. Again, while this was less of an issue for the feasibility study, as a learning experience focusing on "proof of concept", it would be crucial for the credibility of the results of a future AHELO full survey.

Of course, any future AHELO would need to include a field trial to test out both the instruments and the procedures in real-life conditions and identify any practical issues ahead of the full survey. This would also provide a crucial opportunity to test both participants' readiness and external quality control procedures.

© OECD 2012

Electronic delivery

A major success of the feasibility study was the use of electronic delivery. Not only is electronic delivery important for efficient management and processing of the instruments, but it also opens up future opportunities for analysing how students responded (e.g. more or less quickly) as a way of refining the instruments themselves. Nevertheless, while the **use of two different platforms** for the Generic Skills strand did not generate particular problems for delivering the assessments, it **did add a layer of complexity and costs** associated with integrating the two platforms.

A range of technical glitches were encountered during the delivery, although most were resolved very quickly. While this is to some extent to be expected within a feasibility study implemented across many different IT environments, and particularly with a compressed timeline, it underscores the importance of **allowing sufficient time for preparation and testing beforehand** under different operating systems and conditions. **Readiness criteria** (as suggested above) for any future AHELO could include these technical aspects as well, drawing fully on the experiences and challenges encountered during the feasibility study.

One approach worth further consideration could build on the experience of the Slovak Republic (see Chapter 5), which used an innovative approach to minimise the extent of IT preparations and support needed at institutional level by providing a set of dedicated laptops that were transported from institution to institution. This not only minimised their exposure to IT risk, it also provided a different way of managing test security. While such a particular approach might not be feasible in a full survey or in other country contexts, it nevertheless points to the **value of exploring alternative approaches to electronic delivery** and some cost-benefit analysis would shed light on the tradeoffs involved.

Sampling frames and student response rates

The timely provision of an appropriately detailed and verified unit-record student frame is a critical element of a sample-based approach to ensure the integrity of study outcomes and results and to allow the estimation of confidence intervals. Yet the feasibility study showed that most countries and institutions found this a challenging task and **clearer definitions and better protocols for developing sampling frames and managing exclusions were needed.** National Project Managers suggested that a pre-prepared data template for listing students and faculty to assist the construction of sampling frames as well as more time to complete all sampling steps would have helped (Brese and Daniel, 2012).

One of the biggest challenges in delivery was obtaining adequate response rates and despite countries' best efforts, the field implementation yielded a wide range of experiences, some of which were disappointing to countries, leading some National Project Managers to **call into question the overall approach to sampling students.**

Two possible alternative approaches on the student sampling frames were suggested. The first is to reject sampling in favour of a census approach for all in-scope students, since where a census approach was taken, response rates were typically much higher. A census approach would also streamline – to some extent – the work involved in preparing and drawing the

© OECD 2012

sample although it would add to scoring workload and costs. However, while this would seem a straightforward option to consider in a discipline-specific context, it would be less easy to apply to testing generic skills across an entire institution. The second would be to consider drawing sample groups of students (e.g. classes or courses) rather than taking a random selection of students within the HEI. Such an approach would facilitate communication on AHELO and the organization of testing which could be incorporated into the course programme. However, while the operational advantages are obvious, further analysis would be required to ensure the effectiveness of such a sampling approach (Brese and Daniel, 2012).

Countries and institutions employed a wide range of approaches to engage students and raise response rates. Experience from the AHELO feasibility study also showed that the nature and impact of participation incentives offered to students varied widely. But it was not possible to detect any clear link with engagement or response – with the possible exception of institutional strategies that incorporated the AHELO assessment as part of their curriculum. Reflecting on the experience, suggestions that were put forward included:

- administering the assessments as part of university courses/classes
- making participation in AHELO a compulsory part of the university curriculum or linked with broader discussions on learning outcomes and expectations from employers
- shortening the length of the assessment
- redesigning the instruments to provide feedback to students on their own performance both individually and in comparison with other students
- providing students with some form of academic benefits to reward participation – though not performance
- improving the scheduling and timing of the AHELO assessment

While a number of different suggestions and perspectives were offered, no agreement was reached or even sought to resolve these perspectives should AHELO move forward. But, at the very least, it is clear that **further research is needed into what measures would be most effective in raising response rates** in further development of AHELO. Any future AHELO would also require an agreed-upon **policy on incentive structures** to clarify which incentives are appropriate and limit the risk of potential bias.

Scoring

Managing the scoring process by and large worked well and countries invested considerable time and effort into training scorers and endeavouring to increase the reliability and consistency of the scoring process. It also appears that small groups of scorers worked most efficiently. This aspect of the feasibility study perhaps worked so well because procedures for quality in scoring are very well-established in other contexts both within many institutions and at national levels in other contexts (such as in national examinations or international tests such as PISA). However, many participating countries felt that more **time for training and**

© OECD 2012

preparation would have been useful. **Tools and training to support closer monitoring of inter-scorer reliability** would also have been useful and in a future AHELO, more emphasis could be given to these aspects of scoring procedures, including cross-country equivalency.

Estimating costs of participation

Participation in AHELO clearly involved significant costs for participating countries and institutions in the field implementation stage, over and above the contributions countries made to the OECD for the international costs of the feasibility study. While some cost data has been collected by the Consortium and will be analysed, it is not currently possible to provide a full estimate of all the costs incurred. More complete assessments of the costs involved would have been possible if more systematic monitoring of costs within countries had been built into the project from the outset. While the full costs incurred at in the feasibility study would not necessarily be a good guide to the possible future costs for countries of participating in a future AHELO survey, not least because of adaptations made in light of this experience, more complete data about costs would have been useful to identify to carry out some cost-benefit analyses of different study aspects.

Extra value for participating institutions and faculties

A number of countries felt they got an extra bonus out of participating in the feasibility study, because of the deep reflections that this induced about teaching and learning, as the three country examples in Box 6.1 show. Their experience points to perhaps the most important lesson to be drawn from the feasibility study: that the assessment of higher education outcomes is not an end in itself, but rather a stimulus to deeper professional dialogue on desired learning outcomes and the teaching approaches needed to achieve them. These experiences prove that this has indeed been achieved and point to the potential benefits to be gained from taking AHELO development further as a tool to foster improvement in learning outcomes.

Box 6.1 - Value to participating institutions and faculties

Japan - Eye opening

Japan has been involved in the OECD-AHELO Feasibility Study from the very early stages. From our five years of commitment, we have indeed learned much and we would like to highlight the potential value of AHELO to institutions and faculties.

As a member of the AHELO consortium, NIER participated in instrument and scoring rubric development and modification for the engineering strand. Through this experience of international collaboration, we came to understand tangibly and substantively that there is a conceptual framework of engineering competencies and learning outcomes that can be shared globally. Although there are limitations and precautions, we believe that in the field of engineering, we can achieve with deliberation instruments and scoring rubrics that are valid for institutions and faculties around the world.

As the National Center, NIER undertook translation, instrument implementation, scorer training,

© OECD 2012

and scoring. The scorer training and scoring exercise proved to be eye openers for our 13 professors who gathered to mark the student responses. Because scoring requires consensus on what kinds of responses can be identified as correct, it forces scorers to define precisely the scope and level of learning outcomes that students are expected to demonstrate. This exercise was instrumental in generating clearer understandings of the conceptual framework of competencies, and encouraged critical reflections on teaching practices.

NIER, Japan

Canada Changing the way they teach

Ontario universities were excited to contribute to the ideas and possibilities of AHELO – describing it as a timely international conversation on learning outcomes assessment. Their interest was based on the recognition that our campuses are increasingly made up of internationally engaged, and often mobile students and faculty members; and that participating in this project would improve global understandings of compatibility and comparability of teaching and learning

Faculties of Engineering were interested in knowing how their students' knowledge and skills compared to those of other institutions and countries. For example, how would their students fare in generic engineering questions or in areas of practical application? Deans and department chairs were particularly interested in how this information could demonstrate the nature and strengths of their curriculum design and how it might inform curriculum change.

Faculty members were further encouraged by the potential of AHELO when a small group came together to score the constructed responses of the students. The framing of questions to make students 'think like an engineer' was innovative to some – prompting them to consider how they taught their students, what they expected of them and how they were assessed. "I wish I had considered these issues before setting the final exam for my students," noted one professor during the scoring session, "I would have done it much differently."

NPM for Canada (Ontario)

Kuwait - Fostering a focus on assessment

The overall experience of participating institutions has been valuable in that it has both focused and expanded the conversation on assessment in areas of:

- developing measurable mechanisms that can provide the optimal means by which student learning outcomes can be determined;
- standardising these means across compatible institutions and cross-border;
- developing benchmarks against which institutions are measured;
- internationalising expectations with respect to student learning outcomes; and
- internationalising the benchmark that determines quality institutions, quality instruction, and quality learning.

NPM for Kuwait

© OECD 2012

REFERENCES

Brese, F. and T. Daniel (2012), OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study: Report on Quality Adherence, IEA Data Processing and Research Center, Hamburg.

Ewell P., V. D'Andrea, P. Holland, K. Rust and F. Van Vught (2010), AHELO TAG Report, December 2010.

OECD (2012), *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*, OECD Publishing.

© OECD 2012

NOTES

1 In the case of PISA and TALIS, these can be managed through well-established procedures within the schooling system and for the Survey of Adult Skills through established household survey approaches.

© OECD 2012

ANNEX A

THE DEVELOPMENT AND IMPLEMENTATION OF THE AHELO FEASIBILITY STUDY: A COLLABORATIVE EFFORT

Introduction

The AHELO feasibility study would not have been possible without the enthusiasm and support of the 17 participating countries/higher education systems from all continents who elicited to be part of this ground-breaking endeavour, namely Abu Dhabi, Australia, Belgium (Flanders), Canada (Ontario), Colombia, Egypt, Finland, Italy, Japan, Korea, Kuwait, Mexico, the Netherlands, Norway, the Russian Federation, the Slovak Republic and the United States (Connecticut, Missouri and Pennsylvania).

Neither would the AHELO feasibility study have taken off the ground without the foresight and generosity of its sponsors, whose commitment and investment allowed this work to prove its value and importance. The OECD is extremely grateful for their contributions to the AHELO effort.

Notwithstanding the level of enthusiasm and generosity from participating countries/systems and sponsors, the success of the AHELO feasibility study also rested with the 249 forward-looking higher education institutions (HEIs) who volunteered to take part in the AHELO assessments and accepted the risk that the results might be challenging them in their teaching practices. The OECD is appreciative of their open-mindedness as well as the efforts they have put in the study to adhere to prescribed procedures and standards, and is hopeful that the results will be valuable to them in their improvement efforts.

The study has been a collaborative effort, bringing together scientific expertise from a wide range of countries and backgrounds.

The study benefited from a unique governance structure within the OECD insofar as it was jointly steered by governments, HEIs and higher education agencies through the Programme on Institutional Management in Higher Education (IMHE). The IMHE platform ensured that the approaches adopted took account of institutional needs and concerns. The IMHE Governing Board (IMHE GB) was responsible for the broader management of the AHELO feasibility study.

To support the IMHE GB, an AHELO Group of National Experts (GNE) was established to oversee the methods, timing and principles of the AHELO feasibility study. The GNE brought together experts nominated by participating and other interested countries and was the main steering mechanism for the technical aspects of the feasibility study.

© OECD 2012

Through National Project Managers (NPMs), participating countries implemented the AHELO feasibility study at the national level in accordance with agreed procedures. NPMs played a vital role in ensuring that the implementation of the survey is of high quality.

The study also drew on the best internationally available expertise at different stages of the project and in a range of domain areas through a number of ad-hoc expert groups. In the initial stages of reflection, three expert groups were convened in 2007 to reflect on the desirability and feasibility of an international Assessment of Higher Education Learning Outcomes (AHELO), and to provide guidance for a strategy. In 2008 and 2009, another three expert groups were tasked to provide input towards the development of terms of reference for the AHELO Call for Tenders. Two expert groups adapted the Tuning approach for AHELO to identify expected learning outcomes in economics and engineering - the disciplines chosen for the feasibility study - while the third group provided recommendations towards the development of a contextual dimension for the feasibility study. In September 2009, a Technical Review Panel was also set up to review the technical aspects of proposals submitted in response to the AHELO Call for Tenders, and prepare recommendations for the AHELO GNE. Subsequently, another three expert groups were set up by the chosen contractor to develop assessment frameworks and instruments: the AHELO Economics Expert Group, AHELO Engineering Expert Group and AHELO Contextual Dimension Expert Group. These groups met in 2010 and 2011 and experts from participating countries served on the discipline-specific groups to ensure that the instruments developed are internationally valid and reflect the diverse cultural and curricular contexts of participating countries.

An AHELO Technical Advisory Group (TAG) was established in 2010 to provide quality oversight throughout the study. It was consulted and provided advice on major technical issues (e.g. instrument development procedures, sample and test design, translation procedures, as well as scoring and verification procedures) and operational aspects.

In addition, a Stakeholders' Consultative Group (SCG) was set up, bringing together diverse stakeholders with an interest in the quality of higher education. While the SCG did not have a formal steering role in the feasibility study, it proved highly valuable in communicating the goals and progress of the AHELO feasibility study; listening to stakeholders' input, suggestions, concerns, warnings or advice; and providing a forum for multilateral discussion and cross-fertilisation of ideas on the AHELO initiative.

The design and implementation of the assessments, within the framework established by the AHELO GNE, was the responsibility of external contractors grouped into an international Consortium[1] led by the Australian Consortium for Education Research (ACER). The development of the Generic Skills instrument was undertaken by the Council for Aid to Education (CAE). Educational Testing Services (ETS) took the lead in developing the economics instrument while the development of the engineering instrument was carried out by a consortium led by ACER in partnership with Japan's National Institute for Education Research (NIER) and the University of Florence. Finally, the contextual dimension survey instruments were developed jointly between Indiana's University Center for Postsecondary Research (CPR) and the Centre for Higher Education Policy Studies (CHEPS). Linguistic quality control was under the responsibility of cApStAn for instruments developed by the Consortium and Comms Multilingual for the Generic

© OECD 2012

Skills strand performance tasks, while Statistics Canada was responsible for sampling and weighting. The test platforms were developed by SoNET in co-operation with ITS for the Generic Skills strand. Other partners in this consortium include the International Association for the Evaluation of Educational Achievement (IEA) and Westat.

The OECD Secretariat was responsible for the overall management of the AHELO feasibility study and the coordination of these various groups under the guidance of the AHELO GNE, and in line with the policy objectives established by the IMHE GB.

The following lists the members of the various AHELO bodies and the individual experts and consultants who have contributed to the AHELO feasibility study.

© OECD 2012

Sponsors

Lumina Foundation for Education
Compagnia di San Paolo
Calouste Gulbenkian Foundation
Riksbankens Jubileumsfond
Spencer Foundation
Teagle Foundation
Higher Education Founding Council for England
Ireland Higher Education Authority
The William and Flora Hewlett Foundation (support for US participation)

IMHE Governing Board

Chair: Prof. Peter Coaldrake (Queensland University of Technology - Australia)

Deputy Chairs: Tom Boland (Higher Education Authority, Ireland), Steven Egan (HEFCE - United Kingdom), Esa Hämäläinen (University of Helsinki - Finland), Åsa Petri (Ministry of Education and Research - Sweden) and Giuseppe Ronsisvalle (University of Catania - Italy).

OECD Member Countries

Australia: Peter Coaldrake, Queensland University of Technology (*or John Hearn, University of Sydney),* and David De Carvalho, Department of Industry, Innovation, Science, Research, Tertiary Education, *(or Shelagh Whittleston, Australian Delegation to the OECD)*
Austria: Christiane Druml, Medical University of Vienna (*or Andreas Raggautz, University of Graz)* and Florian Pecenka, Science Attaché, Permanent Representative of Austria to the EU
Belgium: Benjamin Van Camp, Flemish Interuniversity Council (*or Kristiaan Verslys, Ghent University)*
Canada: Christine Tausig Ford, Association of Universities and Colleges of Canada (AUCC)
Chile: Patricio Sanhueza Vivanco, Universidad de Playa Ancha
Czech Republic: Stanislav Stech, Charles University and Karolina Gondkova, Ministry of Education, Youth and Sports (*or Adam Krcal*, *Ministry of Education, Youth and Sports)*
Denmark: Peter Plenge, Aalborg University and Birgit Kjolby, Ministry of Science, Innovation and Higher Education *(or Mette Holmstrom Mikkelsen, Ministry of Science, Innovation and Higher Education)*
Estonia: Kristjan Haller, University of Tartu (or Karin Jaanson, Ministry of Education and Research)
Finland: Esa Hämäläinen, University of Helsinki (or Ritva Laakso Manninen, Haaga-Helia University of Applied Sciences) and Reijo Aholainen, Ministry of Education (or Maarit Palonen, Ministry of Education)
France: Sebastien Lobiau, Ministère de l'éducation nationale / Ministère de l'enseignement supérieur et de la recherche (*or Laurence Malpot / Regine Ducos Ecole des hautes études en santé publique - EHESP, Rennes*)
Germany: Marijke Wahlers, German Rectors' Conference (or Sybille Hinze, Institute for Research Information and Quality Assurance)

© OECD 2012

Greece: Ekaterini Douka-Kabitoglou, Aristotle University Thessaloniki and Ministry of Education and Vassilios Tourassis, Democritus University of Thrace, (*or Pandelis Ipsilandis, Technological Education Institute of Larissa*)
Hungary: Béla Kardon, Ministry of National Resources
Iceland: Ingjaldur Hannibalsson, University of Iceland
Ireland: Tom Boland, Higher Education Authority, Dublin (*or Brendan Murply, Cork Institute of Technology*)
Israel: Moshe Vigdor, Council for Higher Education and the Planning and Budgeting Committee(*or Nachum Finger, Council for Higher Education*)
Italy: Giuseppe Ronsisvalle, University of Catania and Marzia Foroni, Ministero dell'Istruzione, dell'Università e della Ricerca
Japan: Shuichi Tsukahara, National Institute for Education Policy Research (NIER)
Korea: Hyunsook Yu, Korean Educational Development Institute
Mexico: Fernando Leon Garcia, CETYS University (*or Enrique Del Val, Universidad Nacional Autónoma de México - UNAM*) and Luz Maria Nieto Caraveo, Universidad Autonoma de San Luis Potosi - representing the Secretaria de Educación Publica of Mexico (*or Carlos Quintal Franco, Universidad Autonoma de Yucatan - representing the Secretaria de Educación Publica of Mexico*)
Netherlands: Marlies Leegwater, Ministry of Education, Culture and Science
New Zealand: Roger Smyth, New Zealand Ministry of Education
Norway: Lars Vasbotten, Norwegian Ministry of Education and Research (*or Ingvild Marheim Larsen, Norwegian Ministry of Education and Research*) and Gunn-Elin Aa Bjorneboe, University of Oslo (*or Per Ivar Maudal, Norwegian University of Science and Technology*)
Poland: Daria Nalecz, Ministry of Science and Higher Education
Portugal: Alberto Amaral, Agência de Avaliação e Acreditação do Ensino Superior (A3ES)
Slovak Republic: Peter Mederly, Ministry of Education, Science, Research and Sport
Slovenia: Duša Marjetic, Ministry of Higher Education, Science and Technology
Spain: Luis M. Delgado Martinez, Ministerio de la Educación (*or Angeles Muñoz Fernández de Bastida, Ministerio de la Educación*)
Sweden: Ann-Caroline Nordström, Stockholm University and Åsa Petri, Ministry of Education and Research, Stockholm (*or Per Magnusson, Division of for Higher Education, Ministry of Education and Research*)
Switzerland: Stéphane Berthet, Université de Genève and Müfit Sabo, Office fédéral de la formation professionnelle et de la technologie (*or Silvia Studinger, Secrétariat d'État à l'Éducation et à la Recherche*)
Turkey: Gokhan Cetinsaya, Council of Higher Education (YÖK)
United Kingdom: Steve Cannon, University of Aberdeen and Stephen Egan, HEFCE
United States: Molly Corbett Broad, American Council on Education

OECD non members / non-membres de l'OCDE

Brazil: Leandro Tessler, UNICAMP
Latvia: Janis Stonis, University of Latvia (*or Alina Grzhibovska, University of Latvia*)
Russia: Olga Moshkova, National Research University - Higher School of Economics *(or Lyubov Maximova National Research University - Higher School of Economics)*

© OECD 2012

AHELO Group of National Experts (GNE)

Chair: Jan S. Levy (Norway)
Bureau: Fiorella Kostoris (Italy), Ryo Watanabe (Japan), Patricia Rosas Chávez (Mexico) and Adrie Steenbrink (Netherlands).

Members and Observers for the participating countries

Australia: Margaret Pearce (Australian Delegation to the OECD) and Philip Aungles (Australian Department of Education, Employment and Workplace Relations)
Abu Dhabi: Hesham Wagih Gomma and Sylvie Wald (Abu Dhabi Education Council - ADEC)
Belgium: Noël Vercruysse (Department of Education and Training, Flemish Ministry for Education and Training), Eduardo Cascallar (Center for Research on Teaching and Training - POOLL) and Filip Abraham (Katholieke Universiteit Leuven)
Canada (Ontario): Mary Catharine Lennon and Harvey Weingarten (Higher Education Quality Council of Ontario - HEQCO)
Canada: Amanda Hodgkinson and Natalia Ronda (Council of Ministers of Education, Canada - CMEC)
Colombia: Julian Patricio Mariño von Hildebrand and Margarita Peña Borrero (ICFES)
Egypt: Ibrahim Shehatta (Egyptian National Center for Measurement and Assessment Project); Galal Abdelhamid Abdellah, Samir Hilal and Rasha Sharaf (Egyptian Ministry of Higher Education), Sanaa Radi and Ghada Mourad (Alexandria University); Ali Huzayyin, Ahmed Soliman and Mohamed El-Manylawi (Cairo University).
Finland: Ossi Tuomi (University of Helsinki), Maarit Palonen (Ministry of Education and Culture) and Jani Ursin (University of Jyväskylä)
Italy: Fiorella Kostoris (Universitá di Roma)
Japan: Ryo Watanabe, Osamu Aruga, Hiroshi Ando and Mizuki Tanaka (Ministry of Education, Culture, Sports, Science and Technology - MEXT)
Republic of Korea: Jeung-Yun Choi (Korean Educational Development Institute)
Kuwait: Asad Al-Rashed (Private Universities Council)
Mexico: Luz María Nieto Caraveo (Universidad Autónoma de San Luis Potosí), Patricia Rosas Chávez (Universidad de Guadalajara)
Netherlands: Adrie Steenbrink (Ministry of Education Culture and Science), Frits Gronsveld (Verdonck, Klooster & Associates), Henriette Maassen van den Brink (Amsterdam School of Economics), Inez Meurs (Ministry of Education, Culture & Sciences / HU Business School Utrecht)
Norway: Jan S. Levy (Norwegian Ministry of Education and Research), Elisabeth Hovdhaugen and Vibeke Opheim (NIFU Nordic Institute for Studies in Innovation, Research and Education), Ole K. Solbjørg (NTNU)
Russian Federation: Tatiana Meshkova (National Research University - Higher School of Economics)
Slovakia: Peter Mederly (Ministry of Education, Science, Research and Sport)
United States: Charles Lenth and Paul Lingenfelter (State Higher Education Executive Officers - SHEEO), Andre Lewis and Frank Frankfort (US Department of Education)

© OECD 2012

Observer countries (including OECD member countries not participating in AHELO and official non-OECD Observers)

Austria: Florian Pecenka (Permanent Representation of Austria to the EU) and Eva Schacherbauer (Federal Ministry of Science and Research)
Bahrain: Hesham Al-Ammal and Ebrahim Janahi (University of Bahrain)
Brazil: Sheyla Carvalho Lira (Ministério da Educação)
Denmark: Mette Juul Jensen (Ministry of Science, Innovation and Higher Education)
England: William Locke (Higher Education Funding Council for England), Matthew Bollington, Brian Johnson and Charles Ritchie (Department for Business Innovation & Skills - BIS)
Germany: Martina Diegelmann (Bundesministerium für Bildung und Forschung), Stefan Hornbostel (Institut für Forschungsinformation und Qualitätssicherung / Institute for Research Information and Quality Assurance - IFQ)
Portugal: Isabel Correia (General Secretariat of Ministry of Education and Science), Sérgio Machado dos Santos (A3ES - Agência de Avaliação e Acreditação do Ensino Superior)
Saudi Arabia: Hamad Almehrej (Ministry of Higher Education)
Singapore: Wei Jian Leong, Wen Pin Leow and Jia Yu PAO (Ministry of Education)
Sweden: Asa Petri (Ministry of Education and Research), Per Gunnar Rosengren and Aija Sadurskis (Swedish National Agency for Education)
EU: Margarida Gameiro and Mads Gravas.

AHELO National Project Managers (NPM)

Abu Dhabi: Hesham Gomma and Sylvie Wald (Abu Dhabi Education Council - ADEC)
Australia: Daniel Edwards (Australian Council for Educational Research - ACER)
Belgium: Raoul Van Esbroeck (Vrije Universiteit Brussel)
Canada: Mary Catharine Lennon Higher Education Quality Council of Ontario (HEQCO)
Colombia: Julian Patricio Mariño von Hildebrand (ICFES), Maria Camila Perfetti
Egypt: Ibrahim Shehatta (Egyptian National Center for Measurement and Assessment Project)
Finland: Jani Ursin (University of Jyväskylä), Heidi Hyytinen (University of Helsinki)
Italy: Fiorella Kostoris (Universitá di Roma)
Japan: Shuichi Tsukahara (National Institute for Educational Policy Research - NIER)
Kuwait: Asad Al-Rashed (Private Universities Council), Anwar Almousawi (Kuwait University)
Mexico: Adrian De Leon-Arias (Universidad de Guadalajara), José Antonio González Fajardo (Universidad Autónoma de Yucatán), José Humberto Loriá Arcila (Autonomous University of Yucatan), Ruben Sanchez-Gomez, Patricia Rosas Chávez (Universidad de Guadalajara)
Netherlands: Inez Meurs (Ministry of Education, Culture & Sciences / HU Business School Utrecht)
Norway: Elisabeth Hovdhaugen, Vibeke Opheim and Rachel Sweetman (NIFU Nordic Institute for Studies in Innovation, Research and Education)
Republic of Korea: Jeung-Yun Choi (Korean Educational Development Institute)
Russian Federation: Tatiana Meshkova and Elena Sabelnikova (National Research University - Higher School of Economics), Oleg Rebrin and Irina Sholina (the Ural Federal University)
Slovakia: Jan Boda (Comenius University Bratislava), Ján Kalužný (Faculty of Materials Science of the Slovak University of Technology), Lenka Mackovicova, Peter Mederly (Ministry of

© OECD 2012

Education, Science, Research and Sport), Roman Nedela (Matej Bel University Banska Bystrica), Magdalena Nedelova
United States: Charles Lenth (State Higher Education Executive Officers - SHEEO)

Participating institutions and Institutional Coordinators (for countries where the information is not restricted)

Strand of participation is indicated in parenthesis.

Abu Dhabi (ENG)

Abu Dhabi University (ADU) - Aly Nazmy
ALHOSN University - Marko Savic
United Arab Emirates University (UAEU) - Civil and Environmental Engineering Department (CEED) - Amr S. El-Dieb
Scorers: Abu Dhabi University (ADU): Rafik Al Sakkaf - ALHOSN University: Osama Mohamed - United Arab Emirates University (UAEU) - Civil and Environmental Engineering Department (CEED): Khaled El-Sawy
Lead Scorer: United Arab Emirates University (UAEU) - Civil and Environmental Engineering Department (CEED): Bilal El-Ariss

Australia (ENG)

Charles Darwin University - Deborah West
RMIT University - Kevin Zang
James Cook University - Vincent Wang
Swinburne University of Technology - Emad Gad
University of Adelaide - Bernadette Foley
University of Melbourne - Graham Moore
University of Technology Sydney - Rob Jarman
University of Western Sydney - Fidelis Mashiri
University of New South Wales - David Clements
University of Newcastle - Bill McBride
Curtin University of Technology - David Scott

Belgium (ECO)

University of Gent - Luc Van De Poele
Katholieke Universiteit Leuven - Luc Sels

Canada (ENG)

McMaster University - Cameron Churchill
University of Ottawa - Dan Palermo
Queen's University at Kingston - Ken Novakowski
Ryerson University - Khaled Sennah
University of Toronto - Bob Andrews
University of Waterloo - David Brush
University of Western Ontario - Ernest Yanful

© OECD 2012

University of Windsor - Faouzi Ghrib
Carleton University - Yasser Hassan

Colombia (ENG, GS)

Universidad Nacional de Colombia - Sede Bogotá - Alba Neira (ENG, GS)
Universidad Nacional de Colombia - Sede Medellin - Alba Neira (ENG, GS)
UNIMINUTO - Corporación Universitaria Minuto de Dios - José Orlando Ugarte (ENG)
Escuela Colombiana de Ingenieria Julio Garavito - María Paulina Villegas (ENG)
Corporación Universitaria de la Costa - CUC - Alejandra Elguedo Pallares (ENG)
Fundación Universitaria Internacional del Trópico Americano - Hernando Salcedo (ENG)
Fundación Universitaria Agraria de Colombia - UNIAGRARIA - Carlos Omar Higuera Bautista (ENG)
Politécnico Colombiano Jaime Isaza Cadavid - Sandra Milena Cardona Galvis (ENG)
Corporación Universidad Piloto de Colombia - Sede del Alto Magdalena - Mireya Sáenz Trujillo (ENG)
Universidad Católica de Colombia - Marcela Cascante Montoya (ENG, GS)
Universidad Cooperativa de Colombia - Sede Villavicencio - Julio César Garzón Calderón (ENG)
Corporación Universidad Piloto de Colombia - Sede Bogotá - Jeanette Bermúdez Rojas (ENG)
Universidad de Los Andes - Laura Cristina Pinilla Vera (ENG)
Universidad de la Salle - Manuel Antonio Tobito Cuberos (ENG)
Universidad la Gran Colombia - Luis Fernando Díaz (ENG)
Universidad Pontificia Bolivariana - Sede Bucaramanga - Aldemar Remolina Millán (ENG)
Universidad Santo Tomas - Myriam Gómez (ENG)
Universidad Distrital Francisco José de Caldas - Carlos Gregorio Pastrán Beltrán (ENG)
Universidad de Sucre - José Rodrigo Hernández Ávila (ENG)
Universidad del Magdalena - Jorge Daniel Aragón Russo (ENG)
Universidad del Quindío - Carlos Arturo García Ocampo (ENG)
Universidad de Cartagena - Modesto Barrios Fontalvo (ENG)
Universidad Industrial de Santander - Germán García Vera (ENG)
Universidad Militar Nueva Granada - Luis Alexander Godoy Estrella (ENG)
Universidad del Cauca - José Fernando Sánchez Fernández (ENG)
Universidad Pedagógica y Tecnológica de Colombia - Angel Eduard Rodriguez Suesca (ENG)
Universidad Nacional de Colombia - Sede Manizales - Alba Neira (ENG)
Universidad de Caldas - José Fernando Quintana Velasco (GS)
Universidad del Atlántico - Graciela Angulo (GS)
Universidad del Valle - Ana María Sanabria Rivas (GS)
Universidad de Nariño - Mónica Benitez Córdoba (GS)
Universidad del Tolima - Jairo Ricardo Mora (GS)
Fundación Universidad de Bogotá-Jorge Tadeo Lozano - Juan Carlos Aponte Romero and José Daniel Bogoya Maldonado (GS)
Universidad EAFIT - María Eugenia Hoyos de Hernández (GS)
Universidad de Manizales - María Piedad Marín Gutiérrez (GS)
Fundación Universidad Autónoma de Colombia - FUAC - Yolando Gonzalez (GS)
Universidad Libre - Jorge René Silva Larrotta (GS)

© OECD 2012

Fundación Universitaria los Libertadores - Juan Vincente Ortiz Franco (GS)
Corporación Universitaria del Huila - Corhuila - Robinson Casallas Montealegre (GS)

Egypt (ECO, ENG, GS)

Alexandria University - Fekry Mohamed Hussein (ENG, GS)
Ain Shams University - Samar Abdel Azim Ahmed (ENG, GS)
Assuit University - Ahmed Abdelhafiz Ahmed Abdelhafiz (ENG, GS)
Mansoura University - Hamdy Nabih Agizah (ENG, GS)
Zagazig University - Mohamad Almurri Mohamad Ismail (ENG, ECO)
Helwan University - Yasser Housny Sakr (ENG, GS)
Minufiya University - Meawad Mohamed Elkholy (ENG)
Fayoum University - Abeer Elsayd Abuzaid (ENG, GS)
Benha University - Sedhom Asaad Sedhom (ENG)
Cairo University - Ezzaldin Omar Abusteit (ENG, ECO)
South Valley University - Abdel Basit Masood Ebied (ECO)
Beni-Suif University - Mohamed Hammad Hendy (ECO)
Kafrelshiakh University - Sherein Hamed Abou-Warda (ECO, GS)
Sohag University - Samir Ahmed Abdalnaeem (ECO)
Port Said University - Khaled Abd Elraouf Ebada (ECO)
MUST University - Adel Sayed El-Toukhy (ECO)
Tanta University - Abozaid Said Elcheweky (ECO, GS)
Suez Canal University - Bayoumi Mohamed Dahawy (ECO, GS)
Minia University - Emad Abdel-Massih Youssef Elsafty (GS)

Finland (GS)

Aalto University - Reetta Koivisto
University of Helsinki - Heidi Hyytinen,
University of Eastern Finland - Sari H. Pitkänen,
University of Jyväskylä - Tuula Maijanen
Lappeenranta University of Technology - Annikka Nurkka
University of Oulu - Eva Maria Raudasoja
Haaga-Helia University of Applied Sciences - Helena Mansio
Central Ostrobothnia University of Applied Sciences - Anna Pulkkinen
Helsinki Metropolia University of Applied Sciences - Leila Lintula
Oulu University of Applied Sciences - Asko Karjalainen
Satakunta University of Applied Sciences - Anne Sankari
Savonia University of Applied Sciences - Tuula Peura

Italy (ECO)

Università degli Studi dí Bergamo - Annalisa Cristini
Alma Mater Studiorum - Università di Bologna - Alberto Roverato
Università degli Studi di Roma “La Sapienza” - Silvia Fideli
Libera Università Internazionale Studi Sociali “Guido Carli” LUISS - ROMA - Gionvanna Vallanti
Università degli Studi di Milano-Bicocca - Biagio Ciao

© OECD 2012

Università degli Studi di Napoli "Federico II" - Riccardo Martina
Università degli Studi del Piemonte Orientale “Amedeo Avogadro” - Eliana Baici
Università degli Studi di Siena - Facoltà di Economia “R.M.Goodwin” - Giiulio Ghellini
Universita degli Studi Di Torino - Facolta Di Economia - Donatella Busso
Università degli Studi di Udine - Marcellino Gaudenzi

Korea (GS)

Gyeongsang National University - Bong-geun An (Team Leader) and Sook-nyun Kim (Staff)
Kyung Hee University - Wonki Hong (Sr. Analyst) and Seungryul Jee (Staff)
Pusan National University - Gi Jae Park (Chief)
University of Seoul - Hak Jin Kim (Team Leader)
Sungkyunkwan University - Sangchun Hwang (Mgr)
Chonbuk National University - Yong Kim (Team Mgr) and Kyunghwan Lee (Staff)
Chung-Ang University - MinSeon Kim (Staff)
Hankuk University of Foreign Studies - Hyo-Young Choi (Team Leader) and Hyo-Sik Won (Mgr)
Handong Global University - Jina Choi (researcher)
Lead Scorer: Eun Young Kim (Korean Educational Development Institute)

Kuwait (GS)

Kuwait University - Ferial Borbaie
Public Authority for Applied Education and Training - Esam Jrag
American University of Kuwait - Rawda Awaad
Gulf University for Science & Technology - Nada Alenzi
The Open University of Kuwait - Shafi Aldouseri
American University of Middle East (partial participation)
Australian College of Kuwait's - Saad Alamri

Mexico (ECO, ENG, GS)

Main institutional coordinator is shown first, strand-specific coordinators are show after.

Universidad Autónoma de San Luis Potosí, Luz María Nieto Caraveo, Gerardo Castañon Nieto (ECO), Arturo Dufour Candeleria and Mirna Sandoval Medina (ENG), Sergio Dávila Espinosa (GS)
Universidad de Guadalajara: Patricia Rosas Chávez, Javier Medina Ortega (ECO), Carlos Pelayo Ortiz (ENG), Marisol Luna Rizo (GS)
Universidad Autónoma de Yucatán, José Antonio González Fajardo and José Humberto Loría Arcila, Luis Alberto Araujo Andrade (ECO), José Ángel Méndez Gamboa (ENG), Brenda Gómez Ortegón (GS)
Tecnológico de Monterrey (Campus Monterrey), José Rafael López Islas, Gerardo Isaac Campos Flores and Marcela Maldonado de Lozada (ECO), Gerardo Isaac Campos Flores and Carlos Enrique Nungaray Pérez (ENG), Gerardo Isaac Campos Flores and Liliana Manrique Cadena (GS)
Universidad Veracruzana, Héctor Julián Vargas Rubin, Silvia María Méndez Maín (ECO), Eduardo Castillo González (ENG), María de Lourdes Watty Urquidi (GS)
Instituto Politécnico Nacional, Daffny J. Rosado Moreno, Victor Daniel Escalante Huitrón (ENG), Tatiana Nayeli Domínguez Mendoza (GS)

© OECD 2012

Instituto Politécnico Nacional (Escuela Superior de Economía), Valente Mendiola Hernández (ECO)
Universidad Autónoma de Coahuila, José Abdón Padilla Hernández, Frederico Ricardo Müller Rodríguez (ECO), Roberto del Toro Walls (ENG), Aurora Bustillo Garfias (GS)
Universidad Autónoma de Chihuahua, Rosendo Mario Maldonado Estrada, Cecilia Mercedes Meléndez Ronquillo (ECO), Javier González Cantú (ENG), Luis Felipe Moriel Acosta (GS)
Universidad Autónoma de Zacatecas, Lorena Jiménez Sandoval, Fernando Robledo Martinez (ECO), Enrique Jiménez Flores (ENG), Lorena Jiménez Sandoval (GS)
Universidad de Colima, José Manuel Orozco Plascencia and Francisco Javier Guzmán, José Manuel Orozco Plascencia (ECO), Francisco Javier Guzmán (ENG)
Universidad Tecnológica de la Mixteca, Agustín Santiago Alvarado, Adriana Mejía Alcauter (GS)
Instituto Tecnológico Superior de Irapuato, Javier Bañuelos Ortega, Sergio Elías Pérez Pizano (GS)
Universidad Politecnica de Aguascalientes, Marco Antonio Paz Ramos, Mónica Terán Díaz (GS)
Universidad de Ciencias y Artes de Chiapas, Santos Jadiel Castro Villatoro (GS)

Norway (GS)

Norwegian University of Life Sciences (UMB), Faye Frances Benedict and Lars Vemund Solerød
Norwegian University of Science and Technology (NTNU), Ole Kristen Solbjørg
Lillehammer University College (HiL), Jens Uwe Korten and Arve Thorsberg
University of Stavanger (UiS), Marit Cecilie Farsund
Vestfold University College (HiVe), Bjørnulf Stokvik

Russia (ENG, ECO)

Far Eastern Federal University, Elena Gafforova and Tatiana Eliseeva (Subst) (ENG, ECO)
Northern (Arctic) Federal University, Georgy Sukhanov (ECO), Yulia Churkina (ENG)
North-Eastern Federal University named after M.K. Ammosov, Natalia Okhlopkova (ECO), Nadezhda Nelakirova (ENG)
Ural Federal University named after the First President of Russia B.N. Yeltsin, Victoria Beliaeva (ECO), Oleg Rebrin and Irina Sholina (Asst) (ENG)
National Research Irkutsk State Technical University, Vera Vlasova (ENG)
National Research Tomsk Polytechnic University, Elena Muratova (ENG)
Tyumen State of Oil and Gas University, Alena Antipova and Larisa Skomorohova (Asst) (ENG)
Ural State Mining University, Lyudmila Gavrilova and Anton Lylov (Asst) (ENG)
Ural State University of Railway Transport, Evgeny Malygin and Yulia Arkhipova (Asst) (ENG)
Don State Technical University, Elena Chernysheva and Pavel Obukhov (Asst) (ENG)
I.I. Polzunov Altai State Technical University, Marina Kaigorodova and Larisa Kapustina (Asst) (ENG)
The Russian Presidential Academy of National Economy and Public Administration, Elena Karpuhina, Sergei Maruev (Subst.) and Elena Leontyeva (Asst) (ECO)
The State University of Management, Marina Chelyshkova and Olga Davidova (subst) (ECO)
Kazan (Volga Region) Federal University, Vladimir Malaev (ECO)
Moscow State University of International Relations, Alexandra Khudaykulovaand Irina Kuzmina (Asst) (ECO)

© OECD 2012

Moscow State University of Economics, Statistics and Informatics, Natalia Dmitrievskaya, Irina Smirnova (Subst) and Anna Vershinina (Asst) (ECO)
National Research Novosibirsk State University, Gagik Mkrtchyan and Naimdzhon Ibragimov (Subst) (ECO)
National Research University - Higher School of Economics, Kirill Bukin, Alexey Pomomarenko, Oleg Zamkov, Vladimir Zuev, Natalia Gorbacheva, Anna Bykoova, and Tatiana Prokhorova (ECO)
Plekhanov Russian University of Economics, Galina Selyanskaya (ECO)
Saint Petersburg State University, Olga Kanaeva (ECO)
Saint Petersburg State University of Economics and Finance, Irina Egorova and Marina Voloshinova (ECO)
Southern Federal University, Irina Mostovaya and Maria Kazakova (ECO)
Altai State University, Maxim Kostenko, Olga Mamchenko (Subst), Olga Alimushkina (Asst), Marina Titova (Asst) and
Lomonosov Moscow State University (phase 1 only), Boris Derevyagin (ECO)
Siberian Federal University, Elena Sevastyanova and Elena Sapozhnikova (Asst) (ECO)
Financial University under the Government of the Russian Federation (phase 1 only), Natalia Dumnaya

Slovak Republic (ECO, ENG, GS)

Slovak University of Technology in Bratislava, František Horňák (ENG)
Constantine the Philosopher University, Režena Žilová (GS)
University of Prešov, Milan Portik (GS)
Matej Bel University, Banská Bystrica, Štephan Porubský (GS), Ladislav Klement (ECO)
University of Veterinary Medicine and Pharmacy, Košice, Jana Mojžišová (GS)
Slovak University of Agriculture, Nitra, Zdenka Gálová (GS)
Police Academy, Bratislava, Ľuboš Wäldl (GS)
Pavol Jozef Šafárik University, Košice, Pavol Sovák (GS)
Comenius University, Bratislava, Ján Pekár (GS), Ján Boda (ECO)
College of Security Management, Košice, Pavel Raschman (GS)
Bratislava International School of Liberal Arts, Samuel Abrahám (GS)
Pan-European University, Bratislava, Radka Sabová (GS, ECO)
Alexander Dubček University of Trenčín, Pavol Hrivik (GS), Jozef Habánik (ECO)
Trnava University, Ľubica Ilievová (GS)
Slovak Medical University in Bratislava, Viktor Majtán (GS)
Slovak University of Agriculture, Nitra, Jozef Repiský (ECO)
Technical University of Košice, Oto Hudec (ECO)
University of Zilina, Miloš Poliak (ECO)
Economics University, Bratislava, Erika Pastoráková (ECO)

United States - States of Connecticut, Missouri, Pennsylvania (GS)

Southern Connecticut State University, Michael Ben-Avie (Sr Research Analyst) and Maureen Gilbride-Redman (MPA)
Missouri State University, Kelly Cara, Assessment Research Coordinator
Central Methodist University, Amy Dykens, Assistant Dean

© OECD 2012

Truman State University, Nancy Asher, Coordinator of Testing and Reporting and Assoc. Budget Officer
Webster University, Julie Weissman, Director of Institutional Effectiveness
Cheyney University of Pennsylvania, Wesley C. Pugh, (Professor)
Clarion University, Susan C. Turell (Assoc Provost)
Lock Haven University, David L. White (Interim Provost)
Millersville University, Joseph R. Maxberry III (Inst'l Effectiveness Mgr)
University of Central Missouri, Michael Grelle (Vice Provost)
Edinboro University, Rene Hearns (Director Grants and Sponsored Prgms)

AHELO Ad-hoc Expert Groups

Washington Expert Group - April 2007

Jo Ritzen (Chair), President of Maastricht University (Netherlands)
Anne Baly, Teaching, Equity and Collaboration Branch Manager, Department of Education, Science and Training (Australia)
Roger Benjamin, President, CAE - Council for Aid to Education (United States)
Bong-Gun Chung, Director-General, Ministry for Education (Korea)
Masao Homma, National Institute for Academic Degrees and University Evaluation (Japan)
Akihiko Kawaguchi, Vice-President, National Institute for Academic Degrees and University Evaluation (Japan)
Salvador Malo, Director of Research, IMCO - Mexican Competitivity Institute (Mexico)
Detlef Müller Böling, Director, CHE - Gemeinnûtziges Centrum für Hochschulentwicklung (Germany)
Mari Pearlman, Senior Vice President and General Manager, Higher Education Division (United States)
Jamil Salmi, Tertiary Education Coordinator (World Bank)
Marshall Smith, Professor and former Undersecretary of State for Education, William and Flora Hewlett Foundation (United States)

Paris Expert Group - July 2007

Roger Benjamin, President, CAE - Council for Aid to Education (United States)
Bong-Gun Chung, Director-General, Ministry for Education (Korea)
Sang-Duk Choi, Director of Office of Higher and Adult Education, KEDI - Korean Educational Development Institute (Korea)
Anne Gregory, Director of Analysis Unit, Higher Education Group, Department of Education, Science and Training (Australia)
Masao Homma, Vice Chancellor and Professor, Ritsumeikan Trust (Japan)
Akihiko Kawaguchi, Vice-President, National Institute for Academic Degrees and University Evaluation (Japan)
Salvador Malo, Director of Research, IMCO - Mexican Competitivity Institute (Mexico)
Detlef Müller Böling, Director, CHE - Gemeinnûtziges Centrum für Hochschulentwicklung (Germany)
Jamil Salmi, Tertiary Education Coordinator (World Bank)

© OECD 2012

Marshall Smith, Program Director, William and Flora Hewlett Foundation (United States)
Dirk Van Damme, Chief Executive, Cabinet of the Minister of Education in the Flemish Community (Belgium)
Thomas Van Essen, Senior Assessment Director, Educational Testing Service (United States)
Frans Van Vught, EUA Board Member, University of Twente (Netherlands)
Marijk Van der Wende, Chair IMHE Programme, CHEPS - Centre for Higher Education Policy Studies, University of Twente (Netherlands)
Matthias von Davier, Principal Research Scientist, Center for Statistical Theory and Practice in the Research and Development Division, Educational Testing Service (United States)
Peter Williams, Chief Executive, Quality Assurance Agency for Higher Education (United Kingdom)

Seoul Expert Group - October 2007

Roger Benjamin, President, CAE - Council for Aid to Education (United States)
Bong-Gun Chung, Director-General, Ministry for Education (Korea)
Yung-Chul Kim, Director of Policy Team, Human Resources Planning Policy Office, Ministry of Education and Human Resources Development (Korea)
Sang-Duk Choi, Director of Office of Higher and Adult Education, KEDI - Korean Educational Development Institute (Korea)
Anne Gregory, Director of Analysis Unit, Higher Education Group, Department of Education, Science and Training (Australia)
Masao Homma, National Institute for Academic Degrees and University Evaluation (Japan)
Akihiko Kawaguchi, Vice-President, National Institute for Academic Degrees and University Evaluation (Japan)
Terrel L. Rhodes, Vice President for Quality, Curriculum and Assessment, AACU - Association of American Colleges and Universities (United States)
Jamil Salmi, Tertiary Education Coordinator (World Bank)
Marshall Smith, Program Director, William and Flora Hewlett Foundation (United States)
Are Turmo, Senior Advisor, Department for Policy Analysis, Lifelong Learning and International Affairs, Ministry of Education and Research (Norway)
Thomas Van Essen, Senior Assessment Director, Educational Testing Service (United States)
Marijk Van der Wende, Chair IMHE Programme, CHEPS - Centre for Higher Education Policy Studies, University of Twente (Netherlands)

Initial Contextual Dimension Expert Group - November 2008 and February 2009

Per Olaf Aamodt, NIFU STEP (Norway)
Jim Allen, ROA (The Netherlands)
Eric Bettinger, Stanford School of Education (USA)
John Brennan, The Open University (UK)
Peter Ewell, NCHEMS (USA)
Gero Federkeil, CHE Centre for Higher Education Development (Germany)
Frans Kaiser, CHEPS (The Netherlands)
Rafael Llavori, ANECA (Spain)
Alexander C. McCormick, Indiana University Center for Postsecondary Research (USA)

© OECD 2012

Naoyuki Ogata (Japan)
Anna Prades (Spain)
Josep Torres, AQU CATALUNYA, (Spain)

Tuning-AHELO Economics Expert Group - May 2009

Project Coordinator: Julia González, University of Deusto (Spain)
Project Assistant: Pablo Beneitone, University of Deusto (Spain)
Chair: John Beath, University of St. Andrews, Royal Economic Society (UK)
Rapporteur: William E. Becker, Journal of Economic Education, SSRN Economic Research Network Educator, Indiana University, University of South Australia, Institute for the Study of Labor - IZA, Bonn Germany (USA)

Mark Freeman, University of Sydney (Australia)
Friedrich Schneider, University of Linz (Austria)
Filip Abraham, K.U. Leuven (Belgium)
Julius Horvath, Central European University (Hungary)
S. Madheswaran, Institute for Social and Economic Change (India)
Cecilia A. Conrad, Scripps College, International Association for Feminist Economists (IAFFE)
Fiorella Kostoris, Universitá di Roma (Italy)
Tatsuya Sakamoto, Keio University (Japan)
Elvio Accinelli Gamba, Universidad Autónoma de San Luis Potosí (Mexico)
Vladimir Zuev, State University- Higher School of Economics (Russia)
Henriëtte Maassen van den Brink, University of Amsterdam (the Netherlands)
W. Groot, University Maastricht (the Netherlands)

Corresponding members
Bernard Cornet, Université Paris 1 Panthéon-Sorbonne (France)
Aswin van Oijen, University of Tilburg (the Netherlands)
Erik Mellander Institute for Labour Market Policy Evaluation - IFAU (Sweden)

Tuning-AHELO Engineering Expert Group - May 2009

Project coordinator: Robert Wagenaar, University of Groningen (the Netherlands)
Project assistant: Ingrid van der Meer, University of Groningen (the Netherlands)
Chair: James L. Melsa, Iowa State University (USA)
Rapporteur: Iring Wasser Accreditation Agency for Degree Programmes in Engineering, Informatics, Natural Sciences and Mathematics - ASIIN E. V (Germany)

Yinghe He, James Cook University (Australia)
Philippe Wauters, FEANI - Federation of European National Engineers' Associations (Belgium)
Paulino Alonso, Pontificia Universidad Católica de Valparaiso, Council of Engineering Deans of Chile - CONDEFI (Chile)
Pierre Compte, Laboratoire régional des Ponts & Chaussées, Commission des titres d'ingénieur (France)
Giuliano Augusti, Università La Sapienza, EUR-ACE, ENAEE (Italy)

© OECD 2012

Kikuo Kishimoto, Tokyo Institute of Technology (Japan)
Eric van der Geer-Rutten-Rijswijk, Eindhoven University of Technology (the Netherlands)
Alfredo Soeiro, University of Porto (Portugal)
Iacint Manoliu, Technical University of Civil Engineering (Romania)
Mats Hanson, KTH- Royal Institute of Technology (Sweden)
Rob Best, London South Bank University (United Kingdom)

Corresponding members:
Andrew Downing, University of South Australia
Roger Hadgraft, The University of Melbourne (Australia)
Francesco Maffioli, Professor of operational research and Coordinator of several European Thematic Networks in Engineering Education - H3E, E4, TREE (Italy)
Peter Wieringa, Delft University of Technology (the Netherlands)
Johan Malmqvist, Chalmers University of Technology (Sweden)
Jim Birch, Engineering Council UK (United Kingdom)
Patricia Daniels, Seattle University (USA)
Edwin C Jones Jr, Iowa State University (USA)

Technical Review Panel (Sept 2009)

Pierre Brochu, CMEC (Canada)
Piero Cipollone, INVALSI (Italy)
Yinghe He, James Cook University (Australia)
Kyung-Sung Kim, Seoul National University of Education (Korea)
Dianne Pennock (Canada)

Economics Expert Group

Chair: Cecilia Conrad, Chair, Pomona College (U.S)
Fiorella Kostoris, La Sapienza Facoltá di Economia (Italy)
Maria de Lourdes Dieck-Assad, Monterrey Institute of Technology and Higher Education (Mexico)
Henriëtte Maassen van den Brink, University of Amsterdam (the Netherlands)
Tatsuya Sakamoto, Keio University (Japan)
Vladimir Zuev, State University—Higher School of Economics (Russia)

Engineering Expert Group

Chair: Robin King, Australian Council of Engineering Deans
Giuliano Augusti, Universita 'La Sapienza' (Italy)
Michael Hoffman, University of Ulm (Germany)
Kikuo Kishimoto, Tokyo Institute of Technology (Japan)
Johan Malmqvist, Chalmers University of Technology (Sweden)
Nobutshi Masuda, Tokyo City University (Japan)
Jim Melsa, Iowa State University (United States)
Lueny Morell, Hewlett Packard (United States)

© OECD 2012

Technical Advisory Group

Chair: Peter Ewell, NCHEMS (USA)
Vaneeta D'Andrea, University of the Arts London (UK)
Stuart Elliott, Board on Testing and Assessment, the National Academies (USA)
Motohisa Kaneko, Tsukuba University (Japan)
V. Lynn Meek, University of Melbourne (Australia)
Paul Holland, ETS (USA)
Keith Rust, WESTAT (USA), member until July 2012
Chris Skinner, London School of Economics and Political Science (UK)
Frans Van Vught, CHEPS (the Netherlands)
Robert Wagenaar, University of Groningen (the Netherlands)

Stakeholders' Consultative Group (SCG)

Association of American Colleges and Universities (AAC&U): Terrel Rhodes
American Council on Education (ACE): Molly Corbett Broad, Patti McGill Peterson and Bradley Farnsworth
Association Européenne des Conservatoires, Académies de Musique et Musikhochschulen (AEC) : Linda Messas
Asia-Pacific Quality Network (APQN): Jagannath Patil and Antony Stella
Business and Industry Advisory Committee to the OECD (BIAC): Irene Seling, Marita Aho, Charles Fadel, Jonathan Greenhill, Carrie Murdoch
Bologna Process Secretariat: Gayane Harutyunyan
Calouste Gulbenkian Foundation: Eduardo Grilo and Manuel Rosa
Council of European Employers of the Metal, Engineering and Technology-Based Industries (CEEMET): Henri de Navacelle, Isabelle Biais, Carla Botten-Verboven and Maurice Pinkus
Coimbra Group: Dorothy Kelly
Compagnia di San Paolo: Barbara Daviero
Council for Higher Education Accreditation (CHEA): Judith Eaton
Education International (EI): David Robinson, Monique Fouilhoux and John Bangs
European Association for Quality Assurance in Higher Education (ENQA): Achim Hopbach and Fiona Crozier
European Association of Institutions in Higher Education (EURASHE): Stefan Delplace
European Economic Association (EEA): Piero Tedeschi and Gemma Prunner-Thomas
European University Association (EUA): Lesley Wilson, Tia Loukkola, Andrejs Rauhvargers, Andrée Sursock and Thérèse Zhang
European Students' Union (ESU): Andrea Blaettler and Nevena Vuksanovic
The Higher Education Authority (HEA), Ireland: Tom Boland
The Higher Education Funding Council for England (HEFCE): William Locke
International Association of Universities (IAU): Eva Egron-Polak and Isabelle Turmaine
International Network for Quality Assurance Agencies in Higher Education (INQAAHE): Iring Wasser and Esther van den Heuvel
Lumina Foundation for Education: Jamie Merisotis
European Society for Engineering Education (SEFI): Wim Van Petegem and Jorg Steinbach

© OECD 2012

The Spencer Foundation: Michael McPherson
Teagle Foundation: Richard Morrill
Trade Union Advisory Committee to the OECD (TUAC): Roland Schneider
Union of Universities of Latin America and the Caribbean (UDUAL): Rocio Santamaria Ambriz

AHELO Contractors

AHELO Consortium

International Project Director: Hamish Coates (2010-2011) and Sue Thomson (since Feb. 2012)
Director Generic Skills strand: Roger Benjamin (CAE)
Director Economics strand: Tom Van Essen (ETS)
Director Engineering strand: Julian Fraillon (ACER)
Director Contextual Dimension: Jon File (CHEPS)
Director Project Management and Survey Operations: Sarah Richardson (ACER)

In addition, the following staff was involved in the different Consortium partner organisations:

Australian Council for Educational Research

Ray Adams
John Ainley
Yan Bibby
Rajat Chadha
John Cresswell
Daniel Edwards
Patricia Freeman
Tim Friedman
Roger Hadgraft
Brad Jackel
Siek Toon Khoo
Jacob Pearce
Doug McCurry
Martin Murphy
David Noga
Wolfram Schulz
Xiaoxun Sun
Ling Tan
Rebecca Taylor
Ross Turner
Yu Zhao

BranTra

Musab Hayatli
Ben Meessen

© OECD 2012

cApStAn

Steve Dept
Andrea Ferrari
Shinoh Lee

Centre for Higher Education Policy Studies

Don Westerheijden

Centre for Post-Secondary Research

Alexander McCormick

Council for Aid to Education

Scott Elliot
Kurt Geisinger
Robert Keeley
Steve Klein
Richard Shavelson
Willy Solano-Flores
Jeffrey Steedle
Doris Zahner

Educational Testing Service

Claire Melican
Rae Jean Goodman

IEA DPC

Falk Brese
Tim Daniel

NIER

Satoko Fukahori
Shuichi Tsukahara

SoNET Systems

Steven Birchall
Mike Janic
Rakshit Shingala

Statistics Canada

Jean Dumais

University of Florence

Claudio Borri
Elisa Guberti

© OECD 2012

Other ad-hoc Experts

Ms. Béatrice Halleux, Hallstat (Belgium)
Professor Marijk Van der Wende, Vrije Universiteit Amsterdam (the Netherlands), former Chair of the IMHE Governing Board

OECD Secretariat

Senior Management

Barbara Ischinger, Director for Education
Andreas Schleicher, Deputy Director for Education, Special Advisor to the OECD Secretary-General for Education
Deborah Roseveare, Head of the Skills Beyond Schools Division
Richard Yelland, Head of the Policy Advice and Implementation Division

AHELO Team

Karine Tremblay, Senior Survey Manager
Diane Lalancette, Analyst
Fabrice Hénard, Analyst
Valérie Lafon, Analyst
Leslie Diamond, Analyst
Cécile Bily, Project Assistant
Sabrina Leonarduzzi, Project Assistant (March-September 2009)

Consultants, seconded staff and interns

We also wish to thank the following consultants, seconded staff and interns for their work on the AHELO feasibility study:

Rodrigo Castañeda Valle
HoonHo Kim
Claire Leavitt
Deborah Nusche
Eleonore Perez Duarte
Alenoush Saroyan
Tupac Soulas
Takashi Sukegawa
Mary Wieder

© OECD 2012

NOTES

[1] An exception is the Council for Aid to Education (CAE) which separately developed and adapted the Generic Skills strand's performance task instrument prior to joining the international Consortium for the implementation phase.

© OECD 2012

ANNEX B

ILLUSTRATIVE ASSESSMENT ITEMS IN THE VARIOUS STRANDS

Exemplar items for each of the three AHELO strands of work are provided below together with their respective scoring rubrics for the purpose of illustrating what the AHELO tests looked like.

In examining these items, readers should however bear in mind that the experience of test developers suggests that an item can have very low face/appearance validity but provide highly reliable and valid measurement of target constructs, and vice versa. In particular, it is important to remember that these items:

- were designed for use as part of an overall secure test rather than as stand-alone features;
- are not indicative of the test overall, of the framework or of other items;
- were not developed or intended for public release, or to be used to depict the test;
- would be modified and enhanced in light of results if they were to go full scale; and
- were developed for a specific purpose and time and for use in specific countries.

Information in the scoring rubrics is partial and has been developed to aid the work of a small community of scorers. The data is not designed to cover a comprehensive range of responses.

© OECD 2012

Generic Skills strand illustrative items

Example of Generic Skills CRT

Practice performance task

In order for students to familiarize with performance tasks, they were provided with an example to practice: Crime reduction (see www.collegiatelearningassessment.org/files/Architecture_of_the_CLA_Tasks.pdf for a description).

Instructions

You are a member of the administrative staff for the City of Milltown. The Mayor's Office has received many inquiries from the public and press regarding the recent discovery of a deformed catfish in Miracle Lake. The Mayor of Milltown, Sally Bigelow, plans to discuss this matter at the Milltown City Council meeting tomorrow night.

To help Mayor Bigelow prepare for tomorrow's meeting, she has asked you to review the documents provided in the Document Library (on the right side of the screen) and answer a series of questions.

Your answers to the questions that follow should describe all the details necessary to support your position. Your answers will be judged not only on the accuracy of the information you provide, but also

- how clearly the ideas are presented,
- how effectively the ideas are organized,
- how thoroughly the information is covered.

While your personal values and experiences are important, please answer all the questions solely on the basis of the information above and in the Document Library.

Write your answers in the box below each question. You can write as much as you wish; you are not limited by the size of the box on the screen.

Questions

The materials in the Document Library suggest at least three explanations for the 3-eyed catfish, namely: inbreeding, parasites, and contamination by Db09 and/or Validium.

Question 1a

What specific information, evidence, facts, and/or theories discussed in the Document Library support the **inbreeding** explanation?

Question 1b

© OECD 2012

What are arguments against the **inbreeding** explanation?

Question 2a

What specific information, evidence, facts, and/or theories discussed in the Document Library support the **parasites** explanation?

Question 2b

What are arguments against the **parasites** explanation?

Question 3a

What specific information, evidence, facts, and/or theories discussed in the Document Library support the **chemical contamination** explanation?

Question 3b

What are arguments against the **chemical contamination** explanation?

Question 4

Are there any other explanations for the 3-eyed catfish (i.e., besides inbreeding, parasites, and chemical contamination) and if so, what are they?

Question 5

What explanation for the 3-eyed catfish is most likely to be correct and why? Your answer should discuss how you might counter (rebut) possible arguments against this explanation.

Question 6

What should the City Council do now? Should it close the mill, close the conservation area, close Miracle Lake to recreation, stop taking drinking water from Bush Lake, do something else, or do nothing at all? What is the basis for your recommendation?

Question 6 is the last question. When you click on the **NEXT** button below, you will move out of this task and on to the survey. You will not be able to return to continue work on any of the questions, so only click on the **NEXT** button when you are completely finished with all of the questions in this task.

Document library

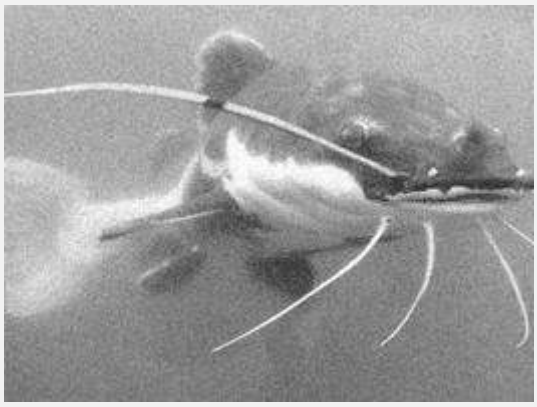
© OECD 2012

Document 1

The Milltown Clarion

Three-eyed Catfish Caught in Miracle Lake

Samantha Stewart, Staff Reporter



The three-eyed "Simpson fish" caught by Miss Lily Stone, 12, at the Milltown Miracle Celebration yesterday. "It's just gross!"

Milltown. Yesterday's Milltown Miracle Celebration at Miracle Lake was marred by the ill-timed appearance of a three-eyed catfish during the children's Fishing Derby. Lily Stone, who caught the fish, was shocked by its appearance. "It's just gross. At first I was excited to catch a fish, but when I pulled it out of the water and looked at it I just wanted to scream." Ms. Stone's father, the environmental activist, Charles Stone, blames Bonaventure Mills. "Its just like the nuclear power plant in The Simpsons that caused three-eyed fish. The mill has done it for real."

Dr. Malcolm Reis at the University Extension Service said the three-year old fish Ms. Stone caught reminded him of those caught recently by some Milltown teenagers ten miles away in Hampton County and of some frogs with extra limbs that Dr. Reis has been studying for several years. Reis said that tests showed that parasites caused these deformities in both cases. He also said that while abnormalities can be caused in many ways, such as from exposure to radioactive material, chemical contamination, and inbreeding, he thinks the most likely cause is parasites.

© OECD 2012

Ironically, Ms. Stone caught her three-eyed monster during the Fishing Derby, a contest intended to showcase the safety of the waters in the Milltown Conservation Area after the Db09 spill 10 years ago, when a break in a pipe at Bonaventure Mills spilled 100 gallons of the potent neurotoxin Db09 into the millpond. The spill killed all the animals in the water between Bush Falls and Miracle Falls. This triggered an aggressive campaign against the mill that resulted in its closure, putting 125 employees out of work. The mill remained closed for two years and recreational use of the waters between the falls was prohibited.

The mill reopened under the direction of Benjamin Bonaventure who completely reorganized the mill, its manufacturing process and its business plan. The mill no longer uses Db09. At the mill's re-opening 8 years ago, Dr. Bonaventure said "Our new manufacturing process does not release toxins of any kind. We also completely redesigned the plant to minimize the risks of spilling untreated waste into the pond. Manufacturing is more expensive now, but we can market our product as an environmentally friendly alternative to those produced the old-fashioned way. We think this gives us a competitive edge and it is the right thing to do." The mill has hired back all the laid-off employees who wanted to return and now has 115 workers.

After the spill, Bonaventure gave the land around the mill to the city to create a conservation area and park. A system of dirt trails now provides access to this beautifully wooded area.

Last winter the city council voted 7 to 1 to reopen the waters downstream of the mill to recreational use and restock Bush Creek and Miracle Lake with trout and bass. Yesterday was opening day.

© OECD 2012

Document 2

The Milltown Clarion

75 cents Monday June 27, 2005

One Monster is Enough

Charles Stone

The Milltown Miracle! A catfish so deformed it gives new meaning to the word grotesque. Can anyone imagine that this is not the inevitable consequence of the unholy alliance between the city council and Dr. Benjamin Bonaventure? The so-called miracle of cooperation is nothing more than the spectacle of corruption. My 12-year-old daughter can tell you what caused this poor fish to grow three eyes. It was chemical contamination. Meanwhile the official scientific consultants, bought and paid for by Bonaventure, hem and haw and say it might be mutation or it might be parasites or someone dumping radioactive waste. Give me a break. We all know what happened.

This is the same pond that Bonaventure Mills poisoned to death 10 years ago. Now it spawns a three-eyed fish so unnatural we have to resort to television cartoons to describe it. And our Mayor wants to "wait for the experts to evaluate the situation." What needs to be evaluated?

And what about the human population living next to this cesspool? Is anyone studying the incidence of birth defects, cancer, or leukemia? Are the wildlife biologists using tests that are relevant measures of effects in humans? Do we want to wait until the three-eyed monsters turn up in the Milltown maternity ward?

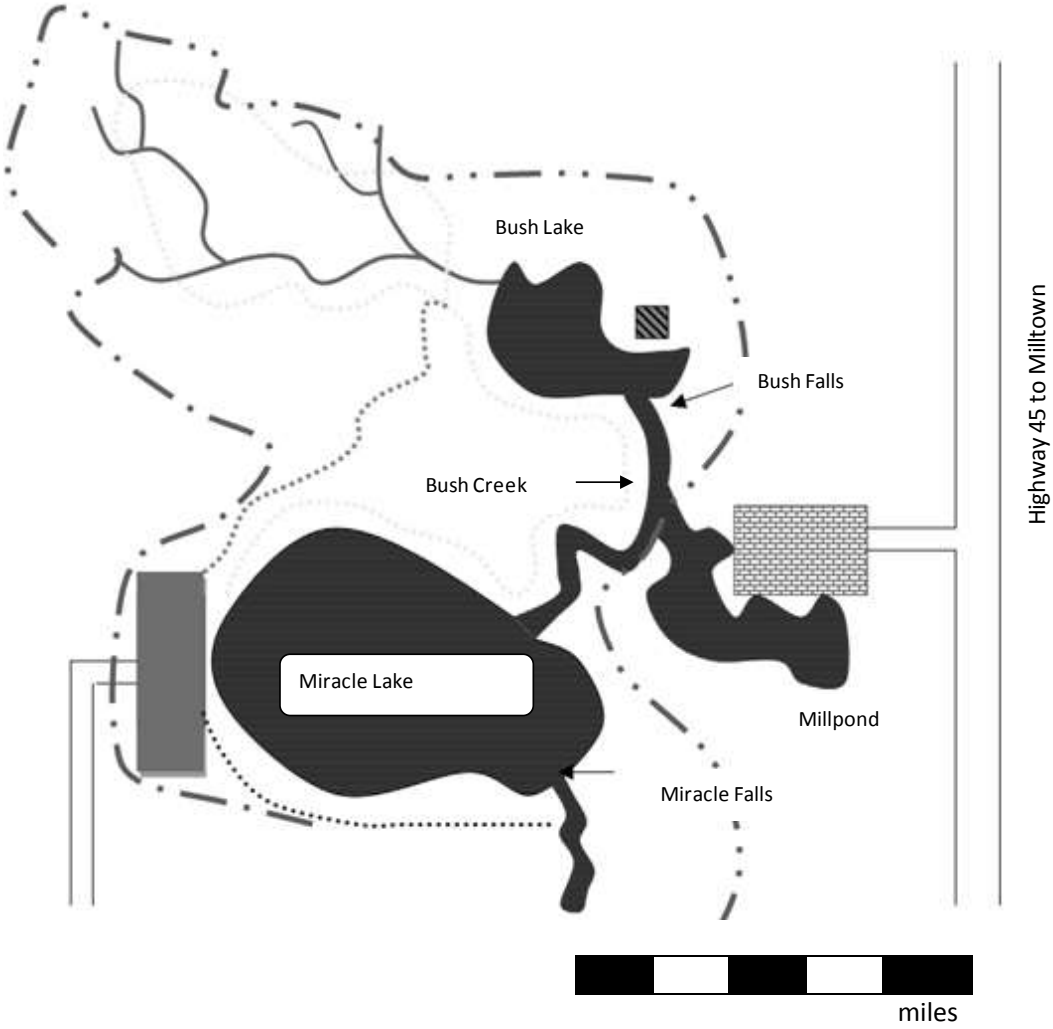
Remember how all the pregnant women who took thalidomide gave birth to deformed children. Do we want that to happen here?

I know that Milltown Elementary has two children undergoing chemotherapy right now. That must be a little unusual in a town this size. Last time that happened was 25 years ago when the mill started using Db09! We must act now. We must close the mill now. Before anyone else is hurt. One monster is enough!

© OECD 2012

Document 3

Milltown Conservation Buffer Area Map



Legend	
	Bonaventure Mills
	Miracle Trail
	Bush Creek Trail
	Marsh Trail
	Conservation Area
	Pump House
	Day Use Parking

© OECD 2012

Document 4

Dr. Seymour Munt
Department of Biology
City College

Dr. Benjamin Bonaventure
President and CEO, Bonaventure Mills
Milltown

Dear President Bonaventure:

Thanks for giving my students access to your mill's waste area and mill pond. They had a terrific experience investigating a real-world situation.

They tested: (1) the waste flow from the mill, (2) water from the mill pond, (3) water from the stream at the bend (about half a mile downstream), and (4) water above Bush Falls, near the city's water intake pipe at the pump house. Using the Cobb Test on the water samples, they found that none of these waters caused mutations significantly above background. That is, water from these sources did not appear to be mutagenic because there was no detectable change in the DNA of the indicator bacteria.

Also, because she grew up in Milltown and remembers the Db09 spill, Sandy Evans wanted to test Db09 itself. Again, results showed Db09 caused no increase in the rate of mutation in the Cobb Test bacteria. This was good news since the rocks and soils in our area are unusually rich in validium (the naturally occurring form of Db09) that can wash into our water supply. Indeed, a colleague of mine recently found that large amounts of validium are present in the sediments at the bottoms of Milltown's streams and lakes, especially where adjacent soils have been exposed.

Again, I thank you for all your help.

Sincerely,

Seymour Munt, Ph.D.
Department of Biology
City College

cc: Sally Bigelow, Mayor of Milltown; Carey Thomas, President of City College

© OECD 2012

Document 5

Excerpt from Saginaw, M. 1998. Introduction to Laboratory Measurement: Biological Sciences. Pillsberg Press.

Cobb Test: A test developed by Bruce Cobb in the early 1970s that investigates new or old environmental chemicals for effects associated with mutagenicity. A compound is said to be mutagenic if it causes a change in the DNA (deoxyriboneucleic acid) of a living cell or organism. The test uses the bacterium Salmonella typhimurium as a test organism, and is used as a rapid and relatively inexpensive first screening of untested chemicals that are suspected of being mutagenic agents. When mutagenic agents are present, the Cobb Test has been shown to detect their presence 90% of the time. When there is no mutagenic agent present, the Cobb test falsely indicates their presence 2% of the time.

© OECD 2012

Document 6

KMIL Radio
Transcripts Dept.

TRANSCRIPT, July 15, 2005 Program: *MidWeek in MillTown*, Interview with Dr. Thomas Leusid

MWMT (or Reporter):	We have with us today the distinguished biologist Dr. Thomas Leusid from Milltown College. Dr. Leusid has been a frequent guest on our show, bringing us interesting news and insights from the natural world. Dr. Leusid and his students have been involved in monitoring the wildlife in the Milltown conservation area since it was established almost 10 years ago. Welcome Dr. Leusid.
LEUSID:	Thank you for having me on your show.
Reporter:	Tell us about the work you do with your students in the conservation parklands.
LEUSID:	My students and I have been monitoring the ecosystem of the mill pond and stream for over 10 years. We use a site upstream of the mill at the base of Bush Falls as a control site. We monitor plants, salamanders, birds and stream invertebrates three times a year. By all our measures the wildlife is doing well. I would like to thank Bonaventure Mills for their support of our projects.
Reporter:	Wait. Are you saying Bonaventure Mills pays for your fieldwork?
LEUSID:	Yes, but it is common in science to have research supported by the people and organizations that are interested in the problem. That's not to say we wouldn't welcome more public funding.

© OECD 2012

Reporter:	I'm sure you've heard about the girl catching a three-eyed catfish at Miracle Lake. Have your students found such monsters?
LEUSID:	I read about the duplicate eye structure on that fish, but I have not seen it myself. Although our research focuses mainly on amphibians, we have noticed increasing numbers of catfish. Since the discovery of the unusual catfish, my students and I have been out in the conservation area doing some catch-and-release studies of the fish at various places. As you know, catfish are bottom-feeders, so we drag the bottom of the lake. In Miracle Lake, most of the catfish seem to be pretty young, most less than 3 years old. Several of them have duplicated structures, though none nearly as dramatic as Ms. Stone's catch.
Reporter:	Yes. Hers had three eyes! It seems really unnatural.
LEUSID:	Such duplications are not as rare as you might think. There have been many reports of duplications of limbs in frogs, caused by a parasite that invades the developing embryo. Parasites also caused the deformities in some catfish that were caught recently in Hampton County. Although we cannot definitely rule out parasites at this time, we haven't found them in our Bush Falls fieldwork.
Reporter:	But what about chemical contamination? Do all these places have Db09 spills?
LEUSID:	No and that's the point. In Reis's studies, these duplications--the ones in frogs and catfish in Hampton County--are caused by parasitic infections. But I don't think that's the case here. In Miracle Lake, I

© OECD 2012

	actually favor genetic mutation and then inbreeding in the mill pond.
Reporter:	How could that happen?
LEUSID:	You remember the Db09 spill years ago that killed all the vertebrates in the mill pond, Bush Creek, and Miracle Lake? We have been monitoring the return of amphibians to these waters. These populations have recovered to their pre-spill numbers. However, Bush Falls and Miracle Falls severely restrict the access for fish. Frankly, I am surprised to find any natural fish populations. The mill pond and Miracle Lake were only recently restocked and then only with trout for the fishing contest.
Reporter:	So if the waterfalls keep fish out, where did the catfish come from?
LEUSID:	I suspect a clutch of fertilized eggs came over Bush Falls. Adult catfish spend most of their time on the bottom, under rock outcroppings and in caves. They aren't likely to be washed over the falls. Even if they did go over they probably wouldn't survive. And you would need a few fish to survive that adventure to get a male and a female. But if a clutch of eggs or even a partial clutch broke loose, then all the mature adults in the mill pond would be brothers and sisters. In other words, the small numbers of relatively large catfish we catch are probably brothers and sisters from the original clutch. The much more frequently caught smaller fish, the one to three year olds, would be their offspring. And up to now, we have only seen the deformities in the young fish.

© OECD 2012

Reporter:	So how is that related to Db09?
LEUSID:	Only that the spill killed all the fish that were there, so that a new population could be established. But if parents are brothers and sisters, their offspring will be highly inbred, and the frequency of visible mutations in the population will be high.
Reporter:	But the mutations might be caused by Db09, right?
LEUSID:	I doubt it. According to our recent observations, all the fish in the pond and Miracle Lake are much too young to have been affected by the chemical spill. And there is no evidence that Db09 causes mutations.
Reporter:	I bet Bonaventure Mills was happy to hear that. So what do you think about Dr. Reis' theory that parasites caused the deformities?
LEUSID:	Dr. Reis and I favor different explanations but neither of us can say for sure what happened. The duplicated structures also may have been caused by some agent we have not identified yet.
Reporter:	What will it take to find out what really happened?
LEUSID:	We need to raise catfish from the mill pond in the lab.
Reporter:	Maybe a listener will fund that work and then we can have you back on the show to talk about it. Thank you so much for your time today.
LEUSID:	Thank you for having me, it is always a pleasure.

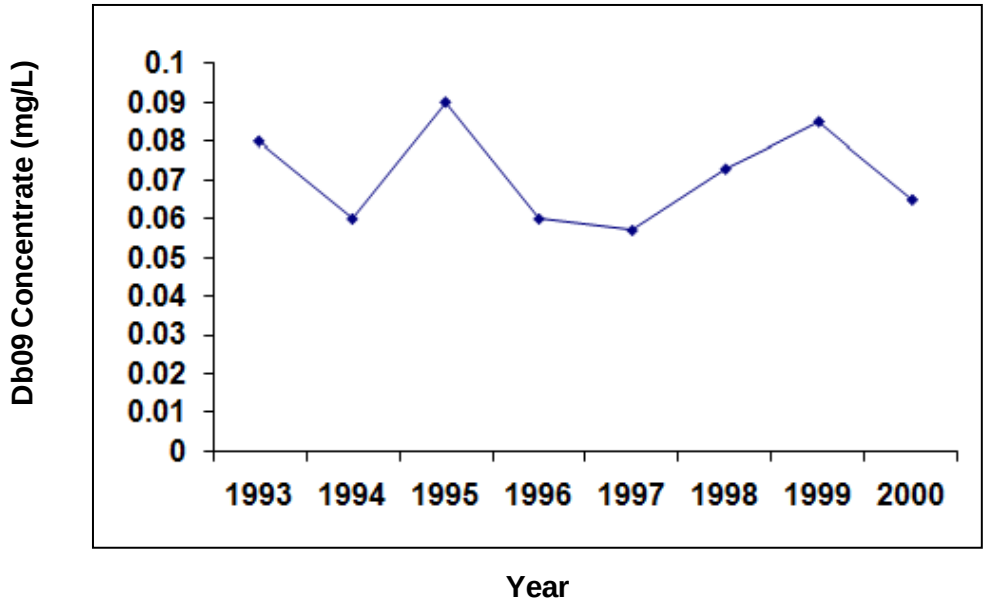
© OECD 2012

Document 7

Water Quality Report

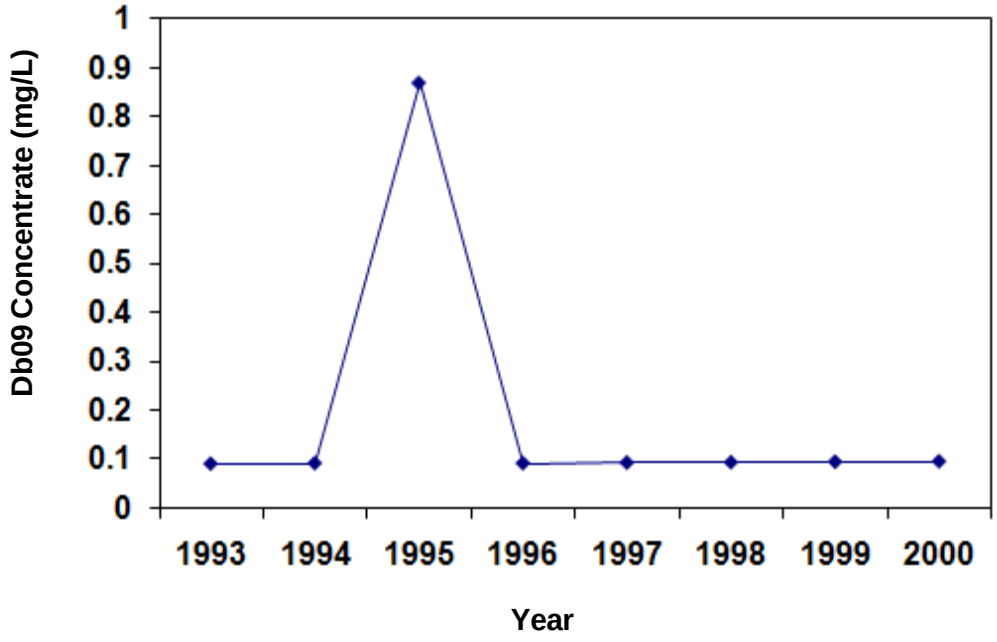
The figures below show the results of annual tests of water at two locations near Milltown between 1993 and 2000. Sample collection sites are the pump house for the city water supply and 0.25 miles downstream of Bonaventure Mills on Bush Creek. The water was tested for various contaminants, including Db09. The presence of Db09 in concentrations below 0.1 milligrams per liter (mg/L) of water is considered safe for recreational use, according to the Division of Health Standards. The tests in 1995 were performed very shortly after an industrial spill into the millpond at Bonaventure Mills.

Db09 Concentrate 0.25 miles Upstream from Millpond at Bonaventure Mills



© OECD 2012

Db09 Concentrate 0.25 miles Downstream from Millpond at Bonaventure Mills



© OECD 2012

Scoring

Analytic Reasoning and Evaluation

Description: interpreting, analyzing, and evaluating the quality of information. This entails identifying information that is relevant to a problem, highlighting connected and conflicting information, detecting flaws in logic and questionable assumptions, and explaining why information is credible, unreliable, or limited.

Scale	**Description**
0	Not scorable. Student made no attempt to answer the questions, so the response cannot be evaluated.
1	Does not identify facts or ideas that support or refute arguments presented in the Document Library (salient features of objects to be classified) or provides no evidence of analysis. Disregards or severely misinterprets important information. Does not make claims about the quality of information and bases response on unreliable information.
2	Identifies very few facts or ideas that support or refute arguments presented in the Document Library (salient features of objects to be classified). Disregards or misinterprets much of the Document Library. May restate information "as is." Does not make claims about the quality of information and presents some unreliable information as credible.
3	Identifies a few facts or ideas that support or refute several arguments presented in the Document Library (salient features of several objects to be classified). Disregards important information or makes minor misinterpretations of information. May restate information "as is." Rarely, if ever, makes claims about the quality of information and may present some unreliable information as credible.
4	Identifies a few facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). Briefly demonstrates accurate understanding of important Document Library content, but disregards some information. Makes very few accurate claims about the quality of information.
5	Identifies several facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). Demonstrates accurate understanding of much of the Document Library content. Makes a few accurate claims about the quality of information.
6	Identifies most facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). Provides analysis that goes beyond the obvious. Demonstrates accurate understanding of a large body of information from the Document Library. Makes several accurate claims about the quality of information.

© OECD 2012

Problem Solving

Description: considering and weighing information from discrete sources to make decisions (draw a conclusion and/or propose a course of action) that logically follow from valid arguments, evidence, and examples. Considering the implications of decisions and suggesting additional research when appropriate.

Scale	**Description**
0	Also 0 if and only if Analytic Reasoning and Evaluation is 0.
1	Provides no clear decision or no valid rationale for the decision. When applicable: Does not propose a course of action that follows logically from the conclusion. Does not recognize the need for additional research or does not suggest research that would address unanswered questions.
2	Provides or implies a decision, but very little rationale is provided or it is based heavily on unreliable evidence. When applicable: Briefly proposes a course of action, but some aspects do not follow logically from the conclusion. May recognize the need for additional research. Any suggested research is vague or would not adequately address unanswered questions.
3	Provides or implies a decision and some reason to favor it, but the rationale may be contradicted by unaccounted for evidence. When applicable: Briefly proposes a course of action, but some aspects may not follow logically from the conclusion. May recognize the need for additional research. Any suggested research tends to be vague or would not adequately address unanswered questions.
4	Provides a decision and credible evidence to back it up. Possibly does not account for credible, contradictory evidence. May attempt to discount alternatives. When applicable: Proposes a course of action that follows logically from the conclusion. May briefly consider implications. Recognizes the need for additional research. Suggests research that would address an unanswered question.
5	Provides a decision and a solid rationale based largely on credible evidence from multiple sources and discounts alternatives. When applicable: Proposes a course of action that follows logically from the conclusion. May consider implications. Recognizes the need for additional research. Suggests research that would address some unanswered questions
6	Provides a decision and a solid rationale based on credible evidence from a variety of sources. Weighs other options, but presents the decision as best given the available evidence.

© OECD 2012

	When applicable: Proposes a course of action that follows logically from the conclusion. Considers implications. Recognizes the need for additional research. Recommends specific research that would address most unanswered questions.

Writing Effectiveness

Description : constructing organized and logically cohesive arguments. Strengthening the writer's position by providing elaboration on facts or ideas (e.g., explaining how evidence bears on the problem, providing examples, and emphasizing especially convincing evidence).

Scale	**Description**
0	Also 0 if and only if Analytic Reasoning and Evaluation is 0.
1	Does not develop convincing arguments. Writing may be disorganized and confusing. Does not provide elaboration on facts or ideas.
2	Provides limited, invalid, overstated, or very unclear arguments. May present information in a disorganized fashion or undermine own points. Any elaboration on facts or ideas tends to be vague, irrelevant, inaccurate, or unreliable (e.g., based entirely on writer's opinion). Sources of information are often unclear.
3	Provides limited or somewhat unclear arguments. Presents relevant information in each response, but that information is not woven into arguments. Provides elaboration on facts or ideas a few times, some of which is valid. Sources of information are sometimes unclear.
4	Organizes response in a way that makes the writer's arguments and logic of those arguments apparent but not obvious. Provides valid elaboration on facts or ideas several times and cites sources of information.
5	Organizes response in a logically cohesive way that makes it fairly easy to follow the writer's arguments. Provides valid elaboration on fact or ideas related to each argument and cites sources of information.
6	Organizes response in a logically cohesive way that makes it very easy to follow the writer's arguments. Provides valid and comprehensive elaboration on facts or ideas related to each argument and clearly cites sources of information.

© OECD 2012

Examples of Generic Skills MCQs

The Generic Skills MCQs are owned by ACER and they are of commercial value. As a result, ACER did not consent to the publication of the Generic Skills MCQs. Sample items are nevertheless available on www.acer.edu.au/documents/CRT-OnlineSampleQuestions.pdf and http://www.acer.edu.au/documents/GSA_SampleQuestions.pdf

© OECD 2012

Economics strand illustrative items

Example of Economics CRT

The quotation below is from a 2010 article by Alejandro Foxley, a Chilean economist, writing about world economic conditions in 2008-2009.

"While middle-income countries have pursued regional trade agreements since the 1960s, these ties are becoming more important as the global economic crisis curtails demand from the United States and other major markets. With the Doha Round of multilateral trade talks stalled, regional trade agreements (RTAs) offer an alternative approach to increase trade, spur stronger economic growth, and lower unemployment rates in participating countries".

Question A

Explain what Foxley meant by his statement.

Model answer:

Foxley stated that countries faced with multilateral trade talks that did not reach an agreement looked for an alternative method of reducing trade barriers and prices of imports, and increasing exports. These countries were in a position to discuss regional trade agreements to expand free trade among countries in a geographical region. These countries moved to regional trade talks rather than talks involving a wider range of countries globally.

Scoring (1 point total):

Code	Description of Response
1	**Recognises that not all nations in all regions were equally affected by global slowdown; AND that some nations, especially developing ones, would be open to trade barrier reduction talks. Since multilateral trade talks were stalled, these countries were looking for regional trade agreements which would increase their trade.** Example responses: • Foxley said that since multilateral trade talks had broken down, some countries were looking for an alternative method of reducing trade barriers such as regional trade agreements. • RTA's increase efficiency in regions by removing barriers to trade and in some cases barriers to entry into a market. Since the US demand for foreign goods is shrinking exports from certain countries are declining. In order to increase exports and efficiency RTA's can be powerful and are becoming very important. • Decreasing demand from major markets has caused multi-lateral trade liberalization to slow which is causing regional trade agreements to become

© OECD 2012

Code	Description of Response
	more important for continued economic growth. • Regional trade offers an alternative to the demand in times of crisis, it seeks to revive the economy and therefore, if there is trade jobs are created.
0	**Fails to recognise that not all nations in all regions were equally affected by global slowdown; OR that some nations would be open to trade barrier reduction talks.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • Foxley stated that developed countries favored regional trade agreements over multilateral agreements. • Usually the countries depend on big countries to make trade but when big countries suffer the middle countries need to get an alliance so they won't suffer. • Regional markets are beginning to focus on the betterment of their individual nations/regions rather than being dependent on the world market. As the global market weakens, industry and regional trade are another way of saving the economy.
9	Missing (no response given)

Question B

Describe the world economic conditions in 2008-2009.

Model answer:

There was a world-wide recession (high unemployment) and financial turmoil worldwide. Many developed nations were moving toward more protectionist policies in the face of the slow growth and high unemployment. There was slower international demand for goods and services faced by developing countries due to slower growth in developed ones. Further, the impact of the recession was not the same across all nations.

Scoring (3 points total):

Code	Description of Response
3	**Describes the world economic conditions in 2008-09 by recognizing and discussing the following three aspects of the global economic situation:** **1 - State of economy worldwide --worldwide recession, high unemployment, slow**

© OECD 2012

Code	Description of Response
	growth, and/or financial turmoil **2 - Protectionist policies by developed nations—such policies were hindering the completion of multilateral trade talks to reduce trade barriers and increase world trade** **3 - Variation in impact of recession—i.e., developing nations vs. developed nations or any indication of variation in impact of recession within a group of nations.** **Response must address all of the above. A response, for example, that only indicated there was a worldwide recession with high unemployment and slow growth would not receive 3 score points since the discussion is just describing the worldwide economy** Example responses: • There was a world-wide recession (high unemployment) and financial turmoil worldwide. Many developed nations were moving toward more protectionist policies in the face of the slow growth and high unemployment. Further, the impact of the recession was not the same across all nations. • Unemployment was increasing in all countries, especially developed ones so they were less interested in reducing tariffs and more interested in promoting the domestic economy. Some developing countries wanted to increase trade by decreasing barriers. • The main feature of the global economic landscape over the years is the financial crisis that originated in the United States because of the irresponsible mortgage loan rate and the subsequent deterioration in the financial market. This item resulted in a fall in trade with Latin American countries which depend on it, and a number of problems in the field of international finance,mainly to Europe, due to the involvement of the foreign exchange market. Many countries into recession.
2	**Recognises two of the following three aspects of the global economic situation:** **1 – State of economy worldwide --worldwide recession, high unemployment, slow growth, and/or financial turmoil** **2 –Protectionist policies by developed nations—such policies were hindering the completion of multilateral trade talks to reduce trade barriers and increase world trade** **3 – Variation in impact of recession—developing nations vs. developed nations** Example responses: • There was a world-wide recession (high unemployment) and financial turmoil worldwide. Many developed nations were moving toward more protectionist policies in the face of the slow growth and high unemployment. • The world economy was struggling. Major markets, such as the US, the UK, France, and generally the entire EU were experiencing high unemployment, low consumer demand, and a general time period of recession as the economy was

© OECD 2012

Code	Description of Response
	shrinking. Countries such as China continued to buy US and EU debt at an increasing rate. • There was a worldwide recession. Developed countries faced with high unemployment were more interested in saving their domestic economy than reducing trade barriers. • During this period, there were great problems of speculation in the area of mortgages, and allocations to units that did not meet the level required to protect the security of transactions, this affected the consumption of developed countries, making the demand for markets to shrink.
1	**Recognises one of the following three aspects of the global economic situation:** **1 – State of economy worldwide --worldwide recession, high unemployment, slow growth, and/or financial turmoil** **2 –Protectionist policies by developed nations—such policies were hindering the completion of multilateral trade talks to reduce trade barriers and increase world trade** **3 – Variation in impact of recession—developing nations vs. developed nations** **Such response might include a detailed description of the state of the worldwide economy (recession, high unemployment, financial turmoil, and/or slow growth) without any discussion of either protectionist policies that were inhibiting the completion of multilateral trade talks or the fact that not all countries were equally impacted by the worldwide recession.** Example responses: • Unemployment, decreased value of currency, decreased demand for goods • Many countries, including the USA and in Europe, are suffering from economic downturns with high unemployment rates. • Based on the extreme drop in the status of the U.S. market and the lack of security in the U.S. currency. Economic conditions around the world were in a standstill, experiencing no growth or even negative growth. Causing recessions and devaluation of economies.
0	**Fails to recognise any of the three aspects of the global economic situation mentioned above.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • Too much lending peaked = higher risk of default which lead to high market instability and increased risk for investors.

© OECD 2012

Code	Description of Response
	• Unemployment was decreasing faster in developed countries than developed ones. • Inflation was increasing worldwide. Governments were trying to keep prices low by promoting international trade.
9	Missing (no response given)

Question C

How might these economic conditions have led Foxley to draw his conclusions in 2010?

Model answer

Developed countries were experiencing severe recessions and were moving toward more protectionist trade policies. Thus, talks to open multilateral trade and, consequently, national economies to engage in increased competition did not occur in this global economic environment. However, developing nations, who were not impacted as much by the financial turmoil, were willing to engage in regional trade talks to lower the barriers to trade.

Scoring (2 points total)

Code	Description of Response
2	**Foxley concluded that regional trade agreements were becoming more important in increasing trade since the multilateral trade talks were stalled. Contributing to his conclusions were the following:** **The recession was worse for developed nations and they didn't want their businesses to have to compete with foreign producers, contributing to the stalling of the multilateral trade talks.** **Developing nations weren't so impacted by the recession and were more willing to engage in regional trade talks. These countries were more interested in increasing trade to help their economies by increasing trade and growth and reducing unemployment.** **Response must include a discussion of both the stalling of the trade talks by developed countries and then the desire of developing countries to engage in regional trade talks.** Example responses: • RTAs were an alternative to multilateral trade agreements for developing nations who wanted to engage in international trade. Because recessions were less severe in developing countries, they weren't so interested in protecting their domestic economy but interested in increasing markets for trade.

© OECD 2012

Code	Description of Response
	• Doha rounds were stalled because reducing trade barriers was not a high priority for developed countries who were more interested in protecting their domestic economies. Developing countries were less impacted by recessions so more interested in RTAs.
1	**Discussion of one of the following as contributing to Foxley's conclusions:** **The recession was worse for developed nations and they didn't want their businesses to have to compete with foreign producers, contributing to the stalling of the multilateral trade talks.** **Developing nations weren't so impacted by the recession and were more willing to engage in regional trade talks. These countries were more interested in increasing trade to help their economies by increasing trade and growth and reducing unemployment.** Example responses: • The economies of big countries such as US were really hurting and they didn't want to change anything that would hurt their domestic economies. • Due to the fact that the demand from major importing countries was shrinking, exporting countries need to have more trade liberalization. By engaging in a RTA a country becomes more open to imports as well as exporting more to other countries (hopefully) which promotes efficiency and hopefully growth. • The worldwide recession was not as bad for some developing nations as many developed ones so they were more interested in engaging in trade talks and increasing trade. • By the fact that countries depended on trade with major countries and, if these are no longer willing to trade and we assume that international trade brings benefits, it is derived as seeking trade with close associates.
0	**Fails to recognise the recession was worse for developed nations, making it unlikely that they would engage in talks to increase competition for the domestic economy AND that some developing countries were not as affected by the financial turmoil and would be willing to engage in regional trade talks/agreements.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • With unemployment and decreased demand in US, the middle income countries are suffering as a result.

© OECD 2012

Code	Description of Response
	• Other countries can then supply such goods and services that the US and other markets can no longer during the crisis. • RTAs were more likely to reduce trade barriers than were multilateral trade agreements.
9	Missing (no response given)

Question D

List two possible advantages of a regional trade agreement and explain why.

Model answer

1. The response includes two advantages, such as the following, and explanations.
2. Increased efficiency because it eliminates some distortions (smaller for advanced economies and larger for poor developing countries).
3. Companies may achieve economies of scale with the removal/reduction of tariffs within the trade agreement region because trade restrictions (tariffs or quotas) reduce competition in the protected market and result in too many firms in the protected industry.
4. Increased competition resulting from free trade provides incentives for learning and innovation.
5. Increased variety of goods becomes available.

Scoring (2 points total)

Code	Description of Response
2	**Lists two possible advantages, such as the following, of a regional free trade area and explains how each is advantageous.** **1. Increased efficiency because it eliminates some distortions (smaller for advanced economies and larger for poor developing countries).** **2. Companies may achieve economies of scale with the removal/reduction of tariffs within the trade agreement region because trade restrictions (tariffs or quotas) reduce competition in the protected market and result in too many firms in the protected industry.** **3. Increased competition resulting from free trade provides incentives for learning and innovation.** **4. Increased variety of goods becomes available.** Example responses:

© OECD 2012

Code	Description of Response
	• Economies of scale – smaller economies are able to focus resources on producing a select few goods, and doing it well, with certainty that they will be able to trade for other goods they need. All parties benefit with higher quality goods. Lower transportation costs – Markets are closer to producers so it costs the producer less to get their goods to the market. This results in cheaper goods for the consumer. • If a nation has to compete with other nations, firms may find cheaper ways to produce the product. A company may achieve economies of scale if they can increase their production by selling not only to the domestic consumers but also foreign consumers. • One advantage is that the price would be lower without any tariffs so consumers benefit. Another is that people in the area can buy goods that they might not be able to have if they didn't trade. • While exists the trade agreement but no tariffs, dead weight disappears and the domestic price is optimized. It increases the quantity traded and increase profits for both nations.
1	**Lists one possible advantage, such as the following, of a regional free trade area and explains how it is advantageous.** **1. Increased efficiency because it eliminates some distortions (smaller for advanced economies and larger for poor developing countries).** **2. Companies may achieve economies of scale with the removal/reduction of tariffs within the trade agreement region because trade restrictions (tariffs or quotas) reduce competition in the protected market and result in too many firms in the protected industry.** **3. Increased competition resulting from free trade provides incentives for learning and innovation.** **4. Increased variety of goods becomes available.**
	Example responses: • Duties and import taxes only cause inefficiency. By engaging in a RTA these taxes are removed and the entire area will become more efficient as specialization would occur. RTA's promote unity within a region and create a sort of self-reliance and independence from other countries. The countries in a RTA can rely on each other and become more independent from the rest of the world because they are producing most of what they need as efficiently as it can be produced. • By increasing production, firms can achieve economies and scale and result in lower costs of production.

© OECD 2012

Code	Description of Response
	• People can get things, such as bananas, that they couldn't get if they didn't engage in international trade. • First, ensure a level of demand and, second, economic and political stability.
0	**Fails to list an advantage of a regional free trade area and explain how it is advantageous. Response may simply list an advantage but fail to explain how it is advantageous.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • Economies of scale. • Increases revenues received by the government.
9	Missing (no response given)

Question E

List two possible disadvantages of a regional trade agreement and explain why.

Model answer

1. The response includes two disadvantages, such as the following, and explanations.
2. Countries are unable to protect home industry from foreign competition.
3. At specific tariff levels, large countries are able to generate a terms-of-trade benefit. With a trade agreement the large country may lose this advantage.
4. Domestic market failure may reduce the benefits of free trade because labor or capital is insufficiently mobile.
5. Trade diversions may cause countries to trade within the region rather than from countries not part of the regional trade agreement where goods may have been cheaper before the regional trade agreement.
6. Countries within a regional trade agreement lose the ability to select their own economic policy.

Scoring (2 points total)

Code	Description of Response
2	**Lists two possible disadvantages, such as the following, of a regional free trade area and explains how each is disadvantageous.** **1. Countries are unable to protect home industry from foreign competition.** **2. At specific tariff levels, large countries are able to generate a terms-of-trade benefit.**

© OECD 2012

	With a trade agreement the large country may lose this advantage. **3. Domestic market failure may reduce the benefits of free trade because labor or capital is insufficiently mobile.** **4. Trade diversions may cause countries to trade within the region rather than from countries not part of the regional trade agreement where goods may have been cheaper before the regional trade agreement.** **5. National governments may not be able to address their macroeconomic problems using their selected stabilization policies.** Example responses: • Fewer trade options – Isolated nations have fewer trade partners in their region so they are limited to what goods are produced by those partners. Monopolized goods – Trade partners are left to trade with fewer partners so price gouging is a concern especially when the trade partner has far more trading options. • Poses a hindrance to growth, increase. Region does not benefit fully from the growth of the world market. Also might have fewer types of goods. • May cause a certain country to become net importer if they cannot produce something more efficiently than other countries in RTA. Falsely keep industries that are less efficient than those from outside countries solely because of RAT agreement (because of tariff on goods from other countries). • Inability to grow: if the region does not have adequate infrastructure, it is difficult for it to grow. Slow growth: Because it is not competent enough.
1	**Lists one possible disadvantage, such as the following, of a regional free trade area and explains how it is disadvantageous.** **1. Countries are unable to protect home industry from foreign competition.** **2. At specific tariff levels, large countries are able to generate a terms-of-trade benefit. With a trade agreement the large country may lose this advantage.** **3. Domestic market failure may reduce the benefits of free trade because labour or capital is insufficiently mobile.** **4. Trade diversions may cause countries to trade within the region rather than from countries not part of the regional trade agreement where goods may have been cheaper before the regional trade agreement.** **5. National governments may not be able to address their macroeconomic problems using their selected stabilization policies**. Example responses: • People in region may not get all the types of products they want. Price of traded goods will be higher than they would be with a tariff. • Governments in the regional free trade area will no longer be able to protect

© OECD 2012

	"infant" industries. • With a multinational trade agreement with specified tariffs, some large countries have negotiated beneficial (to the country) terms of trade. With a regional trade agreement the large countries may lose this. • Mainly, the economy would be closing and could have effects that would cause later saturation and, in turn, a low performance.
0	**Fails to list a disadvantage of a regional free trade area and explain how it is disadvantageous. Response may simply list a disadvantage but fail to explain how it is disadvantageous.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • Incomes of workers will decrease. • Cost of living will increase. Jobs will be shipped overseas. • Tariffs will increase but other trade barriers will decrease. Workers will be unable to move to other countries within the region where jobs are.
9	Missing (no response given)

Question G

Describe the difference between a free trade area and a customs union.

Model answer

A free trade area permits goods made in a participating country to enter and leave the countries that are part of the free trade area without the imposition of tariffs. Member countries can impose their own tariffs on goods exported to non-members. A customs union is an agreement whereby free trade occurs among members of the customs union but in this case the same tariff for specific goods is imposed on countries not part of the customs union.

Scoring (2 points total)

Code	Description of Response
2	**Correctly defines both a customs union and a free trade area.** Example responses: • A free trade area permits goods made in a participating country to enter and leave the countries that are part of the free trade area without the imposition of tariffs but if a country trades with a country outside the free trade area, the

© OECD 2012

	country can set its own tariffs. . A customs union is an agreement that the same tariff for specific goods is imposed on countries not part of the customs union • Both permit free trade between their members, i.e. no tariffs. In a free trade area, individual members may impose different tariffs on the imports of non-member countries. In a customs union, all members impose the same tariff on goods traded with non-member countries. • A customs union is where all parties involved adopt the same external tariffs while allowing free trade between members. A free trade area allows free trade between members with different tariffs imposed externally. • In a customs union countries form a "club" that establishes quantitites and prices to trade, in addition to the members. With free trade there are no barriers.
1	**Correctly defines either a customs union or a free trade area, but not both.** Example responses: • Free trade area is when there is no strong presence of government restrictions. The consumers and producers can freely trade without any interference of a third party such as the government. Customs union involves producers of another region or country to pay a fee or tax to sell in the present market. This is different from free trade because there are no such fees or taxes in free trade area and both the buyers and sellers can exchange goods without any encouragement or hindering by the government. • Free trade area is an area with complete trade liberalization. There are no barriers to trade which results in all member nations specializing in what they can most efficiently produce. A free trade area deals solely with trade. A customs union is a union of countries on the basis of many common traits. A customs union contains some trade liberalization (but may still maintain certain duties) but also includes aspects of political, social, and human rights ideas. A customs union concentrates on much more than just free trade. • A free trade area is the union of several countries where usually the trade is duty free.
0	**Fails to define either a customs union or a free trade area.** **Or provides a vague, incoherent, unrelated or incorrect response.** Example responses: • A free trade area involves a group of neighboring countries who agree to trade only with each other whereas in a customs union, countries may or may not be

© OECD 2012

	neighbours but can trade with each other and/or non-member countries. • In a free trade area, workers and other factors o production are able to move without restrictions between member countries but final goods produced may be taxed. In a customs union, workers and other factors of production are not free to move between member nations without restrictions, but final goods produced are. • Free trade allows markets to mix freely creating a global market that looks at some efficient state (high efficiency) while customs union regulates trade and can cause inefficiency with the market. Customs union sets the price of goods to which all countries must confirm.
9	Missing (no response given)

Examples of Economics MCQs

MCQ 1

Assume a fixed-coefficient production function,

Q = min (L, K),

where

Q = the quantity of output,

K = quantity of capital, and

L = the quantity of labor.

An increase in the price of one input will most likely result in:

A. a decrease in employment of all inputs by the same proportion, and a decrease in production

B. a decrease in employment of the input whose price has increased, an increase in all other inputs, and a decrease in production

C. a decrease in employment of the input whose price has increased, an increase in all other inputs, and unchanged production

D. no change in output because the coefficients of production are fixed

Key: Response A

© OECD 2012

MCQ 2

A French woman signs a contract this year agreeing to work for 240,000 euros next year. The price level is 100 this year, and the probability distribution for the price level next year is given below.

Price Level	Probability
100	0.5
120	0.6

The expected value of her real wages next year is:

A. 200,000 euros

B. 216,000 euros

C. 218,000 euros

D. 240,000 euros

Key: Response B

© OECD 2012

Engineering strand illustrative items

Example of Engineering CRT

The Hoover Dam is a 221-metre 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.

Figure 1: Hoover Dam

The following two images were prepared before the dam was built.

© OECD 2012

Figure 2: The site proposed for the dam in about 1921.

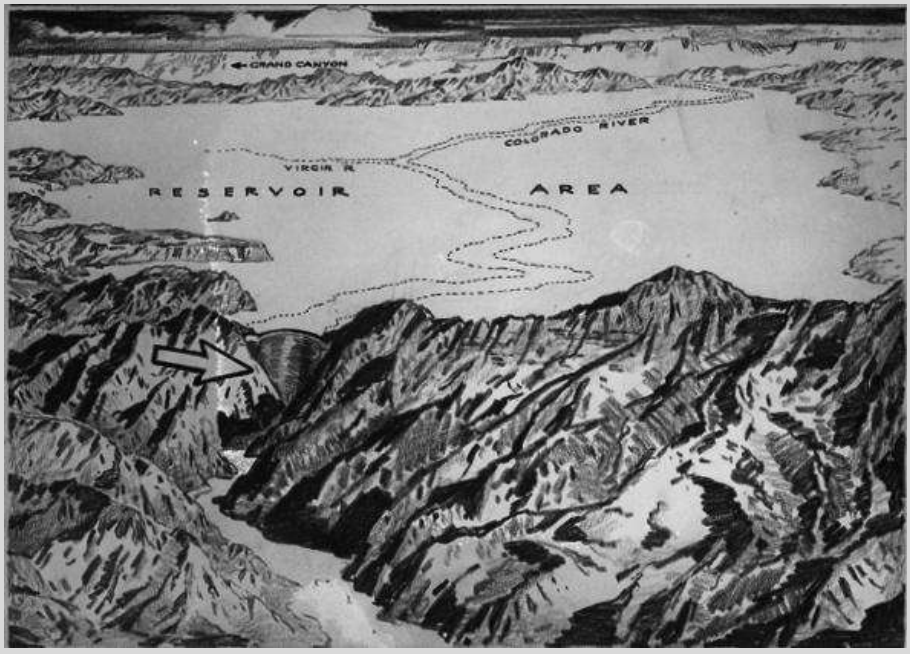


Figure 3: A sketch of the proposed reservoir.

The next two images show construction plans for the dam and power stations.

© OECD 2012

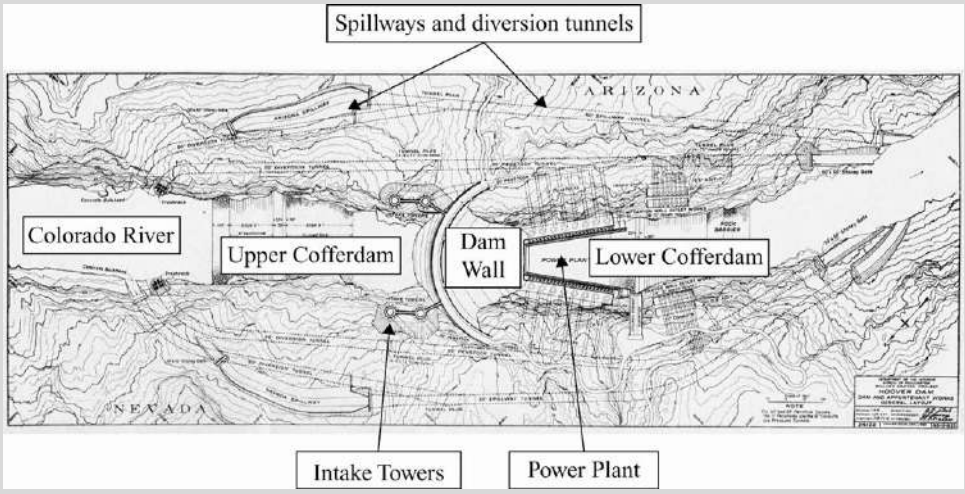


Figure 4: Plans

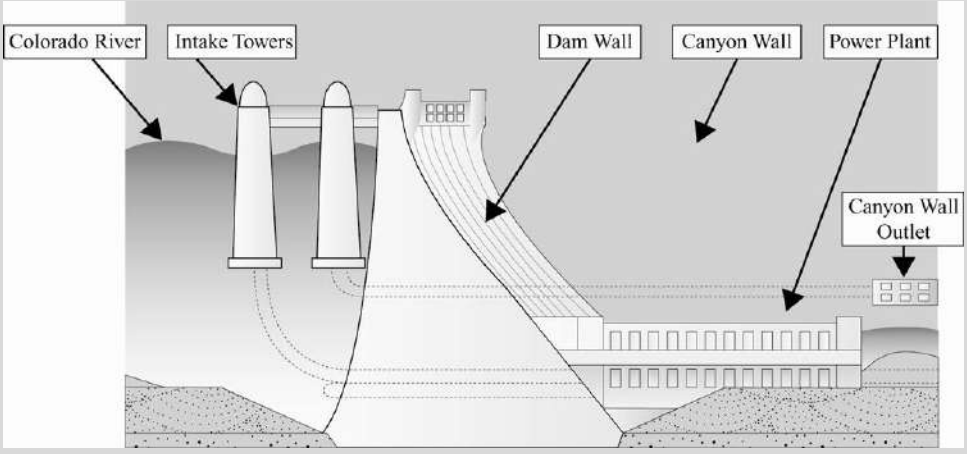


Figure 5: Plans

CRT1 1

Explain why this is a good dam site for hydroelectric power generation. You should discuss at least two aspects.

Scoring Note 1: The question requires students to explain, therefore responses should both list a feature AND provide an indication of why/how that feature makes the site suitable for the dam.

© OECD 2012

Scoring Note 2: Note the emphasis of hydroelectric power generation in the question.

CRTM11(a) Dam height / High potential energy

Code 1: Refers to the height of the dam OR the possibility of using the high potential energy to generate power.

- The dam is 221 metres high. This is possible due to the natural formation of the canyon. This will create an immense amount of kinetic energy when the water passes through turbines after transformation from gravitational potential energy.
- High potential energy from the dam height will be converted into high kinetic energy, which in turn spins the turbines, to generate electricity.
- The canyon formation allows for a large difference in gravitational potential energy. Consequently, there is a large potential for extracting energy.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM11(b) High flow rate of river (amount of water entering the dam)

Code 1: Refers to the beneficial geographic location of the dam site due to the large volume of water available flowing from the river OR the high flow rate of the river.

- This site is situated on a major river. This provides a large volume of water, which will allow for the creation of a large amount of electricity, which makes the site more economically justifiable.
- The geographical location of the dam site ensures a large annual flow rate, which allows for much electricity to be generated.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM11(c) Lake capacity

Code 1: Refers to the lake capacity AND identifies this as a benefit to power generation (must relate the size to a benefit to power generation).

- Large lake capacity enables an ability to shift volumes of water between different seasons so electricity can be generated all year round.
- Large water storage means that even in dry months, there should be water available for power generation.
- The lake capacity will enable the hydro-station to reliably produce its capacity output.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

- The lake has a very large capacity.

CRTM11(d) Minimal social impact

Code 1: Identifies that the dam location will have minimal social impact (people living on the dam site will not need to be relocated OR people living near the dam site will not be adversely affected by dam failure)

© OECD 2012

- There are no residences or human activity in the nearby area (so people don't need to be relocated).
- The surrounding area is dry land, with no human habitation so people would not need to be relocated.
- Because the dam is a considerable distance from a built-up area, the consequences of failure of the dam wall are minimised.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

- There are no residents nearby.

CRTM11(e) Characteristics of rock (could also include hardness and suitable foundation)

Code 1: Identifies that the rock in the canyon is water tight.

- The rock is impermeable so it traps the large body of water in the lake.
- If the rock was highly fractured you may lose the water through seepage.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM11(f) Narrow gorge

Code 1: Identifies that the shape of the gorge allows for efficient construction of the dam.

- The geographical shape of the canyon makes this an appealing location for dam construction as costs would be minimised.
- A small volume of material would need to be moved during construction of the dam due to the narrow gorge.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRT1 2

Explain the two main design features that contribute to the structural strength and stability of the Hoover dam.

Scoring Note: The question requires students to explain, therefore responses should both list a feature AND provide an indication of why/how that feature makes the site suitable for the dam.

CRTM12(a) Arch-shape

Code 1: Identifies that the arch-shape of the dam provides stability by directing the loads into the two sides of the canyon.

- The arch shape of the dam places the loads into the two sides, that is, the rocky material on either side of the canyon. The arched structure improves the stability as it diverts the forces elsewhere.
- The particular shape of the dam (convex) allows the transfer of axial compressive loads to the surrounding walls.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

© OECD 2012

CRTM12(b) Material in canyon

Code 1: Identifies that the material in the canyon walls must be robust enough to take the loads exerted.

- Due to the loads being directed into the valley wall, a strong valley wall is essential for stability.
- The rocky nature of the gorge, and the fact that it is narrow, allows for the load to be distributed into the rock-mass.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM12(c) Weight of concrete

Code 1: Identifies that the sheer weight of the concrete in the dam increases stability.

- The sheer weight of the concrete dam itself adds further stability. Thus, the dam wall may not need to be as strong as other Arch dams. This dam functions as an Arch and Concrete gravity dam simultaneously.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM12(d) Tapered shape of concrete wall/low centre of gravity

Code 1: Identifies that the tapered shape of the concrete wall keeps the centre of gravity low.

- The dam shape mimics the hydrostatic stress distribution, which is triangular.
- The dam wall is thinner at the top compared with the bottom. This makes the centre of gravity low and stabilises the entire dam.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM12(e) Spillways and/or tunnels

Code 1: Identifies that the spillway diverts flood flows past the dam wall, so that the dam is not over-topped AND/OR that that the tunnels are used for controlled releases or to divert flood flows past the dam wall.

- Spillways help floods to flow past the dam wall.
- Diversion tunnels help floods to flow past the dam wall.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRT1 3

The maximum electrical power generated by the turbines at the Hoover Dam is 2.08×10^9 W.

What is the approximate amount of water that flows through the turbines at this power output, if the power station operates at 90% efficiency?

Note: 1 W = 1 J/s, where 1 J = 1 Nm.

© OECD 2012

A. 10^3 L/s

B. 10^4 L/s

C. 10^6 L/s

D. 10^7 L/s

Key

Code 1: C.

Code 0: Other responses.

Code 9: Missing.

CRT1 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.**

Note: The guide assumes that changes to the river flow are understood. Candidates need to provide indications of the consequences of changes to the flow.

CRTM14(a) Habitats

Code 1: Identifies that plants and/or animal species' habitats will be affected.

- Plant systems may be affected both upstream, due to increased water storage, or downstream, due to changes in the flow of a river.
- Animal habits will be affected by construction of a new dam upstream and downstream with loss of habitat and/or changes in flow regimes.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(b) Soil and/or siltation/erosion

Code 1: Identifies that soil conditions will be affected upstream or downstream of the dam.

- Soil conditions may change over time.
- The dam may stop fertile soil from flowing downstream, impoverishing riverbank areas.
- It is likely that siltation will occur upstream.
- Because sand will accumulate within the dam and will not be able to move downstream, the riverbed, the coast, and the beach will be eroded

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(c) Ground stability around the storage itself

© OECD 2012

Code 1: Identifies that the dam may have effects on the stability of the ground at the dam site

- The rock at the dam will have increased water content and could become unstable in parts resulting in landslides
- Increased rate of earthquakes around large water storages.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(d) CO2 Emissions/Greenhouse gases

Code 1: Identifies that building the dam will result in high levels of greenhouse gas emission.

- The concrete used in building the dam wall will result in large volumes of CO_2 emissions.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(e) Aesthetics

Code 1: Identifies that there may be some aesthetic impact.

- The dam will impact scenic beauty.
- The native wilderness will be impacted.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(f) Effluent impact

Code 1: Identifies that there may be some effluent impact (must relate to water quality)

- Inflow of household and/or rural effluents may cause over-nourishment of water within the dam.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM14(g) Community impact

Code 1: Identifies that there may be some impact to the community. This could include existing water users in the river. Also include flood impacts, e.g. very heavy rainfall leading to extreme dam releases.

- New residential areas will need to be developed in order to relocate people residing at the dam site.
- Could impact upstream on farmers.
- Noise from the dam may disrupt the life of people near the dam.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

© OECD 2012

CRT1 5

A taller arch dam, built in 1959 and now disused, is the Vajont Dam in Italy.

In 1963, a massive landslide occurred upstream of the dam, and some 260 million cubic metres of earth and rock fell into the reservoir. Over 2000 lives were lost as water spilled over the dam and caused massive flooding in the valley below.

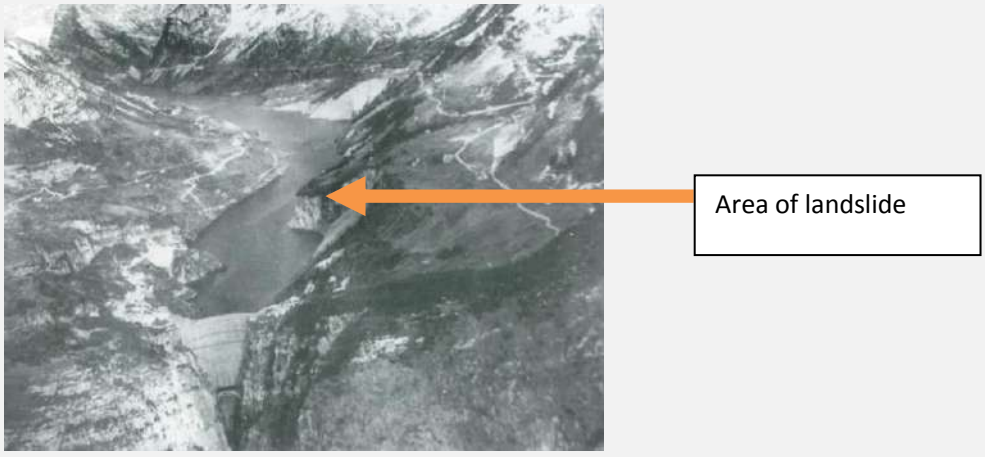


Figure 6: Before the landslide

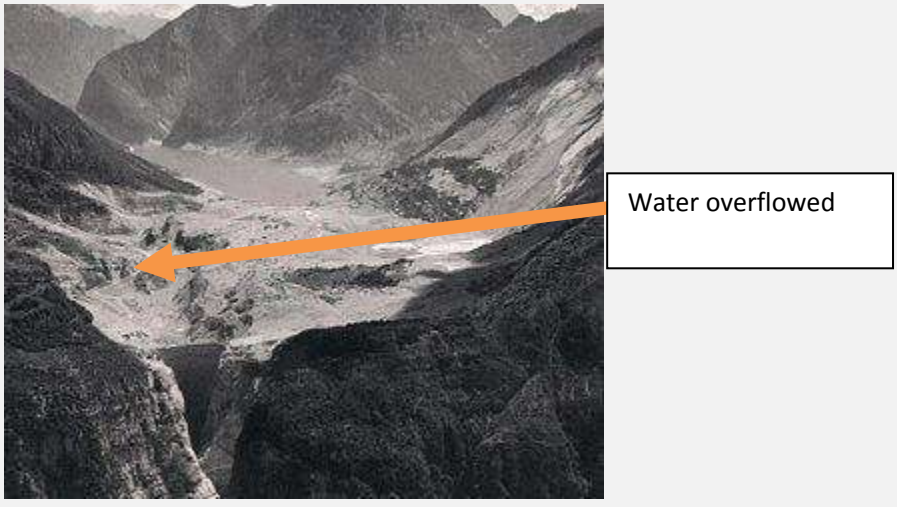


Figure 7: After the landside

© OECD 2012

Briefly explain two geo-technical assessments you would expect to have been done before the dam was constructed.

CRTM15(a) Material assessment

Code 1: Refers to an assessment of the material in the vicinity (must specify outcome measure or name of test).

- Geotechnical soil and rock assessments would have been carried out, such as penetration or laboratory tests such as triaxial test or shear box tests.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM15(b) Stability assessment

Code 1: Refers to an assessment of the stability of the slope or hillside (typically refers to some form of calculations)

- A geotechnical slope stability analysis of the surrounding hills to assess the effect of increased water mass on the soil, for example, under rapid drawdown effect.
- Investigating the impact of additional loads from the new dam on the hillsides.
- Perform a stability analysis, such as a slip circle analysis.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM15(c) Geological assessment

- Code 1: Refers to a geological assessment of the region (does not require reference to calculations, rather refers to mapping-style investigations)
- Analysing the geological setting and other detailed geological investigations may have indicated that a landslide was possible.
- Looking at the formation of the valley and investigating the possible failure modes. For example, existing fault lines.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM15(d) Hydro-geological assessment

Code 1: Refers to a hydro-geological assessment of the region.

- Investigation of water table and groundwater, which might affect stability of the slopes.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRT1 6

Consider the following scenario:

© OECD 2012

Results of the geotechnical analyses performed before the dam is constructed indicate that there is the potential for a serious slope failure risk, but with very low probability of occurrence. The engineering team is presented with two options:

(*i*) Proceed with the current design, on the basis that the serious problem is unlikely to occur;

(*ii*) Re-design the dam, which will incur greater cost, but reduce the risk of failure.

Discuss two aspects of engineering practice that favour option (*ii*).

CRTM16(a) Ethics

Code 1: Refers generally to the professional responsibility of an engineer to act in the best interests of the community.

- The engineering code of ethics states that the responsibility of the engineer is to the welfare of the community before the employer and the profession (of engineering).
- An engineer must alert others to the potential for danger even if this will compromise his job.
- An increase in cost is not as important as the potential for serious failure. An engineer must act in an ethical manner in this situation.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM16(b) Safety/Risk

Code 1: Refers explicitly to engineers' responsibility to ensure safety or minimise risk.

- The engineer has the responsibility/duty to minimise the potential for dangerous situations arising. Safety must be considered in engineering design and implementation through a formal risk assessment.
- The safety of the community is paramount.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM16(c) Action

Code 1: Refers explicitly to the need to make sure that something is done (may be through follow-up or confirmation of implementation).

- It is important that the engineer ensures that their advice is acted upon. The engineer does not absolve themselves of responsibility by merely telling somebody else. They need to make sure that suitable action is taken to ensure the safety of the community.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM16(d) Communicating Risk

Code 1: Identifies that the role of the engineer is to communicate the importance of option (ii) to other parties.

© OECD 2012

- If the engineer deems that the site is unsafe or the dam must be re-designed, it is their responsibility to convey this to all interested parties.
- Team members may be unhappy with increased costs and project delays. However, it is essential for the engineering team to highlight the catastrophic risks, even when the probabilities are low.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRT1 7

After construction of the Vajont Dam, and recognising the possibility of hillside failure, outline two planning measures an engineer could have suggested to minimise potential harm to people.

CRTM17(a) Evacuation procedures

Code 1: Refers generally to the implementation of evacuation procedures / warning system (can include communication of these procedures)

- Implement a warning system, whereby increased geological movement could trigger an evacuation alarm so that the town is warned of the possible disaster. Such warning is commonplace for large earth and rock-fill dams for instance.
  Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(b) Town planning

Code 1: Refers generally to town planning initiatives.

- In a location such as this, the town could have been encouraged to move itself away from the river over time. A 'danger zone' could have been declared so that no new buildings were built over the next 20 years, for example. This is linked to town planning possibilities.
- Relocation of the entire town over a 5 year period could be planned. Although this is not appealing, it may be the only way to ensure safety to the population.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(c) Town protection

Code 1: Refers generally to town protection measures.

- A safety wall could be erected which is designed to withstand any possible overflow from the dam and divert water elsewhere. Major planning and implementation would need to be carried out to ensure that such a measure was viable.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(d) Monitoring

Code 1: Refers generally to monitoring the operation of the dam.

© OECD 2012

- Set up a team of experts who can advise regarding dam operation in the event of imminent dam failure.
- A monitoring program could detect possible dam failure (or slope failure) in the case of the Vajont Dam.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(e) Operation

Code 1: Refers generally to changing the utilisation (operation) of the dam (including abandoning/dismantling the dam)

- Reduce the amount of water in the reservoir by releasing water from the dam. Although this will negatively affect the dam's functional efficiency, it will reduce the risk of massive flooding caused by landslide in the upstream areas.
- Implement a system whereby the dam capacity is reduced prior to the wet season (the likely period of failure).
- Dismantlement/abandonment of the storage

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(f) Communication plan (public education) regarding the risks

Code 1: Make people aware about the possible danger of the dam.

- Educate the public through some form of community engagement so that they are aware of the risks of the dam.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

CRTM17(g) Strengthening/reinforcement of the dam wall and/or the hillside

Code 1: Refers generally to strengthening the dam wall or hillside.

- Rock bolt the embankment into the bedrock.
- Rock bolt the hillside to increase stability.

Code 0: Other responses, including vague, incoherent, or irrelevant responses.

Acknowledgements

Figure 1: Natural Resources Conservation Service, United States Department of Agriculture. Photographer: Lynn Betts.

Figures 2 and 3: Los Angeles Times photographic archive (1921), UCLA Library.

Figure 4: Herbert Hoover Presidential Library and Museum, Contour map of Hoover Dam (1930), National Archives and Records Administration.

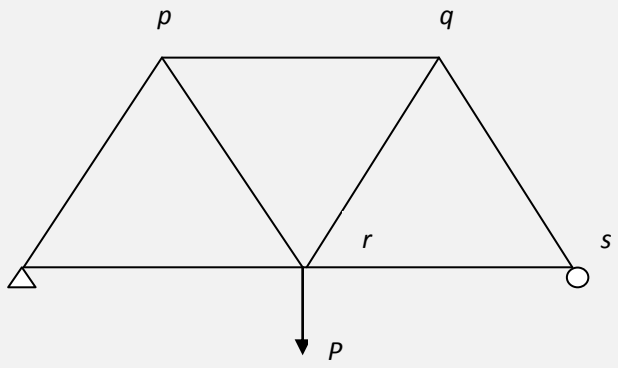
Figures 6 and 7: Semenza, E (1965) Sintesi degli studi geologici sulla frana del Vaiont dal 1959 al 1964. Mem Mus Tridentino Sci Nat 16: 1-52.

© OECD 2012

Examples of Engineering MCQs

MCQ 1

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?

A. Compressive force exists in both the upper-chord member (*p*-*q*) and the lower-chord member (*r*-*s*).

B. Tensile force exists in both the upper-chord member (*p*-*q*) and the lower-chord member (*r*-*s*).

C. Compressive force exists in the upper-chord member (*p*-*q*), while tensile force is applied to the lower-chord member (*r*-*s*).

D. Tensile force exists in the upper-chord member (*p*-*q*), while compressive force is applied to the lower-chord member (*r*-*s*).

Key

	Description: *Identifies the forces on a Warren truss when a load is applied.* Competencies: *BES iii.* *Demonstrates: comprehensive knowledge of their branch of engineering including emerging issues.* *Specialised area: Structural Engineering, including: structural statics; earthquake engineering; maintenance management.*

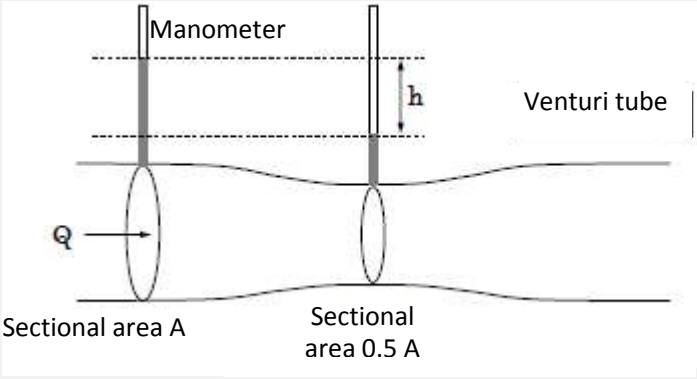
© OECD 2012

Code	Description of Response
1	C.
0	Other responses.
9	Missing

MCQ 2

Two manometers are attached to a Venturi tube as shown below.
Select the correct formula that expresses flow rate *Q* in the pipe channel.

Assume that *g* represents the gravitational acceleration and that energy loss is negligible.



A. $Q = \sqrt{\frac{2}{3}} A\sqrt{gh}$

B. $Q = \frac{2}{3} A\sqrt{gh}$

C. $Q = \sqrt{2} A\sqrt{gh}$

D. $Q = \frac{\sqrt{2}}{3} A\sqrt{gh}$

Key

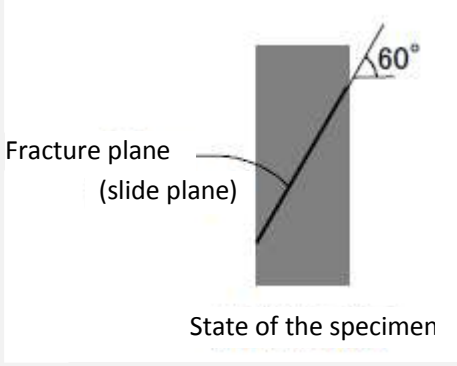
	Description: *Identifies the correct expression for flow rate in a Venturi tube.* Competencies: *BES iii.* *Demonstrates: comprehensive knowledge of their branch of engineering including emerging issues.* *Specialised area: Hydraulic Engineering, including water engineering*

© OECD 2012

	and management; design of components and systems such as water supply systems and sewer networks.
Code	Description of Response
1	A.
0	Other responses.
9	Missing

MCQ 3

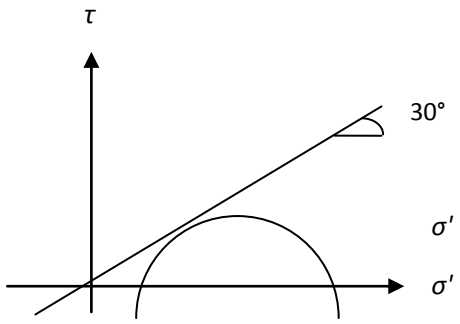
During a consolidated-drained triaxial compression test on some clay, a fracture plane tilting at 60° from the horizontal plane was produced in the specimen as shown in the figure below.



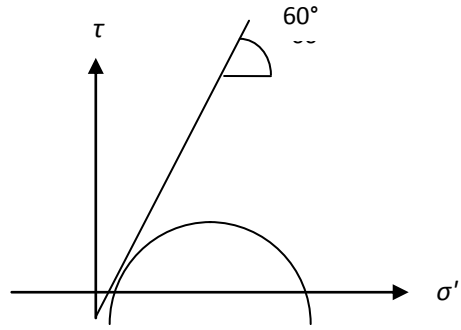
Select the most appropriate illustration that expresses Mohr's circle and Mohr-Coulomb's failure criterion line at the time of fracture of this clay.

© OECD 2012

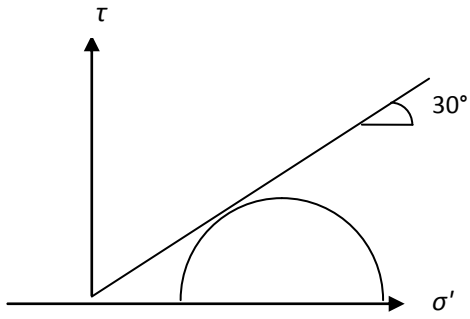
A



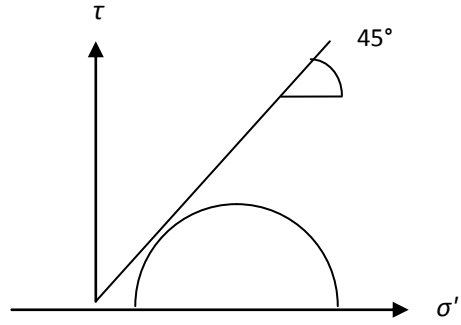
B



C



D



Key

	Description: *Identifies the correct version of Mohr's circle and Mohr-Coulomb's failure based on a slide plane.* Competencies: *BES iii.* *Demonstrates: comprehensive knowledge of their branch of engineering including emerging issues.* *Specialised area: Geotechnical Engineering, including: foundation engineering; dam and tunnel remediation & construction; slope stabilisation.*
Code	Description of Response
1	A.
0	Other responses.
9	Missing

© OECD 2012

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

© OECD 2012

www.oecd.org/edu/ahelo

Over the past 5 years, the OECD has carried out a feasibility study to see whether it is practically and scientifically feasible to assess what students in higher education know and can do upon graduation across diverse countries, languages, cultures and institution types. This has involved 249 HEIs across 17 countries and regions joining forces to survey some 4 900 faculties and test some 23 000 students.

This report presents the design and implementation lessons learnt from this unprecedented experience, as the AHELO Feasibility Study concludes in December 2012. A second volume will be published in February 2013 that will delve further in the analysis of the data and national experiences, while a third volume to be published in April 2013 will present the discussions and insights from the AHELO Feasibility Study Conference (taking place in March 2013).

Contents

Chapter 1 – The rationale for an AHELO: higher education in the 21st century context
Chapter 2 – The beginning of AHELO: decisions and challenges
Chapter 3 – Design and management of the feasibility study
Chapter 4 – Instrument development
Chapter 5 – Implementation
Chapter 6 – Lessons learnt on design and implementation

More information on www.oecd.org/edu/ahelo
Contact us: ahelo@oecd.org

Cover photo © jun.SU./ Shutterstock