

The use of external assessments and the impact on education systems

Olivier Rey



Olivier Rey has been a research analyst at INRP since 2003, where he leads the Veille Scientifique, a department which runs as an observatory on educational research, doing literature reviews and thematic short reports on current topics in education. He is in charge of mapping in NESSE, a network of scholars working on social aspects of education and training, launched in 2007 after a call for tenders by the European Commission). He is also involved in European Educational Research Quality Indicators (EERQI), a research project in the 7th Framework Programme for Research in the Socio-economic Sciences and Humanities Theme (SSH). (<http://www.inrp.fr/vst/LettreVST/SommaireLettre.htm>).

His main fields of expertise are educational policy, quality of educational research, evaluation and assessment, curriculum and higher education. Olivier Rey is also a teacher of higher education policy at Sciences Po, Paris and has worked as an external expert for the agency in charge of the evaluation of higher education and research (AERES, previously CNE). Before joining INRP, he was at the agency for mutualisation for French universities between 2000 and 2003, and was previously a Cabinet member of a French Minister in charge of education, Claude Allégre.

E-mail: olivier.rey@inrp.fr

The use of external assessments and the impact on education systems

Olivier Rey

Abstract

The development of assessing education systems through the performance of their students is a major phenomenon in recent years. At a national level, several European countries, that have not yet established such assessments, have developed standardised assessments to improve the measure of student performance and used these to infer the quality of schools and teachers. At an international level, The International Association for the Evaluation of Educational Achievement (IEA) and Programme for International Student Assessment (PISA) results have become an essential tool for comparing education systems. Their results have launched significant reform movements in some European countries.

In the frame of New Public Management, this development marks a shift from steering by inputs to steering by outputs. It also responds to a widespread concern to monitor and improve the quality of educational services in Europe, in line with the Lisbon 2010 objectives.

While these external assessments are increasingly at the core of education policies, educational research has tackled questions about their usefulness, limitations and risks. Indeed, output assessments offer a set of information to establish a diagnosis of educational systems, identifying their strengths and weaknesses. Many debates, however, have stressed that the information may be incomplete, reflect only a part of learning, suppress socio-cultural diversity and suffer from some methodological bias. Finally, if output assessments become education's targets, or if they are used to legitimise reforms adopted for other reasons, the data gradually loses its informational value.

Allowing researchers to construct and analyse assessment data, according to scientific rules, therefore, is a current challenge in order to ensure their effectiveness in improving educational practices and policies.

Introduction

Continuous (formative) assessment is the traditional form of assessment used in compulsory education. It remains the most common form of assessment method used by teachers and involves assessing in-class activities using tests, observations, homework and oral questioning.

Continuous assessment is sometimes combined with summative assessment to assess the extent and quality of learning at a given point, often for the purposes of assessing student performance before a transition to the next key stage. It may take the form of a diploma.

The object of this article is to study a third form of assessment that involves assessing student results through a centralised system using methods aimed at generating comparative evaluations of student performance at a national, regional (in the case of federal states) or international level.

The methods used in this form of assessment generally involve standardised tests and are based on external assessments without any intervention by teachers involved in teaching the assessed students.

Standardised external assessments may relate to specific curricula or study programmes (as is generally the case in national assessments or IEA tests), or to specific skills (as illustrated by the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) studies).

External assessments are used for a number of reasons, including to provide teachers and education professionals with feedback and comparative evaluations; diagnose the state of the education system as a whole; provide objective information for school users; and assess the effective acquisition of learning by students.

One of the chief purposes of the current use of external assessment is assessing and steering education systems.

The increasing use of external standardised assessments

The assessment of education systems on the basis of student results is a major feature of recent developments in education. A large majority of European countries now use external standardised assessments at a regional or national level.

Until the 1990s, only a small number of countries used national tests in compulsory education (primary and lower secondary education), either for the purposes of transition to the next year (Iceland, Portugal, Scotland, Northern Ireland, the Netherlands, Luxemburg, Denmark and Malta) or for providing diagnostic information about the education system as a whole (Ireland, France, Hungary, Sweden and the UK).

Ten other countries and regions followed suit in the 1990s including Spain, the French Community of Belgium, Latvia, Estonia and Romania.

Since the early 2000s, national tests have been introduced in the Flemish Community of Belgium, Lithuania, Poland, Norway, Slovakia, Austria, Germany, Bulgaria, Cyprus, Denmark and Italy.

According to the most recent Eurydice report (2009), only five countries have not introduced national student assessments: the Czech Republic, Greece, Wales, Liechtenstein and the German Community of Belgium.

The tests used before the 1980s were primarily designed to regulate student transition between levels or to assess key stages.

In the 1980s, the development of external assessments was primarily a result of decentralisation and the increasing autonomy of educational institutions. The assessment of results was the counterpart to a greater freedom for schools and education professionals.

Finally, since the 1990s, the process of decentralisation and increasing autonomy has been extended through an increased use of external assessment as part of a steering policy based on student results. National states seek to carry out more effective assessments of the quality and effectiveness of teaching using more rigorous and more incentive production indicators. In this way, assessment is used as a tool for monitoring education systems.

Different forms and methods of external assessment may co-exist according to specific social, institutional or political circumstances.

France is a good case in point. It is the ancestor of modern standardised assessment, and the education system has a long-standing tradition of using large-scale external assessment at the end of higher secondary education ('baccalaureate') for qualifying and summative assessment (transition from secondary to higher education); and for assessing overall student performance.

The French Ministry of Education introduced the first '*modern*' standardised assessments of student learning in the late 1970s at lower-secondary level as part of the introduction of the *collège unique* (comprehensive lower school) and in order to highlight the diversity of school performance levels.

Systematic diagnostic assessments have only been used since 1989, mainly at the end of primary education and at the beginning of secondary education. Their sole purpose is to generate information about the overall performance of the education system without any local details or indicators, partly because overall performance assessments are opposed by a number of teacher unions, which fear they may result in assessments of teacher performance.

Though annual and systematic large-scale diagnostic assessments are designed to provide teachers with an analysis of individual student difficulties, they are often used to determine the average level of students as a substitute for overall performance assessment.

Other types of overall performance assessment using samples (which tend to be more irregular) are designed for political and educational decision-makers.

Finally, since 2007, the French government has increasingly sought to promote the use of standardised assessment methods for steering education, with an increasing reference to PISA results.

The succession of assessment methods, which differ over time and have distinct aims and objectives, suggests a chaotic history marked by frequent political and professional interference (Mons and Pons, 2006).

While past standardised assessments were designed to assess learning and focus on students, their current use has expanded considerably. Standardised assessments connect the educational realm (their traditional remit) with the political realm, where they are used as a piloting tool (Behrens, 2006).

The influence of New Public Management on the development of standardised assessments

Education can hardly claim to be the only realm affected by the development of external and standardised assessments. Most scholars in the field agree that the introduction of these in education is the result of a transfer of the main principles of public policies since the late 1970s, referred to collectively as New Public Management (NPM) into education.

In the educational realm, four NPM principles have had a direct impact on the development of external standardised assessments at regional, national and international levels (Mons, 2009).

The production generated by public services (for example, education) is amenable to measurement.

Measurements must be carried out using specific instruments. The scientific rigour of these instruments is proven.

Public actors benefiting from increased autonomy, such as schools, are accountable for their actions to the managers of the system (management information model) and/or to citizens (democratic evaluation model).

Public organisation needs to be regulated by results (outputs) and not solely by procedural controls directed towards inputs.

Though initially apparent only in English-speaking countries, the new trend eventually spread to many other parts of the world.

In Europe, the implementation of NPM principles was particularly apparent in England, where reforms initiated since the 1980s had already encouraged using external assessments of student results at different educational levels and key stages. It quickly became apparent that the new standardised assessments were assessing individual school performance, which the new tests were designed to establish on the basis of the combined results of students.

Following the *1988 Education Reform Act*, the principle of accountability was applied to schools individually on the basis that '*schools matter*'. This occurred in the context of national tests generating crucial results for measuring individual school performance, which were quickly translated by the media into '*league tables*'.

In response to the lack of precision in the raw data, the notion of '*added value*' was introduced to measure success on the basis of the predicted success of a school population with identical average characteristics (age and social origin).

Measuring added value gained ground in the 1990s at an official level, entailing the need for richer and more detailed student data beyond that which a school might be capable of providing. The need arose for a system of data collection and centralisation at a national level, resulting, in 2002, in the Pupil Level Annual School Census (PLASC) (Schagen *et al.*, 2006).

At an international level, the external assessment of student results developed largely as a result of the PISA.

The impact of international assessments on education systems: the PISA case

International assessments have become a major focus of national education policies. National governments can use the results of international assessments to

carry out reforms aimed at resolving issues highlighted by the results of the assessments.

The ‘PISA shock’ in Germany

Germany is the most famous example of large-scale educational reform brought about as a result of PISA. Following the publication of the PISA results, the German education authorities engaged in a fundamental reconsideration of citizen perceptions of the German education system, and even of social and cultural values.

The PISA results showed that Germany was one of the countries in which educational success is most closely related to students’ socio-economic and immigrant background. In a country made up of regions that have a history of being fiercely protective of their local educational prerogatives, the ‘PISA shock’ resulted in an agreement by all Länder (federal states) to introduce national standards at the beginning of the 2004-05 school year. The agreement on national educational standards and on an institutionalised structure to ensure that the new standards were observed would have been inconceivable without the impact of the PISA study (Ertl, 2004). The reform was conceptually articulated in a report by Klieme *et al.* (2004) for the German Institute for Educational Research (DIPF). It recommended the introduction of “standards for results”.

International assessments have also had a significant impact on educational discourse and research in education. Here the emphasis is on research aimed at defining the most efficient educational practices and on empirical surveys. This goes against the grain of research and academic traditions that had previously focused on educational values and processes, leading some scholars to recommend a fundamental shake-up of the academic discipline of education in Germany (Bohl, 2004).

However, while the agenda of current reforms has tended to focus on specific issues (national standards, quality assessment and greater emphasis on skills rather than specific academic subjects), other problems have been overlooked. Problems that have yet to be discussed and debated include the highly sensitive issue of selecting higher-achieving students at the beginning of secondary education, and the absence of a comprehensive school system, identified by many observers as one of the main causes of German students’ poor results in the PISA study.

There are many similarities between the German and Czech students’ results in the PISA tests. While the differences of performance between the two countries and their education systems are not statistically significant, the perception of the PISA results has been markedly different.

In the Czech Republic, the PISA results were relatively well received and did not generate any major overhaul or large-scale reform. The policy of selecting higher-achieving students introduced in the early 1990s was largely welcomed and supported by the elite, media and parents. In Germany, the '*PISA shock*' was significant and generated important reforms, though none of the implemented reforms concerned the policy of selecting higher-achieving students at secondary level.

In Germany and the Czech Republic, the real results of the PISA survey have, therefore, been largely ignored. The educational policies introduced in the two countries have tended to consolidate the use of a neo-liberal form of performance assessment and regulation. The new measures only refer to scientific data for the purposes of legitimising policies introduced for other reasons. As noted by the Czech scholar, Greger, there is no evidence-based policy.

Therefore, it seems that both countries drew particular and different conclusions from a seemingly common diagnosis, not only when the PISA results were considered as too '*low*', but also in other cases.

Varying responses in different countries

In Hungary, the PISA results contradict the findings of previous IEA international studies. While previous IEA international studies (based on study programmes and curricula) showed an excellent level of student performance (confirming the elitist tradition of Hungarian education), PISA showed disappointing results in tests aimed at assessing the skills targeted by the OECD. All the conditions of a '*PISA shock*', comparable to Germany's, appeared to be present. Yet according to research, although they have entered public discourse on education in Hungary, the PISA results have not produced a political consensus over potential public-policy decisions or even an academic consensus over the best way to interpret them (Bajomi, *et al.*, 2009). Since 2006, national assessment tests have, nevertheless, been carried out at various educational levels and key stages, and the national assessment of skills has gradually become a knowledge-based regulation tool.

A completely opposite challenge faced Finnish representatives in education: how to manage PISA test scores? The excellent results obtained by Finnish students have, of course, only served to consolidate the Finnish education system, although it is important to note that a wide range of conclusions have been drawn from PISA by different education actors. Teacher unions have highlighted that good PISA results were the product of the expertise and quality of Finnish teachers. Government representatives have tended to view them as a consequence of the reforms introduced in recent years, including the development of a comprehensive school system, university teacher training,

greater teacher autonomy and the decentralisation of education. Yet, while the PISA studies indicate that the excellent results obtained by Finnish students are a product of the homogeneity of student outcomes (with just a small number of top performers), ministerial analyses have presented this as an issue by suggesting measures aimed at taking better account of gifted students (Rautalin and Alasuutari, 2009).

Although Portugal's 17th constitutional government did not have a '*PISA shock*' such as Germany's, PISA was used to justify a number of education reforms and to ensure a degree of '*international*' credibility beyond national controversies. Examples include the New Model of Teacher Assessment, the national programme for the educational promotion of Portuguese, and a national reading plan. PISA data were used to justify and strengthen arguments put forward by the government and to increase the government's decision-making capacities by consolidating the data generated by national assessment tools such as exams and assessment tests. The data have also fed into the government's discourse over the issue of repeating a year, which fails to improve student results (Afonso and Costa, 2009).

In the French Community of Belgium, PISA has also been much discussed in the media, parliament and public political debates. According to research (Mangez and Cattonar, 2009), the results of the PISA study are used as an argument to fuel debates that remain centred around pre-existing conflicts or traditional rivalries between different public policy options.

In other countries, PISA did not cause any injuries. In Denmark, for instance, IEA and PISA assessments have been the subject of many debates over the last 20 years and have fostered a culture of assessment (Egelund, 2008). Approximately ten national tests have been introduced on a range of themes and at different educational levels and key stages since 2007. According to experts, the emphasis on the acquisition of skills for life-long learning in the PISA studies has been well received because a high correlation between the objectives of compulsory education and PISA items was observed.

There has not been a '*PISA shock*' in France, in the sense of a converging and brutal shift of the public debate towards a widespread acceptance of reforms required following the publication of the PISA results. Until 2004, the debate surrounding PISA was primarily an insider affair, partly because the Ministry of Education anticipated the official publication of the results with a view to lessening their impact and channelling the public debate. Experts from the ministry underlined the methodological biases of the study, to the satisfaction of teacher unions and a wide range of education actors who claimed to have reservations about the very principle of an external assessment carried out by the OECD.

However, decision-makers have made significant use of PISA 2006, not least the French Minister of National Education, Xavier Darcos, who was appointed by the new right-wing president, Nicolas Sarkozy, elected to office in the spring of 2007. PISA 2006 marks a distinct break in France, with increased reference to PISA in political discourse and discussions of the survey in the French media. To this extent, a political change resulted in a cognitive change, and not the other way round (Mons and Pons, 2009).

Similarly in Romania, PISA generated very little discussion and had limited impact. Because the PISA results have not been taken up by key social or political actors, the results of the study are a topic of neither public nor professional concern (Fejes, 2009).

PISA, a tool between the political and academic realms

Research on the impact of PISA in different European countries has shown that as a result of OECD reports, experts, scholars and policy-makers have discussed the data of the different PISA waves (2000; 2003; 2006).

The success of PISA may be unprecedented in as much as the resulting data appear to provide a common reference that connects academic research in education and the political realm of public policy (Barroso and de Carvalho, 2008). PISA operates as a mediator between different strands of knowledge from the realm of education and public policy.

In this sense, PISA is an example of a range of public policy knowledge dependent upon a chain of translation processes between different spheres. The use of PISA by national governments is also an example of the way in which external assessment is now largely used as a tool regulating education policies as part of what public policy researchers refer to as '*soft regulation*', with the OECD acting as a '*third-party evaluator*' to assess, classify and compare suppliers of educational services (that is, nation states).

Mangez and Cattonar (2009) even suggest that external assessment surveys provide the OECD with the academic credibility required to promote political orientations on educational issues. Under the appearance of a neutral and objective diagnosis of education systems, PISA might, thus, serve as a means of promoting specific orientations on curriculum issues (for example, promoting skills rather than the transmission of academic knowledge).

How do researchers use the data generated by large-scale external assessments?

The use of external assessments in educational research

The impact of external assessments on the quality of education

In educational research, school effectiveness and improvement research was the first to show an interest in indicators of student output (in the early 1970s). Outputs are seen as a means of addressing issues of school quality, focusing on the importance of reflecting the specific effects of teacher activity and the school's education team. The focus of this research is premised on the notion that the specific activity of a school may 'make a difference'. This goes against the grain of a particular tradition in the sociology of education that argues that disparities and inequalities of educational success need to be understood in terms of specific social and political conditions (that is, the socio-cultural origins of students, system and state structures).

The issue is to identify the practices that may have a causal relationship with the best results obtained in the assessments, but also to use the dissemination of results as an information, motivation and mobilisation tool for education professionals.

The idea is that increased knowledge of the results of teacher activity on student learning (particularly among teachers) will encourage teachers to improve their practice. As summarised by Halasz (2006):

"There seems to be an increasing consensus among both politicians and educational researchers that the best way to improve the quality and effectiveness of education is to combine school autonomy with strong incentives to schools to use the feedback they get for self-improvement."

Online self- and external assessments can be combined to ensure that a reform in education is successful. Self-assessment helps school management and prepares students for external assessment, which gives an overall impression and can be used to confirm that it is in compliance with a reform's objectives. Feedback is the only way of enabling education actors to define their actions on the basis of the defined objectives (Zufanová, 2006).

The degree of flexibility governing the dissemination of external assessment results for the benefit of education professionals and decision-makers varies in different models.

Some argue that a relatively stringent level of accountability, in which results are made public and have significant implications in an educational market, is the best way to improve the quality and performance of educational services (Woessmann, 2007).

Others emphasise the mirror-effect (Thélot, 2002). The idea is to view assessment as an alternative means of market regulation. Assessment should not be required to generate explanatory schemes of educational success but should make education professionals aware of the results of their actions. Teachers and education managers need to understand the results of their actions. If results are not in keeping with teachers' objectives or with the results that they are expected to achieve, education professionals will modify their practice accordingly. For example, diagnostic assessments aimed at evaluating skills in geometry in the first year of French secondary education (*sixième*) over several years was enough to reactivate an interest in this particular area of mathematics (Bardi *et al.*, 2005).

According to Mons (2004; 2007), the use of centralised exams or tests is related neither to the average level of student performance nor to the proportion of pupils who attained the highest (5) or lowest (1) level in PISA assessments. In other words, the introduction of standardised assessments fails to improve the overall level of student performance, reduce educational failure, or identify an elite group of students.

The use of national exams is correlated with high average results, though only if the level of economic development is not taken into account. National tests, therefore, have no specific impact. Overall, the relations between effectiveness and assessment mechanisms are neither uniformly significant nor automatic.

However, standardised assessments appear to have an impact on educational equality by reducing the impact of social origin on educational performance. This involves limiting practices that lower standards in schools with a significant intake of children from working-class backgrounds (Mons and Pons, 2006).

A large number of studies were conducted following the introduction of high-stakes testing in the USA in the 1990s. Initially, the test results obtained by schools improved significantly after the introduction of the tests, before eventually levelling out. More importantly, a number of studies highlighted significant differences between the results obtained in national external standardised tests and the results obtained in local tests used to assess schools. It appears, therefore, that teachers adapted their teaching to meet the specific demands and challenges of the new tests, thus obtaining results strictly limited to the ared assessments – a process known as “*teaching to the test*” (Nichols *et al.*, 2006).

In addition to intensive exam preparation, high-stakes external tests have also been heavily criticised on several other counts: teachers focusing on pupils likely to obtain results that improve the overall school's results in the short term and the curriculum being restricted to the content in the tests, are just two criticisms.

Research has also indicated that English teachers now view the transmission of knowledge and skills as a priority at the expense of personalised support and pastoral care (Osborn, 2006).

It is, therefore, generally agreed that standardised external assessments should comply with a number of conditions to preserve their informative value and not alter the learning processes.

The chief condition is that standards should reflect teaching and learning objectives but should not be focused on the content of teaching. Klieme *et al.* (2004) suggested that results standards should not be used in the context of school careers: the data generated in the context of these procedures should not be used for the purposes of transition to the next level or for passing or failing a student.

In a similar vein, Mons (2009) argued that it is primarily high-stakes testing that generates negative responses from education professionals and leads to the perverse effect of a negative impact on learning.

What new information is likely to be generated by international assessments?

Since the early 2000s, there has been a renewed interest in studies that take account of the results of external assessments (both national and international) to carry out comparative analyses of different education systems and to produce better findings for the analysis and understanding of educational practices.

For example, Linnakyla *et al.* (2004) compared the small proportion of Finnish and Swedish students (7% and 12% respectively) who obtained below-par results in the PISA reading assessments with a view to identifying explanatory factors. The research identified the socio-cultural factors that schools struggle to act upon (for example, social origin and recent immigration) and factors that can be shaped by educational action, such as fostering greater self-esteem in failing students. It was concluded that strategies can be developed to support categories of students who demonstrate these 'risk' factors.

In France, another example is provided by the combined results of the PISA and Progress in International Reading Literacy Study (PIRLS) studies, which highlighted specific characteristics of French students, and particularly obstacles or behaviours that compromise success in assessments because of the way in which students are exposed to writing. French students tend to be less exposed to learning activities aimed at developing textual understanding than the average student taking part in PIRLS studies. It was also found that French students

engaged in relatively unvaried activities often based on the same material (Rémond, 2006).

International assessments (particularly PISA) have also provided comparative views of different education systems and policies on an unprecedented scale.

As noted by Mons (2008), who highlighted long-term variations in student results according to the effects of public policies, the databases generated by large-scale international assessments provide a basis for developing typologies of education systems. In terms of the assessment of education policies, the resulting typologies enable comparative statistical analyses of performance indicators of different education systems and institutional frameworks, which are the product of political choices and decisions.

Countries constituting geographical, historical or cultural units tend to have similar levels of educational inequality on the basis of an analysis of the results of PISA 2006, as noted by Green (2008).

English-speaking countries, particularly the USA and the UK, and continental Europe have high levels of educational inequality, while Northern and Latin European countries and East Asian countries have low levels of educational inequality. Using a qualitative comparative approach to highlight the various features of education systems that may account for regional differences, Green argued that the group of countries with the lowest level of educational inequality have a relatively uniform standardised school system and rarely use secondary level.

Methodological issues

The increasing popularity of external assessments, particularly at an internal level, has sometimes resulted in the use of raw data in public debate without any degree of methodological caution. The use of the same studies for both academic and steering purposes is particularly debatable. In addition to a number of methodological issues, this has given rise to a number of critical analyses in educational research.

Some researchers (for example, Tymms *et al.* (2004)) have complained that international assessments tend to focus on student results during or at the end of schooling irrespective of the progress made by students since the beginning of schooling. Longitudinal studies are, therefore, required to take account of the level of student learning at the beginning of schooling.

Though international comparisons may be relevant in terms of an overall performance assessment of different education systems, Wiliam (2008) argued that caution should be exercised when considering more detailed assessments. It

is impossible to draw specific conclusions concerning the quality of teaching. Assessments such as PISA are, like any standardised assessment, not particularly sensitive to the effects of high-quality teaching. There are three reasons for this.

The rate of progress of individual students is very limited in comparison with the rate of success within a particular age class.

Standardised testing methods tend to eliminate items that are particularly sensitive to variations in teaching.

The use of differentiation items in PISA to identify and exclude items that are not comparable across different languages reduces the sensitivity of assessments to teaching.

Swiss researchers (Moreau *et al.*, 2006) questioned the adequacy of the model and statistical variables used by PISA to highlight local specificities, particularly the differences between cantons that cannot be reduced to socioeconomic or educational disparities. They demonstrated the effects of PISA on the revival of a harmonisation project using standards inspired by PISA skills by showing the virtues and limitations of the transfer. One such limitation is that PISA skills are defined universally and therefore independently of specific educational curricula. By contrast, in a project such as HARMOS, the variety of educational curricula is taken into account in cantons with distinct and varied histories, cultures and languages.

Issues of translation and cultural context

International external assessments are invariably faced with the difficulty of using terms and concepts with different meanings in different languages. While a translation may be technically correct, the connotations and uses of the same term may vary from one country to another.

It is important not to underestimate the biases entailed by translations or different meanings of instructions in different languages and cultural areas (Rémond, 2006). For example, the French word '*style*' has a distinctly literary meaning (for example, the style of a writer), while the English word '*style*' also covers the style of an illustration or typography. French students face particular difficulties in cases where the expected answer requires a writing task that entails subtending meta-cognitive tasks. French students systematically seek to connect any question to routines and prefer not to give an answer if they have any doubt about the type of answer they are expected to give.

French scholars (Bautier *et al.*, 2006) carried out secondary analyses of the PISA 2000 literacy tests on over 800 French pupils and combined these with interviews aimed at providing a better understanding of the strategies adopted by students

in assessments. It was found that while PISA is presented simply as an assessment of reading and written comprehension skills, it also assesses more complex and more diverse skills involving, for instance, a mobilisation of varied spheres of reference extending well beyond learning objectives in reading and writing. The 'good' or 'bad' results obtained by a large number of students in certain PISA items depended on the contexts of interpretation and were less indicative of the level of student skills than the level attained in a range of meta-cognitive skills.

Conclusion: external assessments, between research and policy

Politicians and school administrators are probably the most frequent readers of PISA reports, which may entail the use of confusing phrases and such terms as 'factors', 'effects' and 'causes'. Investigations seeking causal factors will most often require methods other than surveys, such as qualitative in-depth studies and longitudinal surveys based on mixed quantitative and qualitative methods (Egelund, 2008).

Rochex (2008) warned against using PISA as a benchmarking tool and instrument of evidence-based education for conferring academic legitimacy.

However, while the collected data could justify the conduct of secondary or other more-detailed surveys (extending beyond just the standardisation of the results in each country and in order to overcome the noted methodological limitations), international organisations and governments seem reluctant to encourage and fund research that requires academic autonomy.

On a more optimistic note, Gustafsson (2008), while acknowledging that international assessments are liable to be misused and misinterpreted, also argued that they offer possibilities for improving the quality of educational research because the high-quality data generated by these studies can be useful in research on causal factors within or across education systems.

Considering the debates about national and external international evaluations, it is evident that everyone now recognises the usefulness of these evaluations for both researchers and policy-makers. Some methodological points need to be considered, but the key findings from the comparative achievement surveys are widely shared in the research community, at least. For example, the positive impact of a really comprehensive school on PISA scores in Finland, the negative impact of repeating years in France or the early selection in Germany can no longer be ignored. But does this mean that the 'right' curriculum can't be designed from using the results of the PISA items? Does it mean that the

competences at the age of 15 are more important to consider than the level of attainment at 20 years old?

These are the strengths and weaknesses of external evaluations on the level of educational policy. They bring a huge amount of indicators and information on educational systems, but these don't necessarily fit to on-going reforms or government goals. Moreover, governments sometimes only use the ranking and superficial features of the international achievement surveys, not the more detailed analyses. Despite the fashionable trends in evidence-based education, one cannot be convinced that policy has to be driven only by numbers, with the risk that policy-makers have too much say in the indicators used.

The challenge, both for researchers and policy-makers, therefore, is to consider that the aim to improve the information on educational systems has to be pursued for their common interests. Comparative external evaluations may continue to be a key means of doing this, but there needs to be an acknowledgement that policy and research processes cannot always run on the same timescales.

References

Afonso, N. and Costa, E. (2009). 'The influence of the Programme for International Student Assessment (PISA) on policy decision in Portugal: the education policies of the 17th Portuguese Constitutional Government', *Sísifo Educational Sciences Journal*, **10**, 53–64 [online]. Available: <http://sisifo.fpce.ul.pt/?r=25> [19 August, 2010]

Bardi, A.M., Bouzaher, M., Dyckmans, A., Houchot, A., Leblanc, M., Malléus, P., Menasseyre, C., Gauthier, R-S., Bassy, A-M. and Vogler, J. (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école*. Menes: Igen-Igaen [online]. Available: ftp://trf.education.gouv.fr/pub/edutel/syst/igen/acquis_des_eleves.pdf [19 August, 2010]

Barroso, J. and de Carvalho, L.M. (2008). 'Pisa: Un instrument de régulation pour relier des mondes', *Revue française de pédagogie*, **164**, 77–80.

Bautier, É., Crinon, J., Rayou, P. and Rochex, J.-Y. (2006). 'Performances en littéracie, mode de faire et univers mobilisés par les élèves; analyses secondaires de l'enquête pisa 2000', *Revue française de pédagogie*, **157**, 85–101.

Behrens, M. (2006). 'Préface.' In: Mons, N. and Pons, X. (Eds) *Les standards en éducation dans le monde francophone*. Neuchâtel: IRDP.

Bohl, T. (2004). 'Empirische Unterrichtsforschung und Allgemeine Didaktik. Entstehung, Situation und Konsequenzen eines prekären

Spannungsverhältnisses im Kontext der PISA-Studie', *Die Deutsche Schule*, **96**, 4, 414–425.

Egelund, N. (2008). 'The value of international comparative studies of achievement - a Danish perspective', *Assessment in Education: Principles, Policy & Practice*, **15**, 3, 245–251.

Eurydice (2009). *National testing of pupils in Europe: Objectives, Organisation and Use of Results* [online]. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_studies_en.php [19 August, 2010]

Green, A. (2008). 'Le modèle de l'école unique, l'égalité et la chouette de minerve', *Revue française de pédagogie*, **164**, 15–26.

Greger, D. (2008). 'Lorsque PISA importe peu. le cas de la république tchèque et de l'Allemagne', *Revue française de pédagogie*, **164**, 91–98.

Gustafsson, J.E. (2008). 'Effects of international comparative studies on educational quality on the quality of educational research', *European Educational Research Journal*, **7**, 1, 1–17.

Halasz, G. (2006). 'Foreword'. In: Dobbelstein, P. and Neidhardt, T. (Eds) *Schools for Quality: What Data-based Approaches Can Contribute*. Sint-Katelijne-Waver: Consortium of Institutions for Development and Research in Education in Europe (CIDREE). [online]. Available: <http://www.cidree.be/uploads/documentenbank/d3df7cd6c6a61f9d70302ba2a370bb08.pdf> [19 August, 2010]

Klieme, E., Avenarius, H., Blum, W., Dobrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H-E. and Vollmer, H.J. (2004). *The Development of National Educational Standards: an Expertise*. Berlin: Federal Ministry of Education and Research [online]. Available: http://www.bmbf.de/pub/the_development_of_national_educational_standards.pdf [8 September, 2010].

Linnakyla, P., Malin, A. and Taube, K. (2004). 'Factors behind low reading literacy achievement', *Scandinavian Journal of Educational Research*, **48**, 3, 231–249.

Mangez, E. and Cattonar, B. (2009). 'The status of PISA in the relationship between civil society and the educational sector in French-speaking Belgium', *Sísifo Educational Sciences Journal*, **10**, 15–26 [online]. Available: <http://sisifo.fpce.ul.pt/?r=25t> [19 August, 2010]

Mons, N. (2004). *De l'école unifiée aux écoles plurielles: Évaluation internationale des politiques de différenciation et de diversification de l'offre éducative*. Thèse de doctorat, Université de Bourgogne, Dijon.

- Mons, N. (2007). *Les nouvelles politiques éducatives: La France fait-elle les bons choix ?* Paris: Presses universitaires de France.
- Mons, N. (2008). 'Evaluation des politiques éducatives et comparaisons internationales: introduction', *Revue française de pédagogie*, **164**, 5–13.
- Mons, N. (2009). *Theoretical and Real Effects of Standardised Assessment*. Brussels: Eurydice. [online]. Available: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/11EN.pdf [19 August, 2010].
- Mons, N. and Pons, X. (2006). *Les standards en éducation dans le monde francophone: Une analyse comparative*. Neuchâtel: Institut de Recherche et de Documentation Pédagogique.
- Mons, N. and Pons, X. (2009). 'The reception of PISA in France: a cognitive approach of institutional debate (2001-2008)', *Sísifo Educational Sciences Journal*, **10**, 27–40 [online]. Available: <http://sisifo.fpce.ul.pt/?r=25t> [19 August, 2010]
- Moreau, J., Nidegger, C. and Soussi, A. (2006). 'Définition des compétences, choix méthodologiques et retombées sur la politique scolaire en suisse', *Revue française de pédagogie*, **157**, 43–53.
- Nichols, S.L., Glass, G.V. and Berliner, D.C. (2006). 'High-stakes testing and student achievement: does accountability pressure increase student learning?' *Education Policy Analysis Archives*, **14**, 172 [online]. Available: <http://hub.mspnet.org/index.cfm/12605> [19 August, 2010].
- Osborn, M. (2006). 'Changing the context of teachers' work and professional development: A European perspective', *International Journal of Educational Research*, **45**, 4–5, 242–253.
- Rautalin, M. and Alasuutari, P. (2009). 'The uses of the national PISA results by Finnish officials in central government', *Journal of Education Policy*, **24**, 5, 539–556.
- Rémond, M. (2006). 'Éclairage des évolutions internationales pirls et pisa sur les élèves français', *Revue française de pédagogie*, **157**, 71–84.
- Rochex, J.-Y. (2008). 'Pisa et les enquêtes internationales. enjeux scientifiques, enjeux politiques', *Revue française de pédagogie*, **164**, 81–85.
- Schagen, I., Hutchison, D. and Hammond, P. (2006). 'League tables and health checks: the use of statistical data for school accountability and self-evaluation.' In: Dobbelstein, P. and Neidhardt, T. (Eds) *Schools for Quality - What Data-based Approaches Can Contribute*. Sint-Katelijne -Waver: Consortium of Institutions for Development and Research in Education in Europe [online]. Available: <http://www.cidree.be/uploads/documentenbank/d3df7cd6c6a61f9d70302ba2a370bb08.pdf> [19 August, 2010]

Thélot, C. (2002). 'Évaluer l'école', *Etudes*, **397**, 10, 323–334.

Tymms, P., Merrell, C. and Jones, P. (2004). 'Using baseline assessment data to make international comparisons', *British Educational Research Journal*, **30**, 5, 673–689.

Wiliam, D. (2008). 'International comparisons and sensitivity to instruction', *Assessment in Education: Principles, Policy & Practice*, **15**, 3, 253–257.

Zufanová, H. (2006). 'On the meaning of internal and external evaluation for the quality of education during educational reforms'. In: Dobbstein, P. and Neidhardt, T. (Eds) *Schools for Quality - What Data-based Approaches Can Contribute*. Sint-Katelijne-Waver: Consortium of Institutions for Development and Research in Education in Europe. [online]. Available: <http://www.cidree.be/uploads/documentenbank/d3df7cd6c6a61f9d70302ba2a370bb08.pdf> [19 August, 2010].

Further Reading

Berényi, E. and Neumann, E. (2009). 'Grappling with PISA. Reception and translation in the Hungarian policy discourse', *Sísifo Educational Sciences Journal*, **10**, 41–52 [online]. Available : <http://sisifo.fpce.ul.pt/?r=25> [19 August, 2010].

Olsen, R.V. and Lie, S. (2006). 'Les évaluations internationales et la recherche en éducation: principaux objectifs et perspectives', *Revue française de pédagogie*, **157**, 11–26.

Prenzel, M. and Zimmer, K. (2006). 'Études complémentaires de pisa 2003 en Allemagne: Principaux résultats et enseignements', *Revue française de pédagogie*, **157**, 55–70.

Rochex, J.-Y. (2006). 'Pisa: analyses secondaires, questions et débats théoriques et méthodologiques', *Revue française de pédagogie*, **157**, 5–9.

~~Rostas, Z., Kosa, I., Bodó, J., Kiri, and Fejes, I. (2009). *Use and Circulation of PISA in a Romanian Context*. Knowledge and Policy Network [online]. Available: <http://www.knowandpol.eu/index.php?id=245#c403> [19 August, 2010].~~

Schütz, G., West, M.R. and Wössmann, L. (2007). *School Accountability, Autonomy, Choice, and the Equity of Student Achievement: International Evidence from PISA 2003* (OECD Education Working Papers) [online]. Available: <http://dx.doi.org/10.1787/246374511832> [19 August, 2010].

Note

- 1 IEA is an independent, international cooperative of national research institutions and government research agencies. The most well-known reports published by the IEA include TIMMS (Trends in Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study).