

# System-level assessment and educational policy

John Cresswell

## Introduction

Education ministries throughout the world are integrating educational assessment into their education reform processes. Figure 1 below shows that education reform is not a straight-line activity, and that assessment can both inform the discussion on policy reform and give an indication of the effectiveness of policies that have been implemented.

Three different types of assessment are used in the educational process: classroom assessment, examinations and system-level assessment. Each type has a different purpose. While the focus of both classroom assessments and examinations is to measure the learning outcomes of individuals, for system-level assessment the focus is to describe the characteristics of the population as a whole so that policies can be designed and put in place to target observed areas of weakness. This paper attempts to demonstrate how system-level assessments can inform education policy by drawing on examples of countries that have used assessment programs in education reform processes.

System-level assessment programs have been used by education authorities in a variety of ways, and these programs can exert significant influence. In a study by the Organisation for Economic Co-operation and Development (OECD), countries were asked how influential results and analyses from the Programme for International Student Assessment (PISA) had been in informing the policymaking process at their national or federal

level. Of the 37 respondent countries/economies, 17 rated PISA as 'very' influential, and a further 11 rated it as 'moderately' influential (Breakspear, 2012).

## System-level assessment as a trigger for change

As shown in Figure 1, system-level assessment feeds into the reform discussion. An education system may not have undergone changes for many years, and a system-level assessment program may reveal that students are not attaining the expected outcomes.

The results from the first implementation of the PISA survey triggered responses in a number of countries. These responses came to be known as 'PISA shock' when they were particularly dramatic.

In one example, Denmark, whose results demonstrated a lower-than-expected level of student ability, undertook a review of the national education system. The review covered different themes, including (OECD, 2004):

- learning standards, evaluation of student performance and school effectiveness
- the roles and competencies of school leaders
- pre- and in-service professional development of teachers
- the collective agreement regulating the roles and hours of teachers
- opportunities for bilingual and special-needs students.



**Figure 1** The education reform cycle

Denmark subsequently implemented a range of reform policies, including increasing national assessment and evaluation, and implementing strategies to target socioeconomically disadvantaged and immigrant students (Egelund, 2008).

In another example, after a surprisingly low performance in PISA 2000, Germany responded with a complete overhaul of its education system. Prior to 2000, children were selected to go into different types of school: either a Gymnasium (for academic students), a Realschule (for intermediate students), or a Hauptschule (for the less academic students). This separation was thought to be perpetuating inequity, firstly because there was a relationship between the socioeconomic backgrounds of the children and the schools to which they were selected to attend, and secondly because it was difficult for children to move from one type of school to another (Young, 2015).

To address the perceived inequities, measures were taken to adjust the system. These measures included deferring the age when children are assigned to different secondary schools, merging Realschulen and Hauptschulen, and introducing more comprehensive schools, such as lower-secondary schools for which enrolment practices were not based on academic performance (Young, 2015).

A significant proportion of the low performers were migrants who, because of their poor German-language skills, were usually assigned to Hauptschulen. It was thought that the introduction of subsidised all-day schools and comprehensive schools that didn't segregate by ability would provide more language support and scope for integration (Young, 2015).

Being a federal country, each German state has its own education system, making it difficult to implement national programs. Despite this difficulty, there has been a move to standardise curricula and introduce national tests in response to PISA results. In one article, a teacher interviewed stated her belief that due to PISA, lessons had become more interactive with less emphasis on rote learning (Young, 2015). In PISA 2012, Germany's results in PISA increased.

In Japan, also, there was a reaction to PISA that informed changes to educational policy. Following a decline in performance between 2000 and 2003,

there was significant public and political debate on education. This discussion led to the Ministry of Education reversing a controversial curriculum policy and changing the national assessment (Takayama, 2008).

In Norway, PISA results led to reforms of both assessment and curriculum. A national quality assessment system that included national tests and a web-based portal for presentation of data for school evaluation was introduced in 2004. This was followed by the introduction of the National Curriculum for Knowledge Promotion in 2006. Recent policy initiatives have included the Better Assessment Practices Project and a four-year project called Assessment for Learning (Baird et al., 2011).

First-time participation in an international system-level assessment often provokes reform (Gilmore, 2005). One example is Romania's participation in the Trends in International Mathematics and Science Study (TIMSS) in 2003:

*TIMSS findings gave rise to a considerable amount of curriculum reform. Several new topics were added to the mathematics curriculum as a result of TIMSS, including 'statistics and probability', 'data analysis and representation'. An increased emphasis was given to problem solving and a reduced one to deriving formal proofs. An integrated science curriculum was approved for grades 3 and 4. In sciences most of these changes were related to more emphasis on practical investigations, relocation of topics, and more stress on scientific inquiry. New Teachers' Guides for Science (grade 3–4) and Chemistry and new textbooks for students, in both Math and Science, were written based on TIMSS experience — one textbook for Science (grade 4), three textbooks for Physics (grade 6, 7, 8), three for Biology (grade 5, 7, 8), one for Chemistry (grade 9) and a few for Math (grade 1–9). All of them are used in schools following the changes, which occurred after the release of TIMSS results.*

(Gilmore, 2005, p. 34)

Another example is from Turkey, where, after the results of the Progress in International Reading Literacy Study (PIRLS) of 2001 were released, the

Board of Education developed a new curriculum for reading literacy (Gilmore, 2005).

Countries with an extensive record of participation in international system-level assessment have become more adept at tailoring their participation to give themselves the best opportunity to integrate results into the educational policy cycle. In Switzerland, a country with distinct areas based on language, oversampling permits the comparison of performance of students in the different cantons (member states). Participation in PISA, TIMSS and PIRLS has served to accelerate a number of initiatives that had been considered for some time. These initiatives aimed to (Baird et al., 2011):

- harmonise educational structures, curricula and standards across the country, or at least within linguistic regions
- provide information about the education system in its entirety
- monitor the education system regionally and nationally.

Action towards harmonisation began in 2006 with the Harnos concordat, an inter-cantonal agreement on the harmonisation of compulsory education. It established age 4 as the national starting age for obligatory education, and 11 years as the duration of that education (Baird et al., 2011). Harnos also proposed the creation of national educational standards and common curricula within each linguistic region (Delamadeleine, 2008). A common curriculum is being introduced into the French-speaking region, with the German-speaking cantons to follow suit with their own version. There were also debates about the need to set up a program of sample-based system monitoring following the PISA model (Behrens, 2008; Ramseier, 2008).

## System-level assessment for monitoring reforms

Thus far, all examples have discussed reforms that have been introduced in response to the results of international system-level assessments or in accordance with the practices of international system-level assessments. Yet these assessments can also be used for monitoring reforms that have already been implemented. For example, in 1999, the Polish education authorities began a reform to

restructure the education system, especially with regard to the number of years students spend at secondary school. Poland's participation in PISA 2000 and subsequent PISA surveys allowed the authorities to monitor the effects of this policy change over time (Jakubowski, Patrinos, Porta, & Wiśniewski, 2010).

Before the reform, students were in comprehensive primary schools until age 14, and from age 15 they attended academic, general or vocational secondary school. The rationale for this tracking was that students would perform best if they were in classes and schools with similar students. The 1999 reform introduced comprehensive lower-secondary schools for all children aged 13 to 15, thereby delaying the splitting into different types of school for one year. The reform was introduced gradually, so that when Poland participated in PISA 2000, most of the 15-year-old PISA target-age students were still attending either the academic, general or vocational schools; yet when it participated in PISA 2003, 15-year-old students were attending the comprehensive lower-secondary schools.

The PISA results provided an excellent way to monitor the effects of the policy reform. Between PISA 2000 and PISA 2003, the amount of variation in student performance explained by the school attended dropped from 61 per cent (one of the highest of all participating countries) to 20 per cent (below the OECD country average). This is to be expected, because by 2003 the reform had been fully implemented, so students were all attending the comprehensive lower-secondary schools. Less expected was the fact that in PISA 2000, the mean score in reading of Polish students overall was significantly below the OECD mean, whereas by PISA 2009, their mean reading score was significantly above the OECD mean. This suggests that students who were previously divided into the three different school streams were now all performing at a higher level when they were in the same type of school.

Poland's 1999 reform is but one example of how PISA results have informed the debate about tracking policies. PISA results also prompted the delay of tracking in the French-speaking part of Belgium, and influenced policy discussion in a number of other countries, most notably a group of European countries with a history of selection and

tracking, including Austria, Denmark, Germany (as discussed earlier), Hungary, Luxembourg and the Slovak Republic (Breakspear, 2012).

## Responses to PISA results

Some examples of other varied country responses to PISA results include (Breakspear, 2012):

- In Portugal, PISA may not have led to a dramatic restructuring of the education system, but it was influential in guiding the education and curriculum policies in the period up to 2007.
- In Ireland, Germany, Greece and Norway, curriculum standards were revised, often to include and emphasise PISA-like competencies.
- In Japan and Ireland, strategies aimed at specifically improving reading/literacy performance were introduced.
- In Austria, the Flemish-speaking part of Belgium and Hungary, equity has been promoted by addressing school financing.
- In Korea and the European Union, strategies have been introduced to monitor the proportion of students performing poorly and increase the proportion of top performers.
- In Austria, Japan and Korea, student engagement and attitudes have been discussed and strategies have been introduced to improve them.
- In Hungary, competency-based teaching and learning has been implemented.
- In the Slovak Republic, new national measurements of reading and mathematics were developed.

## PISA as a best-practice model

Some high-performing countries/economies used the PISA model of assessment as a focal point for dialogue, validation and innovation in national assessments. For example, since the late 1990s, Singapore has been using PISA items as inspiration as it considers how to assess the learning outcomes of its students. More recently, the delivery of PISA instruments by computer has informed Singapore's own efforts to deliver computer-based versions of its national examinations (Breakspear, 2012).

Other examples of how countries have used PISA as a best-practice model or guide for the design of new national assessments and adaptations to existing assessments include (Breakspear, 2012):

- In Hungary, PISA has had a great impact on the design and the framework of the Hungarian Assessment of Basic Competencies, and data analysis and reporting approaches are also similar to those established by PISA.
- In Korea, test developers and subject experts have tried to benchmark student performance, and to some extent have been inspired by PISA item types, assessment frameworks and test content.

## System-level assessment: Regional perspectives

Interestingly, with respect to developing countries in particular, one detailed review found that the impact of system-level assessment on educational policy was largely confined to certain geographic areas (Best et al., 2013). The review drew most of its material from documentation of national assessments in countries in South America and of regional assessments from sub-Saharan Africa. The review showed that less is known about the ways in which assessments have been used for policymaking in developing countries in Asia and especially in the Pacific. Overall, the review found that:

*Considering the link between assessment programme data and the policy process, and regardless of whether an assessment was sub-national, national, regional or international in type, assessment data were used slightly more often in three stages of the policy cycle, namely (i) policy agenda setting, (ii) policy implementation and (iii) policy monitoring and evaluation than for the stage of policy formulation. In other words, large-scale assessments had a slightly lower impact on the ways in which analytical and political options and strategies for education policies were constructed than on other types of policy activities.*

(Best et al., 2013, p. 63)

The review emphasised the differences between those assessments that had a census of the

target population and those that used a sampling approach, finding that census assessments tended to equally mention the policy goals of quality, equity and accountability, while assessments that used a sampling approach were connected more to quality as a policy goal than to equity or accountability. In addition, the review found that assessments that used a sampling approach had relatively more impact than census-based assessments on the policy formulation stage. Many of the policies that were formulated related to resource allocation aimed at improving teacher quality and increasing teaching materials, improving teacher preparation, and textbook reform and, to a lesser extent, at changing patterns of education funding (Best et al., 2013).

After this review, a further investigation into the use of system-level assessment data in the Asia-Pacific region found that assessments that have an influence on education policy tend to be national rather than international, secondary rather than primary and sample-based rather than census-based (Tobin, Lietz, Nugroho, Vivekanandan, & Nyamkhuu, 2015). This investigation also found that assessments are most frequently used to inform system-level policies and are less likely to be used to inform teaching and learning.

## Conclusion

The aforementioned OECD study presents six main ways in which PISA has influenced educational policy in participating countries (Breakspear, 2012):

- It has provided an assessment that acts as a system evaluation for countries/economies that did not previously carry out national/federal assessments.
- It has led to the formation, or increased the scope, of national/federal assessment systems.
- It has yielded data that can complement national data and validate national results against international benchmarks.
- It has enabled within-country/economy monitoring of sub-national regions or student groups.

- It has served as a best-practice model or guide for the formation and adaptation of national/federal assessment policies and practices.
- It has yielded data that have been used to evaluate the effectiveness of educational reforms.

Any high-quality international or regional system-level assessment might influence educational policy in these ways. Countries are increasingly interested in measuring and monitoring educational quality and equity, and in this climate system-level assessments are likely to continue playing an important role in the education reform processes.

## References

- Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. (2011). *Policy effects of PISA*. Oxford: Oxford University Centre for Educational Assessment.
- Behrens, M. (2008, April). Quelles évaluations dans le cadre d'HarmoS et de la Convention scolaire romande? *L'harmonisation de la scolarité obligatoire* (Politiques de l'éducation et innovations: Bulletin of the Intercantonal Conference of Education of French-speaking Switzerland and Ticino (CIIP)), 22, 13–14.
- Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D., & Tobin, M. (2013). *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Final report*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance* (OECD Education Working Paper No. 71). Paris: OECD.
- Delamadeleine, Y. (2008). Le PER, Plan d'études romand. *L'harmonisation de la scolarité obligatoire* (Politiques de l'éducation et innovations: Bulletin of the Intercantonal Conference of Education of French-speaking Switzerland and Ticino (CIIP)), 22, 11–12.

- Egelund, N. (2008). The value of international comparative studies of achievement: A Danish perspective. *Assessment in Education: Principles, Policy & Practice*, 15(3), 245–251.
- Gilmore, A. (2005). *The Impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries: An evaluation of the value of World Bank support for international surveys in reading literacy (PIRLS) and mathematics and science (TIMSS)*. Amsterdam: IEA.
- Jakubowski, M., Patrinos, H. A., Porta, E. E., & Wiśniewski, J. (2010). *The impact of the 1999 education reform in Poland* (Policy Research Working Paper No. 5263). Washington: World Bank.
- OECD. (2004). *Reviews of National Policies for Education: Denmark 2004. Lessons from PISA 2000*. Paris: OECD.
- Ramseier, E. (2008). Validation of competence models for developing education standards: Methodological choices and their consequences. *Mesure et Evaluation en Education*, 32, 35–54.
- Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, 44(4), 387–407.
- Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., & Nyamkhuu, T. (2015). *Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region*. Melbourne and Bangkok: ACER and UNESCO.
- Young, H. (2015, November 25). What can we learn from the great German school turnaround? *The Guardian*. Retrieved from <http://www.theguardian.com/teacher-network/2015/nov/25/what-can-we-learn-from-the-great-german-school-turnaround?>

The logo for the ACER Global Education Monitoring Centre (GEM) is a purple circle with a white geometric pattern of overlapping lines forming a complex, interconnected network of shapes, resembling a globe or a web of connections.

The ACER Global Education Monitoring Centre supports the monitoring of educational outcomes worldwide, holding the view that the systematic and strategic collection of data on educational outcomes, and factors related to those outcomes, can inform policy aimed at improving educational progress for all learners.

[www.acer.edu.au/gem](http://www.acer.edu.au/gem)