
The Ad Hoc Workshop, “Taking Stock of Education Performance: From Student Testing to System Evaluation” was held in Paris on 21-22 October. It was organised as part of the Education Policy Committee’s output for 2007-2008 ‘Accountability Systems and Evaluating Education Performance’. Earlier work carried out for the Committee found that crude proxies for student assessment systems are strongly related to student performance. The workshop was organised to provide an opportunity for national experts to describe in detail, discuss and compare different approaches currently used in countries to assessing education performance, and to suggest issues that might be addressed in the Committee’s 2009-2010 output on ‘Evaluation and Assessment for Improving School Outcomes’. This document reports on the outcomes of the workshop.

Gregory Wurzburg, Education and Training Policy Division
Tel: +33 (0) 1 45 24 92 95; E-mail: Gregory.Wurzburg@oecd.org
Background

1. The **Ad Hoc** Workshop, “Taking Stock of Education Performance: From Student Testing to System Evaluation” was held in Paris on 21-22 October. Thirty-six experts from 18 countries, BIAC and TUAC participated. The meeting was chaired by Dr. Bill Maxwell, Her Majesty’s Chief Inspector of Education and Training in Wales.

2. The Workshop was organised as part of the Education Policy Committee’s output for 2007-2008 ‘Accountability Systems and Evaluating Education Performance’. Analyses carried out for the OECD drew on PISA data from 2000 and 2003 to investigate the relationship between system-level arrangements regarding parental choice, school autonomy, and student testing regimes, and the quality and equity of schooling outcomes as measured by PISA (Wößmann et al. 2007a, 2007b; OECD 2007). The results suggested that the presence of student assessment systems appear to be more strongly related to student performance as measured by PISA than to system level features related to autonomy or parental choice. These results were broadly consistent with results for PISA 2006 data. However, partly because these analyses used crude proxies for assessment and evaluation, they offer only limited insight into how feedback might be used and eventually influence performance as measured by PISA.

3. Against this background, the workshop was organised to provide an opportunity for national experts to describe in detail, discuss and compare different approaches currently used in countries to assess education performance. Experts discussed how feedback from these different approaches is used and the trade-offs involved in opting for one approach over another. They also exchanged initial views on how to improve information on arrangements for assessment as well as the way in which assessment results are used. Experts also identified issues that might be addressed in the Committee’s 2009-2010 output on ‘Evaluation and Assessment for Improving School Outcomes’. A background paper was distributed to participants beforehand [EDU/EDPC(2008)33].

Main outcomes of the Workshop

4. The meeting was structured to address two issues: what drives the interest in feedback on education performance, and what are the main mechanisms by which countries obtain feedback (see Annotated Agenda [EDU/EDPC(2008)34]. The meeting concluded with an open discussion on possible directions of future work on the topic.

What drives the interest in feedback on education performance

5. This session consisted of interventions describing how international assessment results were treated within countries, the comparative effectiveness of rapid-assessment techniques as a strategy for student improvement, and the use of value-added models for evaluating education system outcomes (the latter being based on a forthcoming OECD report on the subject). Following initial interventions and comments there was general discussion.

6. Several factors appear to affect the usefulness of feedback.

   - Results from international surveys have important impacts when they contrast with expectations or can be used in ways that add a new dimension to widely shared views on system performance. In one country PISA results were influential in shifting the focus of analysis from the overall level of performance (that was already known to be generally good) to the distribution of outcomes; this provided a new and different way of looking at performance.
• Such survey results are also more useful when they are available over time. More generally it was argued that international surveys are most influential when they align with other international surveys as well as national evidence. It was noted that if results are comparatively good, complacency can set in and reforms to improve performance may be harder to enact.

• Evidence suggests that rapid-assessment—a computer-facilitated approach to frequent, brief formative assessment—is far more cost-effective as a strategy for raising education performance than top-down strategies such as comprehensive curriculum reform, raising teacher qualifications or reducing class sizes (see Yeh 2007, 2008). This appears to be explained at least in part by evidence that it increases student engagement and enhances the capacity of teachers to individualise instruction, and better manage a wide array of diagnostic and performance information.

• Formative assessment and other forms of assessment close to the delivery level appear to be particularly effective when feedback is early, frequent and objective. But rapid assessment and other types of formative assessment do not replace other forms of testing and examination because they are not well adapted to serving as tools for accountability.

• There were mixed views on the likely impact of publishing school performance results as league tables. Though there tends to be a high level of interest when such information is initially published, interest levels subside over time, possibly because they are seen to be such crude comparative measures: there is little evidence of much impact on school choice by parents. It was mentioned that in some circumstances published performance results do have an effect on mobility of teachers and the ability of schools to attract teachers; it also can spur school heads to engage in competitive behaviour, such as selecting in strong students and selecting out weak students, that may raise performance of individual schools, but the overall system.

• “Value-added” or “growth” models are intended to compensate for differences in performance by adjusting observed education outcomes for differences in starting points and contextual/background factors that affect learning. In principle, they indicate strengths and weaknesses, and help identify good practice (see OECD 2008 forthcoming).

• Though value-added models can facilitate school improvement directly, they appear to be driven more by an interest in ensuring accountability, for example, through better informing school choice decisions by parents.

• Some participants expressed concern about the complexity of such models, the need for multiple observations, the time required to generate the results, and the need for further refinement regarding the variables that are taken into account. Other options, such as comparisons of similar schools, are seen as offering an expedient compromise. Alternatively it was suggested that weak results on ‘value-added’ models could be used to trigger inspection to obtain more fine-grained pictures of performance and possible weaknesses.

• It was noted that there are important differences in the usefulness of feedback mechanisms for accountability versus school improvement. Mechanisms designed for school improvement focus on diagnosis and performance measurement at the service delivery level rather than the system level, provide more rapid feedback and engage different actors in dialogue about forward-looking questions about improvement.
Overview of mechanisms for gauging education performance

7. During the Workshop participants reported on and discussed national arrangements for four different approaches to obtaining feedback on education performance:

- **National Examinations** refer to examinations that students must take to complete a phase of education and/or to satisfy requirements for entering a new phase of education. Often referred to as ‘high stakes’ examinations, they are universal (taken by all students at a given stage of education), curriculum-based, and often administered and/or graded by authorities outside the school.

- **National Assessments** refer to tests administered to a sample of students to help gauge systemic performance – typically at a national level but also often at a sub-national level (state, municipality, school level, depending on sampling). The results do not have direct consequences for individual students, but may influence pedagogy, curriculum, and education strategy and management at the system level.

- **School inspection** refers to the practice by which central education authorities (though this may be at a sub-national level depending on where political responsibility for education lies) examine inputs to, and management of, education facilities, adequacy of qualifications of teaching and other staff, compliance to relevant laws and regulations, and education outcomes.

- **School self-evaluation, formative assessment and other school-based assessment** refers to various approaches to testing and evaluating student and school performance essentially for purposes of diagnosing progress and adjusting teaching, curriculum, and management practices.

8. These broad feedback categories correspond broadly to those used in the collection of international statistics on system level characteristics of education (through the collection of statistics for INES and the school-level questionnaire used for PISA). In the course of discussion it became evident that the categories are distinct, reflecting different purposes and needs of different users. The main purposes include: signalling levels of proficiency of individual students, thus controlling student flows in education systems; monitoring and compliance of institutions; and improving student, school and system performance.

9. Several more particular points emerged from the presentations and discussions.

- A variety of feedback mechanisms are needed to satisfy the respective needs of actors at different levels of the system:
  
  - Teachers and students need rapid, near real-time feedback to determine whether teaching and learning objectives are being met by individual students, and to shed light on why objectives are not being met when performance falls short.
  
  - Students and parents need to know at certain critical junctures whether students have mastered education curriculum sufficiently to exit one phase of education and enter the next.
  
  - School leaders and regional authorities need feedback on compliance and performance criteria that relate to decisions timed to budget and administrative calendars.
  
  - Policymakers need feedback on the consequences of programme design and resource allocation decisions that are evident only over extended periods of time. The timing and the nature of these
needs dictate different kinds of information requirements, different levels of aggregation, and different strategies for the collection and interpretation of feedback.

- Given these different needs and purposes, different approaches are complementary and vary in the extent to which they can substitute for one another. The feedback they provide varies in transparency, interpretation, suitability and usability to different potential users.

- Countries are quite different in the spectrum of feedback methods developed and the extent to which actors in different levels are equipped and inclined to use feedback.

- Exit exams lend focus and direction to students and teachers when linked to a flexible curriculum.

- There is progressively greater use of sample-based national assessments for measuring education and learning outcomes, using such feedback to measure overall system performance, and providing insights into reasons for observed strengths and weaknesses.

- The role of school inspection is evolving in at least some countries. It has shifted away from accountability based on monitoring inputs and compliance with rules and regulations towards a greater emphasis on school improvement. Inspection is also moving away from prescribing towards advising. This shift from what has been close to an adversarial relationship in some countries to cooperation and collaboration between inspectors and schools has been accompanied by schools themselves taking a more active role in self evaluation. Countries that have taken this approach report high levels of satisfaction.

- In a context in which schools assume certain responsibilities for improving performance, some countries are also using inspection more strategically, focusing on schools ‘at risk’. There is ongoing research to determine which observable risk factors are associated with subsequent poor performance. Risk factors cited include high concentrations of ‘at-risk’ children (e.g. from low socio-economic background, children of migrants, children in remote areas), high staff turnover rates, schools not inspected for a long time, schools where problems have been identified and schools where financial management has been problematic. However, there is a danger that this approach will have the effect of stigmatising schools chosen for inspection.

- In decentralised systems where schools have greater autonomy and also assume greater responsibility for self-evaluation, independent inspection serves a positive role in validating and legitimising what schools do.

- It was noted that where inspection occurs at uneven time intervals and on a case-by-case basis, inspection reports are less useful as sources of representative insights into system performance.

- The role of inspection in teacher evaluation varies from country to country; it appears to be increasing; it is not clear how changes in the other roles of inspection affect this particular role or how changes in this particular role affects the quality of teaching.

- Greater reliance on self-evaluation by schools (as well as the capacity to link school-level information to national test and assessment results) intensifies pressure for teachers and school leaders to possess the skills and know-how to collect and use information needed for diagnosis and performance measurement. Though external consultants and inspectorates may have some role to play in helping teachers and school leaders interpret the feedback and use it to improve
performance, these competencies will eventually need to be built into pre- and in-service training programmes.

- Where self-evaluation takes place in decentralised settings, it is necessary to take into account the role of municipalities or other relevant sub-national levels of government.

- However, more evaluation by itself does not improve performance (“weighing the pig doesn’t make it fatter”); it needs to feed into diagnoses of the problems underlying poor performance, and formulation and implementation of remedies; that, in turn, hinges on building coalitions for winning the needed political and financial support.

- Formative assessment appears to work particularly well when it is possible to relate results of individual students to national standards and nation-wide performance as measured through national curriculum-based exit exams and/or national assessments.

- The successful implementation of formative assessment strategies also has important consequences for pre- and in-service training of education professionals.

- It also was noted that the more general increased emphasis on testing and assessment intensifies the workload of teachers and may detract from learning by students; this situation is made worse if the purpose and methods over-emphasise accountability and under-emphasise improvement.

**Concluding Discussion on Future Work on Feedback**

10. The Secretariat explained that the outcomes of the Ad Hoc Workshop would be forwarded to Education Policy Committee for its consideration, and also would provide part of the foundation of the Committee’s future work on assessment and evaluation frameworks for improving education outcomes. That future work, drawing on the base of description of evaluation and assessment approaches provided by the Workshop, would explore the strengths and weaknesses of assessment and evaluation and the processes by which results are incorporated into policy and practice. In preparation for that new work, the Secretariat might need to collect more detailed information on evaluation and assessment arrangements than is presently available through the collection of system-level information carried out for INES and PISA.

11. In the ensuing discussion participants addressed three questions related to the contents and method of future work.

*Are there other important feedback mechanisms that should be considered?*

12. Participants expressed concern about over-reliance on standardised tests and examinations that fail to capture the full spectrum of outcomes of the school curriculum. They suggested taking into account at least three other kinds of feedback:

- **Public consultations.** These are carried out in some countries as a means for obtaining from various stakeholders their views on education systems and performance. Such consultations are useful for gauging perceptions about the systems and for building consensus, though they are subject to manipulation in some circumstances.

- **National education reports.** These are prepared to take stock of various aspects of education systems. They are typically based on a variety of forms of feedback including results from examinations, national surveys, and inspection. They may be prepared by standing bodies or *ad*
hoc commissions; they may be prepared on a regular or occasional basis; and they often lead to reforms.

- **Surveys of student views and post-schooling experience.** These provide feedback on aspects of and attitudes towards schooling that are not captured in examinations and assessments; as such they can help explain outcomes. Surveys that follow individuals after they leave school provide important information on how schooling performance relates to subsequent education and labour market experience.

13. Participants also indicated that in order to better understand the role of evaluation and assessment in improving education performance it would be helpful to know about the mechanisms that countries have in place such as agencies to review and broker research, for identifying useful feedback approaches and disseminating information on such approaches to various users. As well, they mentioned the value of having better information on how well-trained and equipped are teachers and school leaders to develop feedback mechanisms and effectively analyse and use the evidence they generate.

*What advice can be conveyed to those who will examine how to strengthen the role of feedback in school improvement? What questions should be addressed?*

14. Participants stressed repeatedly the importance of distinguishing between mechanisms that obtain feedback for ensuring accountability *versus* feedback for improving education. The latter requires different kinds of feedback (including differences in the timeliness and level at which feedback is made available) and different capabilities to fully utilise such feedback. It would be useful to understand the differences more thoroughly. Participants noted that though the purposes of accountability and improvement are complementary, they are not identical and in recent debates the former has overshadowed the latter. It would helpful to know how that imbalance might be redressed.

15. More detailed information is needed on interactive forms of assessment -- particularly those closely linked to formative assessment -- and how their role in improving education might be strengthened. Technology-linked forms of assessment such as computer adaptive testing and rapid assessment techniques should be included. It would be helpful to document these approaches and consolidate evidence on their cost effectiveness as well as the costs and feasibility of scaling up experiments that appear effective. It would be helpful as well to identify the pre- and in-service training required for teachers and school leaders in order for them to take a more active role in developing effective feedback mechanisms and fully utilising the information that they generate.

16. It would be helpful to examine how the inspection role has evolved in different countries, and to identify the strengths and weaknesses of the different approaches for improving education outcomes. Participants attached great importance to this line of inquiry because they see the complexity and potential benefits of improved feedback as intensifying the pressure on education professionals in schools, at precisely the same time when the trend towards decentralisation increases their autonomy. Inspection, because of its proximity to schools, offers a valuable means for channelling know-how to the teachers and school leaders who need it.

*What kinds of questions can be dealt with effectively at an international level and where should OECD concentrate its efforts with respect to collecting information, data and indicators?*

17. There was a widely shared view among participants that it would be useful for the OECD to map the different kinds of feedback to different users, indicating the methods by which feedback are collected, and the uses to which they are put (including distinguishing between accountability and improvement).
18. Participants were sceptical about the usefulness of indicators presently available in *Education at a Glance* as a tool for doing this. They argued that they mask and distort the complexity of present arrangements, and fail to capture information for a large number of countries. A deeper understanding of assessment and evaluation arrangements is needed before trying to distil internationally-comparable indicators from them.

**REFERENCES**


