



ROADMAP FOR THE OECD ASSESSMENT OF HIGHER EDUCATION LEARNING OUTCOMES (AHELO) FEASIBILITY STUDY

As of 3 July 2008

TABLE OF CONTENTS

Background context.....	2
AHELO goals and challenges	2
Strategic goals	2
Challenges to address in AHELO.....	3
Why the OECD?.....	3
The AHELO approach and its rationale	4
A multi-dimensional definition of quality.....	4
TEIs as units of analysis.....	5
A target population of students at the end of their undergraduate degree	5
Reporting at institutional level to permit a wide substantive coverage with limited response burden	6
The AHELO feasibility study	6
Expected outcomes	6
Several strands of work reflecting several aspects of quality.....	7
Maximising insight from a broad geographic and linguistic coverage in a manageable way	8
A transparent and collaborative process.....	8
A cooperative funding mechanism	9
Proposed activities in the different strands of work	9
Generic skills strand: international pilot test of the Collegiate Learning Assessment (United States) ...	9
Discipline strand.....	14
Value-added measurement strand.....	15
Contextual strand.....	16
References.....	16

1. The purpose of this paper is to describe the roadmap for work to be undertaken as part of the Feasibility Study for an OECD Assessment of Higher Education Learning Outcomes (AHELO). This roadmap was put together following the advice of experts that met on three occasions in 2007 to provide guidance for a strategy [see EDU(2007)8, EDU(2007)9 and EDU(2007)14]. Planning for the AHELO feasibility study is now underway, although still at early stages for some of the strands of work involved. Hence the roadmap reflects current thinking and will be expanded and updated over time as the planning unfolds.

Background context

2. Following decades of rapid growth in tertiary education numbers of students and institutions and its growing internationalisation, there is increasing recognition that greater attention should be paid to quality and relevance. Policymakers as well as the public devote considerable attention to the outcomes of tertiary education given its importance for human capital development, its cost to public finances as well as to students and their families, and the needs of business and industry. Developing measures that give due weight to teaching and learning has thus become essential.

3. For some aspects of tertiary education outputs and outcomes, such as the type and number of degrees awarded, the research outputs produced or the labour-market returns to tertiary education studies, numerous indicators exist at both national and international levels. However, for other aspects, and most notably for the learning outcomes of tertiary education, available data are very limited (see Nusche, 2007 for a summary of existing assessments). This topic attracted particular attention from OECD Ministers of Education at their Athens meeting in June 2006 and Ministers invited the OECD to advise on how to better understand and evaluate the various dimensions of quality in tertiary education.

4. In response, the OECD embarked on a series of three experts meetings in 2007, gathering top international specialists to define the scope of the task and discuss options for data development. The latter noted that a few countries have developed national instruments to measure, assess, and report on student achievement and institutional performance outcomes, yet no such instruments exist that allow international comparisons to be made on learning outcomes at institutional level. Nevertheless, rankings/comparisons of universities were being made that attracted considerable attention. As a result, the experts advised that it was desirable to launch a feasibility study to explore the scope for developing an international Assessment of Higher Education Learning Outcomes (AHELO) that would measure learning outcomes in ways that are valid across cultures, countries and languages¹. The OECD is now working on the planning of this feasibility study with the support of both governments and tertiary education institutions (TEIs), to determine by the end of 2010 whether such an assessment is scientifically and practically possible.

AHELO goals and challenges

5. The purpose of the AHELO feasibility study is to assess whether it is possible to measure what undergraduate degree students know and can do at the international level, in order to provide better information to TEIs, governments, and other stakeholders including students and employers.

Strategic goals

6. The underlying motivation for an AHELO is that this information could contribute to TEIs' knowledge of their teaching performance, and thereby provide a tool for development and improvement. As such, the AHELO central emphasis is on the improvement of teaching and learning and in providing tertiary education leaders with tools to empower them and foster positive change and enhanced learning.

¹ See para. 23.

7. The feasibility study has two key aims. First, to test the science of the assessment – whether it is possible to devise an assessment of the outcomes of tertiary education which enables to make reliable statements about the performance/effectiveness of learning in TEIs of very different types, and in countries with different cultures and languages – and second, to test the practicality of implementation and of motivating TEIs and students to participate. In addition, the feasibility study will involve related work exploring other options for capturing measures of tertiary education quality indirectly.

Challenges to address in AHELO

8. The development of an AHELO presents a number of scientific and practical challenges. It is of crucial importance that an assessment has both reliability and validity. Constructing an assessment that is valid across TEIs, cultures and disciplines implies taking into account:

- The diversity of TEIs;
- Differences between national systems of tertiary education;
- The absence or presence of selection in the system or certain TEIs;
- Variations in the duration and content of programmes; and
- Cultural and linguistic diversity.

9. Practical and operational challenges that also have to be addressed are how to motivate students and TEIs to participate and ensuring a fair assessment of the results.

Why the OECD?

10. The OECD is well-placed to lead and implement the feasibility study given its institutional framework for cooperative and international comparative work, as well as its credibility and demonstrated expertise in developing large scale international assessments such as the Programme for International Student Assessment (PISA) at school-level and the Programme for International Assessment of Adult Competencies (PIAAC) which focuses on the skills of the adult workforce.

11. Within the OECD, the Programme on Institutional Management in Higher Education (IMHE) provides a platform for the engagement of TEIs with Governments in AHELO in a way which will ensure that the approaches adopted take account of institutional needs and concerns. IMHE is open to all recognised TEIs in OECD countries, as well as associations of TEIs and representatives of Governments. It can therefore provide a locus for consultation and steering of the feasibility study which also allows countries which are not directly involved in it to have their say and monitor progress.

12. The feasibility study will be managed by a relatively small team of analysts and senior staff at the OECD. However, the bulk of the technical work involved in the different strands of the AHELO feasibility study work will be carried out through collaboration with a number of contracted outside experts, consultants and organisations given the scope of the task as well as the high level of highly-specialised expertise required. A number of countries and organisations have already expressed interest in participating and/or supporting this feasibility study work in-kind or through voluntary contributions. These include TEIs' associations, quality assurance agencies, foundations as well as countries interested in taking part in the feasibility study. These organisations will be invited to support the various strands of the feasibility study while countries taking part in the different strands of the feasibility study will contribute through voluntary contributions as well as national implementation costs.

The AHELO approach and its rationale

A multi-dimensional definition of quality

13. Measures of tertiary education learning outcomes hold out important promises for students, TEIs and public policy more generally. The extent to which the information needs of the various stakeholders can be met will depend on the units from which data are collected and for which outcomes are reported. The approach underlying the AHELO feasibility study considers these units in the form of a three-dimensional matrix:

- The first dimension of the matrix comprises users, *i.e.* the stakeholders who would be served and their information needs. These include: *i) individuals*, such as students wishing to make better-informed choices or employers seeking to benchmark qualifications against actual skill measures; *ii) TEIs, departments or faculties* seeking a better understanding of their comparative strengths and weaknesses; and *iii) public policymakers* seeking to quantify stocks and flows in high-level skills to obtain better insights into the quality, equity and efficiency of tertiary education services and to assess the impact of policy decisions.
- The second dimension of the matrix relates to the use of the measures on learning outcomes, ranging from summative comparisons of institutional performance to diagnostic tools at programme and faculty level.
- The third dimension of the matrix considers the kind of instruments needed to serve the respective users and uses.

14. Recognising the multiplicity of potential uses and users, the OECD does not aim for the establishment of a single performance measure that would then be used for a uni-dimensional ranking of TEIs or countries. The OECD also acknowledges that any effort to bring together all TEIs on one standard, would risk driving the assessment down to the lowest common denominator. The AHELO aim is rather to establish a “multi-dimensional quality space”, in which quantifiable criteria for quality establish the dimensions of the space. Within this concept of the “quality space” tertiary education systems, TEIs, departments and faculties can then be situated depending on the prevalence of the different quality attributes. Consumers would then be able to choose programmes and TEIs depending on the configuration of the quality attributes that are 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. It would also become possible to portray policy trajectories of TEIs and systems over time, as they change their position on the different dimensions of this “quality space”, which in turn could become a powerful tool for public policy.

15. For instance, two kinds of outcomes measures can be considered at the level of TEIs or departments, both of which the OECD is committed to explore to meet the information needs of the various groups of stakeholders:

- Individuals, whether prospective students or employers, would want to know the “bottom line” of the performance of TEIs, departments or faculties, in terms of the absolute performance or raw scores students enrolled in these would attain on an assessment, recognising that such an assessment would not only measure the quality of educational services provided, but also other aspects such as the effects of selection and the socio-economic makeup of students.
- Meanwhile, individuals, TEIs and public policymakers wishing to assess the quality of the teaching services provided by TEIs would primarily be interested in the “value-added” or “gain”

provided by the TEIs, *i.e.* the scores an institution would attain after accounting for the quality of prior schooling or the degree of selectivity of the programmes and TEIs.

16. The OECD also recognises that learning outcomes are only one component of the quality of TEIs, and that the “quality space” would need to recognise other dimensions of the quality of outcomes as well, such that TEIs and systems could be appropriately represented in the space, in accordance with their respective missions. Other dimensions for which various instruments have been established include curriculum and consumer ratings. Alumni ratings, such as used by the German Centre for Higher Education (CHE), are considered one way to complement the assessment of competencies linked to “employability” with actual labour-market outcomes. Institutional factors, non-cognitive characteristics that are known to be tied to successful study and achievement, and measures of institutional efficacy are also seen as relevant. Last but not least, there is a wide area of direct and indirect measures of research outcomes that can be utilised to complement the work undertaken by the OECD.

TEIs as units of analysis

17. For several reasons, the OECD does not consider it feasible to develop internationally comparative information on tertiary education learning outcomes at the system-level.

- First of all, any system-level assessment would have to confront issues of differentiation within the system, as variation in institutional structures challenges the establishment of cross-nationally comparable classes of TEIs.
- Second, in many countries governments have limited options to incentivise the participation of TEIs in assessments, particularly those carried out at an international level.
- Third, even if it were possible, large differences in enrolment rates would raise questions about how to interpret results since, other things being equal, it would only be natural that those countries with more selective systems and lower enrolment rates would likely perform better than countries where tertiary education has become largely universal.
- Last but not least, mandated assessment in the form that would be required to obtain a system-wide representative sample of TEIs, may lead to administrative response, not faculty response and would thus not be effective as a tool for improvement at the level of service provision.

18. The AHELO approach therefore centres on the establishment of measures of learning outcomes at the level of TEIs, departments or faculties, the idea being to combine the definition of OECD measures of quality with reliable assessment methods to which TEIs could, with an appropriate set of incentives, voluntarily subscribe and which could progressively find acceptance in a widening range of TEIs. Eventually, if an OECD assessment gathered pace and found wide acceptance, issues of assessing system-level performance might be addressed, and this would make the work relevant to a much wider range of stakeholders in the longer term.

A target population of students at the end of their undergraduate degree

19. The OECD has devoted consideration to what would constitute meaningful target populations for an international AHELO. One possibility, that would facilitate comparisons across TEIs and countries, would be to focus the assessment on comparable age bands. However, this would make it very difficult to link results to national degree and qualification structures and thus make it difficult to interpret them in the national institutional context. Such an approach would also be very difficult to implement, as it would require the selection of age-based samples that may spread widely across years of study. The approach

therefore considers an assessment towards the end of a three or four-year degree as a more practical solution for the feasibility study.

Reporting at institutional level to permit a wide substantive coverage with limited response burden

20. The OECD recognises diverging views on how the knowledge about learning outcomes in TEIs can and should be used. Some see such information primarily as a tool to reveal best practices, to identify shared problems among TEIs and to encourage collaboration and lateral capacity building among research and teaching personnel. With this approach, emphasis would be placed on the relevance of performance information for the institutions themselves and on contextualising performance data with other information on the learning environment in TEIs. Other views extend 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 TEIs publicly available to facilitate choice.

21. Irrespective of where the focus would eventually be placed in different national contexts, an AHELO would not seek to report data at the level of individual students. This approach has important technical implications since it allows the use of matrix sampling techniques² through which the coverage of the assessment areas can be widened without unduly extending the response time demands on individual students.

The AHELO feasibility study

Expected outcomes

22. There is currently considerable interest within institutional, political and scientific circles for measures of tertiary education learning outcomes, but uncertainties and doubts on the part of some actors as to whether it is scientifically and operationally feasible to measure learning outcomes across TEIs of very different types, and in countries with different cultures and languages. In order to answer this question, a number of international experts have been consulted over the past 18 months. Three meetings were held in 2007, bringing together international specialists in the field³. The main conclusion of the experts was that while it might be both desirable in terms of public policy and theoretically possible to assess and compare central components of education outcomes, it would be necessary to conduct a feasibility study to test this proposition before undertaking any more systematic assessment. The feasibility study would have to do two things:

² This approach allows significant gains in testing time and is intended to overcome the conflicting demands of limited student testing time and broad coverage of assessment domains. The difficulty of test items are linked to the abilities of respondents (via a Rasch model) by a logistic function which makes it possible to compute the probability that a student with a given level of ability succeeds on an item. The Rasch model thus creates a scale on which every item and every respondent can be located. Due to this probabilistic link, it is no longer necessary to administer the whole battery of assessment items to every respondent, and it is possible to assign a subset of the items pool to certain sub-samples of students provided that some link items are administered to several sub-samples of students. In this way, the performance levels attained by respondents on their subset of items provide a reliable estimate of their performance had they responded to the whole battery of items.

³ The expert meetings took place in Washington (28 April), Paris (5-6 July) and Seoul (26-27 October). The Washington meeting was primarily about the usefulness and desirability of an OECD international assessment of higher education learning outcomes, the Paris meeting focused on the conceptual possibility, and the Seoul meeting concerned how to move from possibility to feasibility. See www.oecd.org/edu/ahelo for the summary records and lists of participants from these meetings.

- **Test the science of the assessment** – whether it is possible to devise an assessment of the outcomes of tertiary education which enables to make reliable statements about the performance/ effectiveness of learning in TEIs of very different types, and in countries with different cultures and languages; and
- **Test the practicality of implementation** and of motivating TEIs and students to take part in such a test.

23. The feasibility study is expected to demonstrate the feasibility – or otherwise – of comparing TEIs’ performance from the perspective of student learning rather than relying upon research-based measures which are currently being used across the globe as overall proxies of institutional quality. To this aim, it will implement several instruments at the international level. It is anticipated that the assessment will be based on a written test of the competencies of students who are almost at the end of a Bachelor programme, and will be computer delivered. Because of the national differences in academic year, a window of testing time would need to be allowed, giving flexibility to the TEIs to ensure that the timing does not coincide with major exams.

24. The main criteria to assess the success of the feasibility study would be to provide a proof of concept that the various instruments considered can be applied in diverse institutional, cultural and linguistic settings and yet provide valid, reliable and free-of-bias measures of student learning outcomes as well as indirect measures of tertiary education quality. With respect to implementation, success would be defined in terms of meeting minimum student participation and response rates thresholds.

25. The feasibility study evaluation plan will involve careful psychometric analysis, technical reviews by international experts as well as an international conference gathering different stakeholder groups to discuss whether and how to take the results from the feasibility study forward. In case of positive results, this proof of concept would constitute a key pillar for longer-term work as it would assist OECD countries in deciding whether to launch an AHELO at the international level.

Several strands of work reflecting several aspects of quality

26. Any assessment of tertiary education learning outcomes will need to define and operationalise criteria for what constitutes quality. The OECD acknowledges that there is no generally accepted definition of what tertiary education outcomes ought to be, but considers that there are promising ways to examine various facets of learning outcomes underway. The feasibility study will thus involve different kinds of activities to explore the feasibility of directly measuring or indirectly capturing various aspects of quality. Several strands of work are thus foreseen.

27. First, the feasibility study will consider different types of assessment instruments to measure learning outcomes directly, and explore their validity in an international context. In this respect, the OECD considers it important to cover both discipline-related competencies and transversal higher-order competencies manifested or required by particular disciplinary contexts. The feasibility study will thus encompass two assessment strands, which together would embrace a wide spectrum of learning outcomes: a **generic skills strand** and a **discipline strand**. But summative measures of learning outcomes are only one facet of quality, and the OECD is also committed to explore outcomes measures in terms of value-added to meet the information needs of the various groups of stakeholders. The feasibility study thus comprises a **value-added measurement strand**. However, it has to be recognised that assessing the value-added or marginal gain by TEIs imposes layers of complexity that, though theoretically well understood, are difficult to resolve in large scale assessments. Given the complexity of measuring marginal gain, the proposed approach will first explore the value-added perspective conceptually and through a thorough examination of potential data sources, methodologies and psychometric evidence. The pilot measurement

of value-added will then be incorporated into the design at a second stage. Finally, reflecting the OECD recognition of the need for a multi-dimensional approach to tertiary education quality, the feasibility study also includes a ***contextual strand*** of work that will explore the development of indirect measures and contextual information indicators at institutional level of the kind developed by the Centre for Higher Education (CHE) in Germany, which have already been successfully applied across borders in Europe. A more detailed description of each of the strands is provided below.

Maximising insight from a broad geographic and linguistic coverage in a manageable way

28. In order to maximise information gains while recognising the need to ensure a broad geographic and linguistic coverage as well as various practical and resource constraints the different strands of work will be carried out separately but coherently, and each will involve about 10 TEIs as diverse as possible from 3-4 countries. This number is large enough to assess the measurement properties of the various instruments but small enough to keep the process manageable and avoid that validity gains are sacrificed over efficiency gains.

29. This approach also allows the involvement of a larger number of countries and TEIs in the overall feasibility study, thereby contributing to a broader sense of ownership for the outcomes of the feasibility study.

A transparent and collaborative process

30. The support of both governments and TEIs is necessary for the success of this initiative. It is thus essential for the AHELO instruments to be perceived and used as a tool for improvement and that the work be led and steered jointly by the tertiary education sector and governments, making the process transparent to all stakeholders – irrespective of whether they are involved in the feasibility study or not.

Governance and management

31. To facilitate the joint steering of the AHELO feasibility study, its management is placed under the immediate supervision of the Governing Board of the OECD Institutional Management in Higher Education (IMHE) Programme which brings together both governments and TEIs. This mechanism allows countries which are not directly involved in the feasibility study to influence decision-making and to monitor progress. The IMHE Governing Board will, in accordance with the principles governing OECD work on education, report on the conduct and outcomes of the feasibility study to the Education Policy Committee which will bear responsibility to take decisions about whether and how to proceed with the work once the feasibility study is complete.

Transparency

32. Given the high level of interest for tertiary education quality assurance issues, international rankings and the AHELO initiative in relation to these, a communication plan is essential in ensuring transparency on the goals and progress of the feasibility study. The OECD is already engaging with a wide range of stakeholders in communicating the purpose and practicalities of the initiative (officials, TEIs, media, students). Interest is high and although the issues are complex the response from the sector is generally supportive.

33. Several approaches are foreseen to enhance the transparency of the feasibility study process. These include regular reports detailing progress, web communication with a Website dedicated to the work [www.oecd.org/edu/ahelo], and participation of the OECD Secretariat in major conferences to reach relevant audiences. Dissemination will also take place as part of the IMHE conferences.

34. In addition, a final conference will be organised to disseminate the outcomes of the feasibility study and to inform subsequent decision-making. The targeted audience would bring together international experts, quality assurance experts, representatives of TEIs, students, businesses and policymakers. International experts will present the outcomes of the various strands of the feasibility study, and will report on the technical feasibility of developing an international AHELO. Workshops will also be organised to provide participants with a chance to interact and discuss operational issues, political economy constraints and the potential impact of such an assessment.

A cooperative funding mechanism

35. The cost of the feasibility study will be substantial (see Annex I). OECD staff involved in the management and conduct of the study is funded from available resources, but the bulk of the money for development work and implementation needs to be raised. Indeed, OECD Ministers are keenly interested in measures of learning outcomes, but given uncertainties as to the scientific and operational feasibility of such an assessment, the OECD needs to demonstrate that it is feasible to measure learning outcomes in valid and reliable ways across cultures, countries and languages before significant resources can be invested in the development of an AHELO as part of the regular OECD programme of work and budget.

36. Therefore, as is commonly the case for new cutting-edge OECD projects such as PISA or PIAAC, the initial steps of the AHELO feasibility study are to be funded from a range of grants and voluntary contributions. In light of the international experts' endorsement of the AHELO feasibility study, the OECD is now moving forward and seeking funds for the exploratory and development phase of this project. It is envisaged to seek support from foundations for development work and voluntary contributions from countries involved in the feasibility study for national implementation costs.

Proposed activities in the different strands of work

37. The different strands of work of the AHELO feasibility study are at different stages of planning, hence are described with varying degrees of detail. Subsequent updates of this roadmap will provide further information on the comparatively less developed strands of work.

Generic skills strand: international pilot test of the Collegiate Learning Assessment (United States)

38. The generic skills strand is an essential component of the feasibility study. Indeed, competencies in critical thinking, analytic reasoning, problem-solving, or the generation of knowledge and the interaction between substantive and methodological expertise are widely viewed as critical for the success of individuals and of rising relevance in the information age. It is therefore important for an AHELO to measure these transversal higher-order competencies that are necessary for success in both academic and business contexts – not only cognitive knowledge. A key advantage is that such competencies are largely invariant across occupational and cultural contexts and could be applied across TEIs, departments and faculties. Moreover, a focus on higher-order skills allows the coverage of a more diverse population representing the whole undergraduate student body whereas the discipline strands will only cover a subset of students enrolled in given disciplines.

39. International experts gathered in 2007 and reviewed the various initiatives taken in countries to assess tertiary education learning outcomes (Nusche, 2007). They were impressed with the Collegiate Learning Assessment (CLA) approach taken by the Council for Aid to Education (CAE) in the United States. The CLA is a national effort to assess the quality of undergraduate education by directly measuring student learning outcomes. It measures student ability and learning in critical thinking, writing, and synthesising quantitative and qualitative data. The measures are focused on skill sets that students will need as they graduate and enter the work force, namely critical thinking, analytical reasoning, problem-

solving and written communication. These skills are intertwined. Thus the CLA measures are holistic: they require students to use these skills together to respond to tasks. All CLA measures are administered online, using open-ended prompts that require constructed responses. Each task also has an accompanying library of information which students are instructed to use in preparing their answers. Tasks often require students to marshal evidence from these diverse quantitative and qualitative sources and exercise judgment on their relevance. All tasks are appropriate for students across a wide range of undergraduate academic majors and general education programmes.

40. A major strength of the CLA approach is to adjust student performance for the level of ability of entering students – measured by SAT and ACT college-entrance exams or a Scholastic Level Exam for freshmen students – thereby permitting to grasp institutional value-added net of the cognitive and socio-economic characteristics of the student intake.

41. As a result, the experts recommended, as part of an AHELO feasibility study, to implement an international pilot test of this instrument to assess the extent to which higher order skills of the type measured by the CLA can be validly measured across different cultural, linguistic and institutional contexts. ***The generic skills strand will therefore look at the CLA in terms of validity, cross-cultural appropriateness and linguistic transferability.*** As with PISA this direct assessment of student knowledge and ability will be complemented by contextual information which will then enable policy- and practice-related conclusions to be drawn at institutional level.

42. With respect to cultural adaptability, existing studies carried out in the United States using the CLA data suggest that the performance of international students enrolled at TEIs in the United States does not differ markedly from that of domestic students, thereby suggesting that the CLA instrument would not create major cultural adaptability bias. Yet, this assumption needs to be tested more systematically, by analysing the performance of students from various national groups in TEIs of the United States and by testing out the CLA instrument in TEIs from a couple of non Anglo-Saxon countries – preferably reflecting diverse linguistic and cultural backgrounds. The main point would be to have – in addition to the United States for which data from TEIs participating in the CLA can be used – one or two Asian societies, cultures and languages represented and one or two non English-speaking European countries represented.

43. It is also envisaged – as part of the cultural adaptability broader agenda – to organise an international item development workshop that would provide an opportunity for interested countries – mainly non English-speaking ones – to submit items suited to their cultural context for future use if an AHELO is to be implemented upon completion of the feasibility study.

44. With respect to linguistic transferability, the OECD experience with PISA and PIAAC provides a solid basis for adopting a similar approach in the feasibility study and for the international pilot test of the CLA.

45. The implementation of an international pilot test of the CLA in 10 TEIs from 3-4 countries will involve the following activities:

Planning (June-December 2008)

- Develop plan for piloting the CLA in 10 TEIs from 3-4 countries (other than the United States)

In 2007, international experts reviewed existing assessments of tertiary education learning outcomes and concluded that the CLA offered the most promising venue for an international assessment due to its focus on critical thinking, synthesising quantitative and qualitative information, analytical reasoning, problem-solving and written communication. Yet, the

application of the CLA instrument to very different national, cultural, linguistic and institutional contexts is not straightforward. It is important to note that the goal at this stage is not to compare TEIs' performance, but rather to provide a proof of concept that CLA-types of instruments can credibly be applied in different national and institutional settings. This can be achieved by piloting the CLA in a relatively small number of voluntary TEIs in each of a small number of countries, and collecting both cognitive and psychometric information on how well task items perform in a cross-national context.

- Design template for CLA performance task adaptation

The international adaptation of the CLA instrument will require restructuring some performance tasks and creating some new ones. This will involve a review of existing CLA measures by an international committee to be established by the OECD. The goal here will be to see to what extent the existing measures are regarded as valid across the 3-4 country set. Test validity involves the accumulation of evidence that supports inferences about the characteristics of individuals/groups. Therefore, in order to identify potentially problematic tasks ahead of the international committee meeting, an analysis of existing CLA data will be carried out in order to gather psychometric information contrasting the performance of domestic and international students from the various OECD countries on each task.

- Design schematic for international CLA results database
- Adapt Internet test delivery system for international delivery

Using computer-based assessments considerably reduces the organisational complexity of paper-and-pencil tests as well as test duration, and allows the production of reliable and valid measures in a minimal amount of time. For the international pilot test of the CLA, it is proposed to use the TAO⁴ platform which is being used by some 20 countries in the PISA 2009 "Electronic text Reading Assessment" option and is being explored for the delivery of the PIAAC study, meaning that national capacity has already been developed in several countries likely to take part in the feasibility study. TAO is a generic open-source platform that provides a broad range of functionalities for collaborative and distributive creation, management, and delivery of computer-based assessments. It has been applied in various test situations and offers adaptive testing possibilities as well as plugin and web service capabilities. This enables students to register on-line (through a web form) and take the appropriate tests in the sequel (triggered by a web service call to the TAO platform). Another strength of the TAO platform for the international pilot test of the CLA is that it supports multilingual tests and includes characters of non-European alphabets such as Japanese. TAO also allows the automatic scoring of short-text answers and includes a Hypertext Builder (HTB) as an authoring tool for creating innovative stimuli to assess problem-solving skills – *e.g.* simulating web sites and that can be used like them.

- Recruit pilot TEIs in 3-4 pilot countries

Participating TEIs will need to be sufficiently different to provide a cross section of the sector for each pilot country even though they will not be fully representative of the diversity of the sector. The recruitment of pilot TEIs in the 3-4 pilot countries will therefore seek to ensure that prominent TEIs are included along less well-known TEIs, and where appropriate, that the pilot

⁴ TAO (Testing Assisté par Ordinateur = computer-based assessment) is the French acronym for Technology-Based Assessment (TBA).

includes a mix of public and private or urban and rural TEIs. In countries with a binary tertiary education system, a couple of vocationally-oriented TEIs will be recruited as well.

- Recruit academic test development teams in each pilot country

Ultimately, the success of implementation of the CLA – just as for any international assessment – will critically depend on having a well-trained team in each pilot country. The recruitment and training of academic test development teams will therefore be an integral part of the planning phase.

- Meetings to design local pilot testing

Equally important will be to carefully plan local pilot testing, taking into account country and institution-specific constraints (*e.g.* exam dates). As a result, a series of meetings in pilot countries is foreseen as part of the planning phase so as to ensure that the pilot test implementation will run smoothly further down the road.

- Develop budget for subsequent Phases
- Develop criteria to assess success of the international pilot test

Decisions regarding the development of an international AHELO across OECD countries will depend on the outcomes of the feasibility study – including the international test of the CLA as part of the generic skill strand. In this respect, it is important that the criteria for assessing the success of the international pilot test be developed upfront, when planning the initiative. It has to be borne in mind that the ambition at this stage is not to compare TEIs' performance. Rather, success can be defined in terms of providing a proof that the CLA instrument can be applied in diverse institutional, cultural and linguistic settings and yet provide valid and reliable and free-of-bias measures of student learning outcomes. With respect to implementation, success will be defined in terms of student participation and response rates. But criteria for assessing success need to be more precise. The CAE, together with the OECD and international experts, will thus develop a list of technical criteria and threshold measures to assess the degree of success of the international pilot test of the CLA.

Development and piloting of a set of performance tasks for 3-4 pilot countries (January-March 2010)

- Country-specific Task Development

Each international team will develop one or two performance tasks that characterise both topics/issues germane to that society as well as incorporating “misconceptions” and distractors characteristic of CLA tasks, possibly adding “misconceptions” and distractors embedded in that culture. CAE technical staff will lead the teams on performance task construction.

To help facilitate this process, an international item development workshop will be organised. The CLA technical team will present the technical requirements for performance task development to the members of the committee and make them as knowledgeable as possible about the CLA performance tasks through the CLA in the classroom academy programme which is now up and running successfully throughout the United States. To meet the goal of ensuring that the performance tasks reflect a cross-linguistic and cross-cultural perspective, participants from countries other than those involved in the pilot will be invited so that a culturally diverse group of professional test developers will take part in the revision and writing of items. Such an

effort is expected, not only to increase technical knowledge at the national level, but also to help establish collaborative working relationships between countries participating in the pilot, CAE staff and other interested countries. Countries will be encouraged to share ideas and content for new tasks they would like to see included in the pilot (or towards an AHELO for other countries). The aim will be that participants suggest changes or alterations to the existing instrument to meet the needs of individual pilot countries, as well as develop a strong set of new items for field testing. The workshop will thus help to strengthen the validity of the item pools by increasing the range of linguistic and cultural representation. This workshop will also provide an opportunity to enhance communication and transparency about the feasibility study.

The CLA technical team will then meet with the pilot countries' teams to review their candidate performance tasks. These meetings will take place over a 3-4 day period to specify in as much detail as possible, several of the best candidate performance tasks. The surviving performance tasks will be pilot-tested in their home countries and refined based on findings. The pilot test of performance tasks will comprise 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 will be probed to identify particular steps that might have been problematic.

- Task Translation

Once the final set of performance tasks have been established and refined through pilot testing in their corresponding pilot country, the tasks will undergo translation into the languages of the pilot countries. Quality control during the translation process is an integral part of instrument development and is critical to help ensure that the international pilot test of the CLA provides data that are comparable across countries. To this aim, the translation process will be carried out using techniques and procedures that have been successfully implemented over the past three cycles of PISA. They consist of a double translation from the English source to the destination language, followed by a process of reconciliation, verification and a final layout check by the national team. The translation quality plan will not be limited to the performance tasks. It will also cover the context questions and the computer interface. Pilot countries will be provided with a procedure to verify the test delivery software in their language prior to the pilot test.

- Piloting of tasks in the 3-4 pilot countries

The tasks will be piloted in the 3-4 countries along with CLA short essay writing prompts (for calibration purposes), scored, and analysed particularly focusing on validity and reliability issues. The results will be reported back to the international committee. The pilot test will also seek to collect information on implementation – *e.g.* response rates, problems encountered *etc.*

Analysis of pilot results (April-September 2010)

- Analysis of pilot results

The pilot results will be analysed by psychometricians to test the validity, cross-cultural appropriateness and linguistic transferability of the CLA instrument. Outcomes of the pilot test will be benchmarked against the technical criteria and threshold measures defined during the planning phase to assess the degree of success of the international pilot test. This analysis will also be reviewed by a group of international experts to provide conclusions on the technical feasibility of adopting a CLA-type of instrument across the OECD.

- Reporting to IMHE Governing Board and Education Policy Committee on the feasibility of adoption across OECD countries to inform decisions on next steps by the end of the 2009-10 biennium

The views of international experts on the feasibility of adoption across OECD countries will then be conveyed to the IMHE Governing Board and Education Policy Committee as background information to inform decisions on next steps from the OECD perspective.

Discipline strand

46. A discipline strand is intended to complement the generic skills strand. Indeed, the main limitation of an approach entirely limited to generic competencies, is that it would not assess the kind of subject-matter competencies that many tertiary education departments or faculties would consider their primary objective. There would thus be a risk that what is measured becomes too far removed from what goes on in faculties and departments and does not capture the competencies that are uniquely the province of TEIs.

47. The discipline strand will seek to assess discipline-related competencies. The challenge of this approach is that it requires highly differentiated assessment instruments, which would make comparisons across TEIs and countries difficult. It also excludes competency areas that are not easily amenable to large-scale assessment or that are not sufficiently invariant across cultures. However, the experts who convened in 2007 considered that interesting examples do already exist in selected disciplines, such as science, technology, engineering or health care/professions allied to medicine (STEM), on which a feasibility study could be based. It would therefore be conceivable to initially focus the feasibility study in one or two of these areas, and then progressively expand the range of subject areas covered over time. The advantage would be that this approach would cover disciplines that are common among TEIs in OECD countries, less likely to be influenced by unique cultural features, and reflect the dynamics of disciplinary change. One caveat to this approach that the experts noted was, that many countries have already established assessments in the STEM disciplines and that the benefits of international collaboration would be greatest in areas that were yet unexplored and where national work was least likely, such as the humanities and the social sciences.

48. The experts suggested that, whatever the disciplines chosen, the aim will 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 novel contexts unfamiliar to them, an approach that is similar to PISA. The intent will be to develop forward-looking definitions of learning outcomes, which acknowledge change in the disciplines and establish instruments around expert thinking about labour-markets, the economy and social well-being over the next decade. This will be facilitated by using a representative set of disciplines that are at the cutting edge. For the purpose of the feasibility study, engineering and economics could be the focus to get insight into the feasibility of measuring discipline-specific skills in both scientific and social sciences domains, with the understanding that an AHELO would aim at expanding the number of disciplines covered over time.

49. As for the generic skills strand, ***the discipline strand will explore the feasibility of directly measuring learning outcomes in selected disciplines and across different cultural and linguistic contexts.*** Planning for this strand of work is now starting. International curriculum and assessment experts will shortly be convened to review existing assessment instruments that focus on discipline-specific skills, and to recommend options towards an international pilot in 3-4 countries. Possible instruments for consideration include the Brazilian Provão (*Exame Nacional de Cursos*) or ENADE (*Exame Nacional de Desempenho*), the Mexican EGEL (*Exàmen General para el Egreso de la Licenciatura*) or the United

States' *Measure of Academic Proficiency and Progress* (MAPP) or *Major Fields Tests* (GRE). Assessment instruments developed by professional associations will also be examined.

50. In terms of practical implementation, the engineering and economics strands of work will follow a similar model as for the generic skills strand described above, once the instrument to be piloted will have been selected.

Value-added measurement strand

51. As indicated above, two types of learning outcomes measures can be considered at the level of TEIs or departments:

- Individuals, whether prospective students or employers, would want to know the “bottom line” of the performance of TEIs, departments or faculties, in terms of the absolute performance or raw scores students enrolled in these would attain on an assessment.
- Individuals, TEIs and public policymakers wishing to assess the quality of the teaching services provided by TEIs would primarily be interested in the incremental learning deriving from tertiary education attendance, *i.e.* the “value-added” or “gain” provided by the TEIs measured by the scores an institution would attain after accounting for the quality of prior schooling or the degree of selectivity of the programmes and TEIs.

52. The OECD is committed to exploring both kinds of outcomes measures in an AHELO to meet the information needs of the various groups of stakeholders. However, it has to be recognised that assessing the value-added or marginal gain by TEIs raises a number of scientific and practical issues and the measurement of value-added would impose layers of complexity that, though theoretically well understood, are difficult to resolve in large scale assessments. Two sets of strategies can be considered in this area:

- One strategy is to focus the assessment on skill environments that students would typically not encounter prior to university entry, such as science, technology, engineering or health care/professions allied to medicine (STEM).
- The other strategy involves the *post-hoc* collection of data on prior learning as, for example, done in the Collegiate Learning Assessment (CLA), the synthetic linkage of outcome information of incoming students to outgoing graduates, or the longitudinal follow-up of cohorts of students. Each of these approaches has its merits and drawbacks (Nusche, 2007).

53. Given the complexity of measuring marginal gain, the proposed approach for the AHELO feasibility study is to explore both types of outcomes measures sequentially, and to examine the value-added perspective only conceptually in a first stage, while the pilot measurement of value-added in an international context would be incorporated into the design at a second stage. ***The value-added measurement strand will therefore scrutinise possible methods for capturing marginal learning outcomes that can be attributed to TEI's attendance, both at a theoretical level and in terms of psychometric approaches.*** It will build upon similar work carried out at school level by the OECD and review options for value-added measurement in tertiary education. Researchers will be invited to study potential data sources, methodologies and psychometric evidence on the basis of datasets existing at the national level, with a view to providing guidance towards the development of a value-added measurement approach for an AHELO.

Contextual strand

54. While the main focus of an AHELO would be to provide TEIs with a tool for development and improvement of teaching and learning, the OECD acknowledges that other types of uses are relevant for different purposes or groups of stakeholders. Therefore the AHELO approach seeks to develop multi-dimensional measures of tertiary education quality, with different emphases to serve the different uses and users. In recognition of this multidimensional approach, the AHELO feasibility study will also include a **contextual strand of work that will explore the development of instruments to capture contextual measures at institutional level as well as appropriate indirect proxies of learning outcomes.**

55. A possible model for this strand of work could be to devise contextual information indicators and indirect measures of outcomes of the kind developed by the German Centre for Higher Education (CHE), and apply them in an international context. The CHE has developed a number of indicators based on a mix of factual data as well as students and teachers' opinion surveys. These indicators cover various aspects of tertiary education quality, including:

- Academic studies and teaching (contact between students, counseling, courses offered, opportunities for e-learning, study organisation and teaching evaluation);
- Equipment (IT-infrastructure, library, computer workstations, spending for books and journals, rooms);
- International orientation (support for stays abroad);
- Job market and career orientation (employment market related programmes, practice support);
- Research (number of doctorates, publications and internationally visible publications, extent of third party funding);
- Study location and TEI (amount of sport, level of accommodation rent, size of TEI); and
- Overall opinions (study situation, reputation for academic studies and teaching, research reputation).

56. The emphasis for this strand of work will be to develop internationally comparable indicators of a similar kind. The aim will be to capture institutional characteristics, access patterns (such as the socio-economic makeup of the student body), students' perceptions of their tertiary education experience as well as a range of outcome measures such as the labour market outcomes of graduates.

57. Table 1 below summarises the main features of the various AHELO feasibility study strands of work.

References

Nusche, D. (2007), Assessment of Learning Outcomes in Higher Education: A Comparative Review of Selected Practices, OECD Education Working Paper No.15, accessed 3 June 2008 from www.oecd.org/dataoecd/14/8/40257354.pdf

Table 1. AHELO feasibility study strands of work

Strands of work	Focus		Methodology	Instrument used	Countries involved	Expected outcomes
Generic skills strand	Generic skills		Direct assessment of students' competencies	Collegiate Learning Assessment	To be determined	1) Proof of concept of the validity and reliability of measuring higher-order skills across TEIs, languages and cultures. 2) Insight on feasibility of implementation.
Discipline strand	Discipline-specific skills	Engineering	Direct assessment of students' competencies	To be determined	To be determined	1) Proof of concept of the validity and reliability of measuring discipline-specific skills across TEIs, languages and cultures. 2) Insight on feasibility of implementation.
		Economics		To be determined	To be determined	
Contextual strand	Access and contextual indicators at institution level Labour market outcomes of graduates Students' perceptions (or employers'), satisfaction rates with teaching		Contextual and indirect measures of quality at institutional level based on quantitative data, meta data and/or surveys of students and employers	CHE model	To be determined	1) Insight on international comparability of survey data 2) Insight on feasibility of implementation. 3) A multidimensional dataset of contextual and indirect measures of tertiary education outcomes/quality.
Value-added measurement strand	Conceptual and empirical measurement of value-added or marginal gain at TEI		Conceptual and psychometric research	Existing datasets of learning outcomes – mostly national datasets	Based on data availability	Research results and evidence on methods for capturing value-added in teaching.

ANNEX I: ESTIMATED BUDGET FOR THE AHELO FEASIBILITY STUDY, 2008-2010

The cost estimates presented in the table below constitute approximations – especially with respect to the discipline, value added and contextual strands of work for which the planning of activities is just starting. The estimates for national implementation costs are also based on a number of crude assumptions, *i.e.* extrapolations from PISA average national implementation costs per student adjusted for the duration of the various pilot assessments and the extent of marking involved. These estimates also assume an assessment of 1,000 students per pilot country (10 TEIs times 100 students) for each strand. More detailed explanations are available upon request.

		Estimates in EUR
Project administration and management (OECD)		1,800,000
Generic skills strand (international pilot test of CLA)		4,100,000
<i>of which</i>	<i>Development work</i>	<i>2,500,000</i>
	<i>Total national implementation costs</i>	<i>1,600,000</i>
	Country 1	400,000
	Country 2	400,000
	Country 3	400,000
	Country 4	400,000
Discipline strand		6,500,000
<i>of which</i>	<i>Development work - Engineering</i>	<i>2,000,000</i>
	<i>Total national implementation costs - Engineering</i>	<i>1,000,000</i>
	Country 5	250,000
	Country 6	250,000
	Country 7	250,000
	Country 8	250,000
	<i>Development work - Economics</i>	<i>2,500,000</i>
	<i>Total national implementation costs - Economics</i>	<i>1,000,000</i>
	Country 9	250,000
	Country 10	250,000
	Country 11	250,000
	Country 12	250,000
Conceptual value-added strand		350,000
<i>of which</i>	<i>Development work</i>	<i>350,000</i>
Contextual strand		1,900,000
<i>of which</i>	<i>Development work</i>	<i>1,500,000</i>
	<i>Total national implementation costs</i>	<i>400,000</i>
	Country 13	100,000
	Country 14	100,000
	Country 15	100,000
	Country 16	100,000
Total costs		14,650,000