

## **ASSESSING HIGHER EDUCATION LEARNING OUTCOMES SUMMARY OF A FIRST MEETING OF EXPERTS**

### **Introduction**

1. Policy makers as well as the public devote considerable attention to the outcomes of higher education. For some aspects of higher education outputs and outcomes, such as the type and number of degrees awarded, the research outputs produced or the labour-market returns to higher education studies, numerous indicators exist at both national and international levels. However, for other aspects, and most notably for the learning outcomes of higher education, available data are much more limited (a draft summary of existing assessments compiled by the OECD is available for reference). This topic attracted particular attention by OECD Education Ministers at their meeting in June 2006 and Ministers invited the OECD to explore how this information gap could be filled. With two expert meetings, the OECD seeks to explore the feasibility of further work in this area. The first expert meeting was held on 18 April 2007 at the OECD offices in Washington (see the Annex for a list of participants).

2. At this meeting, the experts identified considerable challenges for the development of internationally comparative measures of higher education learning outcomes and acknowledged that there was no clear roadmap for overcoming these – some compared the situation with when Columbus set sail. However, none of the experts considered the goal unreachable and all recognised that reliable information on learning outcomes would only rise 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.

3. The following summarises the discussions of the experts.

### **Determining relevant units of data collection and analysis**

4. Measures of higher education learning outcomes hold out important promises for students, higher education institutions (HEIs) 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 experts considered these units in the form of a three-dimensional matrix:

- The first dimension of the matrix comprised the stakeholders who would be served and their information needs. These included: *i) individuals*, such as students wishing to make better informed choices or employers seeking to benchmark qualifications against actual skill measures; *ii) higher education institutions, departments or faculties* seeking a better understanding of their comparative strengths and weaknesses and; *iii) public policy makers* seeking to quantify stocks and flows in high level skills, to obtain better insights into the quality, equity and efficiency of higher education services, and to assess the impact of policy decisions.
- The second dimension of the matrix related 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 considered the kind of instruments needed to serve the respective users and uses.

5. For several reasons, the experts suggested that it would, at least initially, not be feasible to develop internationally comparative information on higher 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 institutions. Second, in many countries governments have limited options to incentivise the participation of HEIs 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 these 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 higher education has become largely universal. Last but not least, the experts considered that mandated assessment, in the form that would be required to obtain a system-wide representative sample of institutions, would tend to lead to administrative response, not faculty response and thus was not effective as a tool for improvement at the level of service provision. Obviously, if an OECD assessment gathered pace and found wide acceptance, issues of assessing system-level performance could then be addressed, this would make the work relevant to a much wider range of stakeholders.

6. The subsequent discussions therefore centred on the establishment of measures of learning outcomes at the level of institutions, departments or faculties, the idea being to combine the definition of OECD measures of quality with reliable assessment methods to which institutions could, with an appropriate set of incentives, voluntarily subscribe and which could progressively find acceptance in a widening range of institutions. The experts considered two kinds of outcomes measures at the level of institutions or sub-institutions:

- Individuals, whether prospective students or employers, would want to know the “bottom line” of the performance of institutions, departments or faculties, in terms of the 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 aspects.
- Individuals, institutions and public policy makers wishing to assess the quality of the teaching services provided by HEIs would primarily be interested in the “value added” by the institutions, 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.

7. Assessing the value added by institutions would impose layers of complexity that, though theoretically well understood, are difficult to resolve in large scale assessments. The experts considered two sets of strategies as promising in this area:

- One strategy would involve the post-hoc collection of data on prior learning as, for example, done in the Collegiate Learning Assessment Project (CLA) or the synthetic linkage of outcome information of incoming students to outgoing graduates.
- The other strategy would 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).

8. The experts suggested that an OECD feasibility study should explore both of these complementary strategies.

9. The experts also considered what would constitute meaningful target populations for an international assessment of Higher Education learning outcomes. One possibility, that would facilitate comparisons across institutions 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 experts considered an assessment towards the end of a three or four-year degree as a more practical solution for a feasibility study.

10. 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 institutions 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 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. This issue is given further consideration in the subsequent section on the “quality space”. One point on which the experts agreed was that there was little point in reporting data at the level of individual students. Therefore, the assessment could consider the use of matrix sampling techniques through which the coverage of the assessment areas can be widened without unduly extending the response time demands on individuals.

11. The experts considered that the feasibility study should include institutions that do not have the power of selection, by means of reputation, but that are known to succeed in adding value in getting people to the degrees they aspire to.

12. There was some discussion whether the feasibility study should be limited to universities or cover tertiary education more widely. In this context, the experts noted that an international assessment should not get too far entangled in national institutional structures and that that contextualising the outcomes of universities in a broader setting of tertiary learning provision could be accomplished more effectively by PIAAC, with its coverage of the entire adult population.

### **Defining and operationalising higher education learning outcomes**

13. Any assessment of higher education learning outcomes will need to define and operationalise criteria for what constitutes quality. 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.

14. The experts also underlined that getting the “science for the assessment” right would critically determine the credibility of the exercise.

15. The experts reviewed existing experience with the establishment of measurable criteria for the quality of learning outcomes. The experts considered that it was most suitable to put the emphasis on both discipline-related competencies and transversal higher-order competencies manifested or required by particular disciplinary contexts. They therefore 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 the generation of knowledge and the interaction between substantive and methodological expertise, competencies that are widely viewed as critical for the

success of individuals and of rising relevance in the information age, to be assessed in the context of selected disciplines. The kind of competencies to be covered and the methods to be employed could be similar to those used for the Collegiate Learning Assessment Project (CLA). A key advantage is that such competencies are largely invariant across occupational and cultural contexts and could be applied across higher education institutions, departments and faculties. Since such assessments capture, at least to some extent, the cumulative learning outcomes of prior schooling, they should be combined with data on prior learning, for which the CLA provides methods as well. The main limitation of this approach is that it does not assess the kind of subject-matter competencies that many higher education departments or faculties would consider their primary objective. In other words, an approach entirely limited to generic competencies would risk that what is measured becomes too far removed from what goes on in faculties and departments and not capture the competencies that are uniquely the province of the institutions.

- A second strand would seek to assess discipline-related competencies. The challenge of this approach is that it would require highly differentiated assessment instruments, which would make comparisons across institutions and countries difficult. It would also exclude competency areas that are not easily amenable to large-scale assessment or that are not sufficiently invariant across cultures. However, the experts considered that interesting examples do already exist in selected disciplines, such as the STEM areas, on which a feasibility study could be based. It would therefore be conceivable to initially focus the feasibility study on one or two of these areas, and subsequently expand the range of subject areas covered progressively over time. The advantage would be that the approach would cover disciplines that are common among institutions in OECD countries, are likely to be less influenced by unique cultural features, and reflect the dynamics of disciplinary change. One caveat to this approach that the experts noted was, however, 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. 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 novel contexts unfamiliar to them, an approach that is similar to PISA.

16. For either of the two strands, the experts suggested that efforts of the OECD should not duplicate existing assessments and that links between international and national assessments of HE learning outcomes should be explored.

17. The experts acknowledged that the outcomes from these two strands would speak in different ways to the stakeholders identified above, with individuals and consumers perhaps most interested in transversal cross-cutting competencies, but faculty most likely interested in measures of their disciplines that relate most proximally to their work. However, the experts also considered that, even in a disciplinary context, assessing competencies in a broad way could link what is important to what is taught and thus become a powerful driver for improving the quality of teaching in the disciplines.

18. The experts suggested that an international assessment would provide an important opportunity to provide a forward-looking definition of learning outcomes, a definition that would 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 would be facilitated by using, under the second strand above, a representative set of disciplines that are at the cutting edge, such as bio technology. The experts suggested that adding a longitudinal perspective to the assessments would, over time, allow to assess the external validity of the outcome measures, in terms of how well they predict the success of individuals and how

they relate to subsequent individual labour-market outcomes, social outcomes and well-being, as well as assessing to what extent institutions make a difference to global education outcomes.

19. There was, however, recognition that the value of competencies may be perceived differently by different external stakeholders, with the human resource manager likely to focus on transversal competencies, knowing that the employee will assume various roles in his or her career, and the hiring manager primarily concerned with the disciplinary competencies required to matching the employee to a given task. These multiple perspectives would need to be reflected in the quality space.

20. The experts suggested that simplistic assessment tools and an impoverished definition of assessment would pose significant threats to the credibility of the exercise. There was also concern about the “tunnel vision” that could result from narrow assessments driving policy and practice. The experts suggested that one way to counter such tendencies would be to focus the feasibility study on a few institutions, rather than aiming initially at large scale assessment. This would allow thinking broadly about adequate technologies of assessment, having the feasibility study act as a repository of experience on good assessment tools, ensuring that considerations about unit costs do not dominate the exercise and avoiding that information gains are sacrificed for efficiency gains. It would also involve the institutions closely in the development of the assessment instruments. The experts viewed the development of measures and instruments that engage students and the faculties they serve as an essential condition for the success of an OECD exercise. The experts acknowledged that this would be easier for the discipline-specific measures identified under the second strand above, as the content-specific nature would involve faculties in defining the problems for the assessment.

21. The experts acknowledged that due consideration would need to be given to determining an appropriate balance with which an assessment would focus on assessing the top end of the performance distribution, on the one hand, and the capacity of institutions to provide baseline competencies, on the other.

22. The experts suggested that it would be useful to establish a more systematic overview of the lessons learned from existing higher education assessments and that any work in this area should be closely linked to PIAAC and PISA.

### **Establishing a “quality space” in which institutions and systems can be meaningfully situated**

23. The OECD made clear that it does not aim for the establishment of a single performance measure that will then be used for a uni-dimensional ranking of institutions or countries. The OECD also acknowledged that any effort to bring together all institutions 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 criteria for quality 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 in 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 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 institutions 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 and provide an inroad for a gradual process, by which first higher education institutions participate which are in the same space of international competition.

24. The key question would be which dimensions of diversity an international assessment by the OECD would capture. In this context, the experts recognised that learning outcomes are only one component of the quality of HEIs, and that the “quality space” would need to recognise other dimensions of the quality of outcomes as well, such that institutions and systems could be appropriately represented in the space, in accordance with their respective missions. The experts reported on experience with several aspects in this area, including curriculum ratings and consumer ratings. Alumni ratings, such as used in the CHE, were 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 were also seen as relevant. Last but not least, experts noted that there is a wide area of direct and indirect measures of research outcomes that could be utilised.

25. The experts noted that the further elaboration of the concept of the “quality space” would also need to consider various trade-offs, trying to find an appropriate balance between establishing quality dimensions that are as comparable as possible, while at the same time making these as country-specific as is necessary to allow for the reflection of historical, systemic and cultural differences between institutions and systems; portraying the dimensions of the quality space in ways that are as simple as possible while remaining as complex as necessary to reflect multi-faceted educational realities; and making the exercise manageable will require to keep the number of dimensions as small as possible while, on the other hand, making the “quality space” as comprehensive as necessary to situate institutions and systems in different cultural contexts.

26. A critical question would always be how outcomes from such assessments would be reported, i.e. what types of reporting would prove most effective the various stakeholders, in terms of raising outcomes and engaging personnel and institutions in improvement. This also involves the question to what extent the information individuals and institutions receive should go beyond the performance of their own institution. One view was that there should be an orientation towards making public all evidence from the evaluation of public policy (with appropriate analyses) in order to provide evidence to taxpayers and the users of higher education on whether institutions 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 behavior. The experts suggested that the OECD should actively pursue the concept of the “quality space” in disseminating any outcomes from such work. This could best be addressed with the OECD taking charge of the ranking issue, defining intelligent ways of communicating competency profiles, with a good model inciting others to use it, and thus leading to accepted good practices on the conduct on rankings. This could be facilitated by collaborating with relevant media to establish meaningful ways to report and disseminate institutional comparisons.

### **Next steps**

27. Given the complexity of the issues set out above, the experts suggested that the development of a strategy would need to start conservatively and proceed by way of developing exemplary material.

28. The chair proposed to explore the terrain with a feasibility study involving a limited set of volunteering institutions in a limited set of countries that provide a reasonable geographic and geopolitical balance.

29. The experts suggested that, as a next step, the OECD should compile prototype assessment material that the second expert meeting could then consider to validate the general principles established at this meeting and to establish the criteria by means of which the success of the feasibility exercise would be

established. The experts suggested making a deliberate effort to engage institutional academic leaders and faculty members in the early stages of the feasibility study.

30. This could be followed by technical and stakeholder meetings that would assess the framework and technical architecture of the feasibility study.

31. The OECD should also begin to clarify programme constraints, in terms of sample size, response time and assessment technology.

**ANNEX: PARTICIPANTS IN THE FIRST EXPERT MEETING**

Name	Title
<b>Jo Ritzen, Chair</b>	President of Maastricht University, Netherlands
<b>Anne Baly</b>	Branch Manager, Teaching, Equity and Collaboration Branch Department of Education, Science and Training Australia
<b>Roger Benjamin</b>	President US Council for Aid to Education responsible for the Collegiate Learning Assessment US
<b>Bong-Guen CHUNG</b>	Director-General, Ministry for Education Korea
<b>Masao Homma</b>	National Institute for Academic Degrees and University Evaluation, Japan
<b>Akihiko Kawaguchi</b>	Vice-President, National Institution for Academic Degrees and University Evaluation Japan
<b>Salvador Malo</b>	Mexican Competitivity Institute Mexico (IMCO ) Mexico
<b>Detlef Müller Böling</b>	Geschäftsfürher, CHE Gemeinnütziges Centrum für Hochschulentwicklung Germany
<b>Mari Pearlman</b>	Senior Vice President and General Manager Higher Education Division (HED) US
<b>Jamil Salmi</b>	Tertiary Education Coordinator, World Bank, US
<b>Marshall Smith</b>	Professor and former Undersecretary of State for Education The William and Flora Hewlett Foundation US

Name	Title
<b>Barbara Ischinger</b>	OECD, Director for Education
<b>Andreas Schleicher</b>	OECD, Head of Indicators and Analysis Division
<b>Stephan Vincent - Lancrin</b>	OECD, Analyst for Centre of Educational Research and Innovation