POLICY EFFECTS OF PISA

Oxford University Centre for Educational Assessment

Jo-Anne Baird, Talia Isaacs, Sandra Johnson, Gordon Stobart, Guoxing Yu, Terra Sprague & Richard Daugherty

May 2011

Table of Contents

Policy Effects of PISA

Overview

International surveys of student achievement are becoming increasingly popular with governments around the world, as they try to measure the performance of their country's education system. The main reason for this trend is the shared opinion that countries will need to be able to compete in the 'knowledge economy' to assure the economic wellbeing of their citizens. Whilst benchmark indicators of knowledge economy 'supply' variables, such as investment in education as a proportion of GDP, have been available for a long time, countries had no way of comparing the effect of their investments and schooling in general upon students' knowledge and skills.

Since 2000, OECD addressed this by producing a three-year cycle of 'curriculum-independent' tests of reading, mathematics and science. Each cycle has major and minor emphases, with one of the three skill areas consuming the majority of the assessment every nine years. Questionnaire enquiries are also conducted among students and teachers to measure contextual factors. PISA tests are taken at age 15 by representative samples of pupils in participating countries/economies, of which there were 75 in 2009. Age 15 was selected because it was the end of compulsory schooling in many OECD member countries when the programme was in the planning stage, but that is no longer the case in most of those countries today.

Since the launch of PISA, OECD has had considerable policy impact upon countries through their publications and commissioned studies on specific countries. Indeed, with such a large body of data and depth of expertise behind the OECD reports and conclusions, policy-makers and practitioners could find themselves at the receiving end of 'steering at a distance' through 'corporate governance' from this global institution. The narrative of the global knowledge economy imperative is seductive, but it is questionable whether policy reactions to PISA findings take the same form in different country contexts.

This report is the product of a short pilot project looking at the policy response to PISA in six case countries or regions. Deeper research questions have been generated from this project and we describe them in a separate report. High-performing countries (such as Canada, Shanghai-China) were contrasted with European countries that generally performed towards the average, but in which we knew there had been interesting policy impacts of the tests (as in England, France, Norway and Switzerland). As a methodology, comparisons through international tests do not celebrate difference – they are more likely to produce convergence in terms of what is seen to be valuable in educational terms. In the regionalized countries of Canada and Switzerland, there was some evidence of national pressure for more converged structures or data collection. Although assessment was undertaken in a number of Chinese provinces, only the results for Shanghai-China and Hong Kong-China were published in 2009 and there is little information about the policy response in substantive terms as yet. What we see, however, in the policy response in Shanghai-China is a reflective learning approach to PISA results, which is similar to that reported in Finland (not included as a case in this project as it is heavily reported elsewhere – eg Välijärvi *et al.*, 2002).

It is striking that PISA results seem to have been used for sabre-rattling political rhetoric to drive through educational reforms in some of the countries (as in France and possibly currently in England). Research on policy rhetoric relating to PISA would help us to unpick the political from attempts to create policy that will directly address the perceived failings of education systems as measured by PISA. Equally, a more thorough study of the temporal relationship between policy thrusts and PISA results would shed light on the plausibility of the claim that PISA causes policy. The relationship between substantive policy content and PISA results is not always immediately apparent in these cases, with broad-ranging reforms being linked with PISA results as a rationale – further analysis of these connections are also needed. Very different policy responses seem to have resulted from reactions to PISA. Some countries have clearly suffered shock in reaction to international test results (e.g. France and Norway). Furthermore, an investigation into the relationship between PISA and countries' performances in terms of knowledge economy measures is needed to find out whether this oft-made link is warranted. We propose a broader and deeper study of these issues in a wider range of countries, including some whose performances have been weak and those who have improved dramatically. An outcome from the longer-term study would be a taxonomy of modes of policy responses from different countries. Such a study would raise awareness of the variety of narratives that can be adopted in response to international tests, permitting a better-contextualised critique of policy responses in particular contexts, as well as a wider and more nuanced view of the governmental influences of the global institution, OECD.

ACKNOWLEDGEMENTS

We would like to thank the researchers, PISA administrators and policy-makers who commented upon the draft country profiles for this report.

- Stephen Dobson, Lillehammer University College, Norway
- Mark Gierl, Professor Educational Psychology, University of Alberta, Canada
- Jean-Pierre Jeantheau, ANLCI, France
- Marion Sainsbury, former PISA administrator, England
- Urs Vögeli-Mantovani, SKBF-CSRE, Switzerland
- Minxuan Zhang, Deputy-Director General, Shanghai Education Commission and PISA Coordinator for Shanghai

Sophie Pollard at the University of Bristol provided administrative support for the project. The research was funded by Pearson UK.

The views expressed in this report should not be taken as those of the Pearson organisation. This is an independent report commissioned from the Oxford University Centre for Educational Assessment.

INTRODUCTION

The OECD's Programme for International Student Assessment (PISA) was launched with its first international 'attainment and learning environment survey' in 2000, after several years of planning and organisation. There are several major players in PISA, other than the OECD itself. Surveys are coordinated by the governments of participating countries, who meet as the PISA Governing Board. The assessment materials are developed by leading subject experts, and the fieldwork is designed and managed by an international study centre (ISC); for PISA 2009 the ISC was a consortium led by the Australian Council for Educational Research (ACER) and CITO in the Netherlands.

PISA surveys are sample-based and focus on 15-year-olds. They take place every three years in the OECD member states, and in a growing number of non-OECD countries and economies. In each survey student achievement is assessed in three key 'life skills' domains: 'reading literacy', 'mathematical literacy' and 'scientific literacy' (see later for definitions). Each domain is given major emphasis every nine years (Table 1), being allocated two-thirds or more of the available testing space (OECD, 2009a).

Table 1: Timeline for major/minor literacy emphasis

	2000	2003	2006	2009	2012	2015
Reading literacy	Major	Minor	Minor	Major	Minor	Minor
Mathematical literacy	Minor	Major	Minor	Minor	Major	Minor
Scientific literacy	Minor	Minor	Major	Minor	Minor	Major

As well as student testing, questionnaires are used to gather information from students and teachers about the teaching-learning environment (including resourcing and curriculum), information about student learning outside the school, and attitudes towards subjects and learning. The information is used to contextualise the attainment findings and to explore the potential effects of background variables on student achievement.

This short report presents the findings of a small-scale pilot project whose principal purpose was to overview the PISA experience to date of six countries, or rather of five countries – Canada, England, France, Norway and Switzerland – and one country region, Shanghai-China. These particular countries were chosen for their varying approach to obligatory secondary education as well as variation in PISA performances, and they are countries about which we have a depth of understanding. We look briefly at national characteristics, including population size, educational structures, organisation of schooling, curriculum and assessment practices, as well as PISA performance. We also explore policy reactions to PISA findings, and begin to analyse the extent to

which PISA has directly or indirectly contributed to educational reforms that are currently underway. Before this, however, it is useful to describe the PISA programme itself and PISA performance trends of our selected case countries.

MOTIVATION FOR PISA

The motivation for PISA's existence was a perceived need to fill a gap in the extensive set of indicator-based information on education systems that the OECD has been providing now for around two decades in its annual *Education at a Glance* reports (see, for example, OECD, 2010a). The information in these reports extensively covers system input and process – financial and human resource investment, access rates, learning environment and school organisation – along with some readily measurable system outcomes – including progression rates through the different levels of the current system. Each report is considered to offer

> … a rich, comparable, and up-to-date array of indicators that reflect a consensus among professionals on how to measure the current state of education internationally.
>
> (OECD, 2010a, p.17)

The gap in this information was a set of indicators with which countries might evaluate the outcomes of their educational provision in terms of their students' achievement. As outlined below, many would argue that this gap has not yet been filled by PISA because of ongoing issues to do with measuring student attainment and interpreting the results.

National assessment programmes pre-existed PISA in some of the most developed countries in the world, including the USA, The Netherlands, the UK, Canada and New Zealand. These could have provided relevant achievement data, but the information from such programmes was available for a mere handful of OECD countries, and it was not always available for the same stage in schooling or in the same form. The subject frameworks underpinning the various national programmes differed, reflecting differences in national curricula and values, as therefore did the tests that were used. In other words, the outcome evidence available from national sources was only partial, and was not directly comparable. It could not be used to address the gap in the OECD's indicator set.

The other potential source of information was the International Association for the Evaluation of Educational Achievement (IEA), which has a long history of trans-national attainment surveys, and which by the late 1990s had been able to put two survey programmes onto an established footing. The Trends in International Mathematics and Science Study (TIMSS) has been running a four-year survey cycle since 1995 (for the international reports on the 2007 survey see Mullis *et al*., 2008a, b). The Progress in International Reading Literacy Study (PIRLS), meanwhile, is operating on a five-year survey cycle. Surveys took place in 2001 and 2006, with a third underway in 2011 (see Mullis *et al*., 2007 for the international report on the 2006 survey). However, both of these survey programmes rely on participating countries to cover the costs of their participation, and there is no guarantee that any particular country will participate on every occasion. Moreover, PIRLS operates only in the primary sector (10-year-olds), whilst TIMSS focuses also on the lower secondary school (14-year-olds). The two programmes also, by design, tie their assessment frameworks as closely as they can to the 'international commonality' in school curricula, something that the OECD wanted to avoid.

As a result, PISA was born. The programme was conceived and designed to plug the outcomes indicator gap by measuring the attainment of 15-year-olds in the key areas of reading, mathematics and science. Age 15 was chosen as the target age for testing, because it marked the end of obligatory schooling in most OECD countries, although it is no longer the age at which most young people complete their education in the OECD countries. The choice maximised the policy value of cross-border attainment comparisons for this reason – it also usefully facilitates the successful implementation of surveys, by avoiding any need to assess young adults within a less captive environment than the school.

In practice, not all the students tested in a PISA survey will have reached the end of obligatory schooling by the time they are assessed. Target students – universally referred to as 15-year-olds – are actually required to be between the ages of 15 years and three months and 16 years and two months at the start of the assessment period, irrespective of where they might be in the education system by that time. In some countries, including France, a proportion of students tested will not be about to finish their obligatory schooling at age 15, having been held back for a year or more during their secondary schooling because of 'inadequate performance' (in France almost half the age-group is affected in this way, with boys repeating years in greater proportions than girls – Bourny *et al*., 2002). Wagemaker (2008, p.273) shows that 15-year-olds are distributed across a range of school grades within a particular country and that this distribution varies across countries.

PISA'S ASSESSMENT FRAMEWORK

It was intended that PISA would not be tied to school curricula, but that it would focus more broadly on the knowledge and skills that young people would need in order to function adequately in society throughout their adulthood,

> PISA represents a commitment by governments to monitor the outcomes of education systems through measuring student achievement on a regular basis and within an internationally agreed common framework. It aims to provide a new basis for policy dialogue and for collaboration in defining and implementing educational goals, in innovative ways that reflect judgments about the skills that are relevant to adult life.
>
> (OECD, 2009a, p.9)

This commitment to 'skills for life' is intentionally reflected in the labels given to the subject domains assessed by PISA. These are defined as follows:

- *Reading literacy*
  An individual's capacity to: understand, use, reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

- *Mathematical literacy*
  An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen.

- *Scientific literacy*
  An individual's scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to draw evidence-based

> conclusions about science-related issues, understanding of the characteristic features of science as a form of human knowledge and enquiry, awareness of how science and technology shape our material, intellectual, and cultural environments, and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen.
>
> (OECD, 2009a, p.14)

Despite the intention to look beyond curriculum, what is taught in schools will impact upon students' performances and there will be more or less similarity between what is taught and the PISA skills domain across different countries (Nardi, 2008). As far as possible, PISA avoids a heavy emphasis on the assessment of factual recall and information retrieval, and focuses more than the IEA surveys do on knowledge application. It is the difference between what students have learned through their schooling (IEA) and what they can do with the results of that learning (PISA).

The consequence of this approach for surveys is a relatively high proportion of text-heavy multi-item questions with open-ended item formats, in place of what otherwise would be text-light more 'atomistic' supply or select item types. This has been noted by, among others, Ruddock *et al.* (2006). The latter observed, after comparing the style of mathematics and science items used in PISA 2000 and 2003, TIMSS 2003, key stage 3 tests in England, and the General Certificate of Secondary Education (GCSE), that:

> One significant difference between PISA and the other assessments is the amount of reading required. This is apparent in all the PISA science questions and in some of those in mathematics, and is not something that students in England would be familiar with.
>
> (Ruddock *et al.* 2006, p.122)

It has also been observed that a high level of numeracy is demanded in PISA reading tests, along with 'a significant requirement to interpret diagrams' (Ruddock *et al.*, 2006, p.29), both of which are features of reading assessment that are not familiar to English students, nor those in some other countries.

SURVEY DESIGN

The sample size per country was originally intended to be around 4,500 15-year-olds. Some countries, including the UK, have occasionally encountered difficulties meeting PISA's tight sampling criteria, because of school recruiting problems. Others have not experienced such challenges, and have actually chosen to augment their samples from the very first survey so that they could profit from the occasion to gather comparative attainment data for their different regions. For example, the French-speaking region of Switzerland in particular boosted its PISA sample to over 4,500 students, even though its share of the Swiss population is just around 20%. In Canada PISA participation totalled 23,000 students in 2009 (Knighton *et al.*, 2010). This large sample size was necessary to yield reliable indicators for the ten Canadian provinces: in some cases for school systems in both official languages (English and French).

Students are selected for testing using a conventional two-stage stratified sampling procedure. The school population is first stratified, usually by school type and size, but other variables can be brought in, depending on the country (England, for example, has additionally used prior school performance as evidenced in key stage 3 results – though this is no longer an option, given the

abandonment of cohort testing at this key stage). Schools are selected at random from within the various strata to produce a school sample representative of the population. Within each school around 30 students are randomly selected for testing from a list provided by the school; where schools are too small to provide 30 students all the 15-year-olds are tested.

The resulting school samples are often heterogeneous in terms of student ability, socioeconomic background and curriculum studied. Curriculum heterogeneity can arise from different ability-related examination preparation, such as foundation or higher tier GCSEs in England, or preparation in quite different types of school for academic or vocational qualifications, as in France and Switzerland, with students following different interest-related options within each type.

What does PISA do with the students once they are selected? Each student does not sit exactly the same PISA test. In each survey the number of items administered in total would consume several hours of testing time if every student were to be required to attempt them all, something that neither the students nor their schools would tolerate. PISA allows two hours of testing time per student, plus 30 minutes to complete a questionnaire. To maximize use of the allotted time, the assessment materials are bundled into several 30 minute 'clusters' within each domain, and individual students are randomly allocated four of these in the form of two pencil and paper test booklets. Cluster allocation to students uses 'matrix sampling', and is carried out in such a way that there is cluster overlap from one cluster grouping to another to facilitate the placing of all the individual items within a particular domain onto the same 'Rasch scale' (OECD, 2011c).

The Rasch scale is an Item Response Theory (IRT) model that depends on valid application based on the very strong assumption of 'item invariance': individual students are assumed to have inherent fixed 'abilities', and items have inherent fixed 'difficulties'. In principle, item difficulty cannot vary across students or student groups, irrespective of interests, culture or curriculum experience. Using relatively opaque techniques, the item analysis eventually locates both students and items on a common scale – a scale, by the way, with no absolute 'zero'. In PISA the scale is transformed and 'stretched', to have a mean of 500 and a standard deviation of 100. The effect of this is to exaggerate even the smallest differences in country performance estimates, differences that even if 'statistically' significant are not always educationally significant, but which have at times been over-interpreted by policy makers and politicians, leading to unnecessary concern about their countries' relative 'standards'.

PROBLEMS WITH MAKING INTERNATIONAL COMPARISONS USING PISA TEST SCORES

OECD claims that PISA is based upon,

> 1) strong quality assurance mechanisms for translation, sampling and test administration;
> 2) measures to achieve cultural and linguistic breadth in the assessment materials, particularly through countries' participation in the development and revision processes for the production of the items; and
> 3) state of the art technology and methodology for data handling.
> The combination of these measures produces high quality instruments and outcomes with superior levels of validity and reliability to improve the understanding of education systems as well as students' knowledge, skills and attitudes.
>
> (OECD, 2009a, p.10)

Notwithstanding, a number of issues have been identified that render international comparison of survey results problematical. These include: materials translation, the measurement model used for analysis, student sampling, domain representation, student motivation, and consequential validity (for overviews see, Goldstein 2004a, 2004b; Hopmann *et al*. 2007; Thomas & Goldstein 2008).

The originating language of the assessment materials is predominantly English, and the materials are translated into more than 40 languages for survey use. It is not only the quality of translation that could introduce comparability challenges, but also the inherent characteristics of the different languages themselves (see Hilton 2006, for examples and comment).

The validity of applying the Rasch model is another issue. Items are pretested (though not in all countries, since some countries choose not to incur the expense involved in this stage), one of the purposes being to identify 'differential functioning', that is items for which student performance is unusually high or low in one country or another, or for one gender or the other, compared with model expectations. Such items are removed from the set for survey use or revised, with potential consequences for the validity of the literacy domain that is ultimately assessed. Additional issues relate to the degree to which remaining 'model fit' is applicable to all student subgroups within individual countries (e.g. boys-girls, low-high ability, ethnic groups, and so on). Goldstein (2004a) argued that the unidimensional Rasch model may be limited and demonstrated that a two-factor multilevel model for mathematics PISA items produced potentially interesting and useful differences between countries.

Student sampling, i.e. population representation, is another problematical area, though one that has not attracted the same degree of attention in the literature (however, see Prais, 2003, 2007). One important feature to note is that PISA only assesses students who are currently undertaking schooling. While this is not a particular issue for developed countries, it can pose comparability difficulties where developing countries are concerned. In some countries the PISA target population can be much smaller than the actual population of 15-year-olds, and since students not in schooling tend to be from low socioeconomic groups the 'population attainment' results provided by PISA will not be directly comparable with those of countries with almost 100% coverage.

Other sampling issues, that affect all countries, concern the representativeness of the student samples that are tested in surveys, representativeness that is of the population of 15-year-olds in schooling. Imbalances occur occasionally when schools refuse to participate, and through pupil absence on the day of testing; these are addressed statistically through data weighting. But no amount of data weighting can control 'invisible' biases in a sample, such as absent pupils tending to be the lowest achievers. The testing motivation of students is a different issue, but an important one, as is general familiarity with the assessment format used in the surveys.

Apart from the question of domain reduction through the exclusion of 'differentially functioning' items after pretesting, another issue for domain representation is the number of items that are actually used in surveys, even in major domains. The relatively small numbers of items – for example, 131 in reading, 35 in mathematics and 53 in science in 2009 – have been noted as a weakness by several commentators, including Hutchison and Schagen (2007), and Ruddock *et al.* (2006).

SIX CASE STUDIES

In the following section we present and discuss the PISA attainment results for our six case countries/regions: Canada, England, France, Norway, Shanghai-China and Switzerland. Five countries are in the OECD, four of them being European. England, France and Norway have national systems of education in place, while Canada and Switzerland have federated systems. Switzerland, in particular, has a high focus on vocational education and training, with two-thirds of its upper secondary students following vocational programmes of one sort or another in the country's dual system.

Country profiles are included later in the report, providing comparative information about educational structures, assessment practices and PISA policy impact. One characteristic noted in the profiles is the nature of lower secondary education in each country. Four 'styles' of lower secondary education have been identified (Mons, 2007): 'separated', 'unified', 'group-adapted' and 'individualised':

- 'Separated' is where students are allocated to different types of secondary school on exit from primary education. At its simplest, one type of school is 'academic' and leads eventually to academic school leaving qualifications and university entry, while the second is vocational/technical and leads to vocational qualifications and apprenticeships or work. China-Shanghai has a separated system of secondary education. Switzerland has this model operating in the majority of its cantons, mainly in the German-speaking region, though in other cantons lower secondary education is group-adapted or uniform.
- A 'unified' system is one in which all students follow the same programme in the same type of school. France has a uniform system.
- In a 'group-adapted' system all students study in the same kind of school, but they might be streamed or set for different educational programmes or individual subjects. Canada and England have predominantly group-adapted systems.
- An 'individualised' education is one in which all students in principle follow the same programme in the same kind of school but teaching is targeted to suit individual students' needs. Norway, like other Scandinavian countries, has an individualised system.

TRENDS IN CASE PERFORMANCES

Five of the six cases profiled in this short report participated in all of the PISA surveys, i.e. 2000, 2003, 2006 and 2009. For these five countries, therefore, we can look at student attainment in all three literacy domains (Table 2). Shanghai-China participated for the first time in 2009, so that we have a recent 'major domain' reading literacy result in this case also, but no mathematics or science results.

Table 2 records the latest 'major domain' performance results for our countries/regions: mathematics in 2003, science in 2006, and reading in 2009. We note right away that Canada showed the best performance in all three literacy domains, with Switzerland doing essentially as well as Canada in mathematics in 2003. In science in 2006, France was virtually at the OECD average, Switzerland

and the UK above average, and Norway below average. In reading in 2009, all four European countries produced similar results, all close to the OECD average. Shanghai-China produced a particularly high performance for reading literacy in 2009, coming top of the international league table (see Table 3).

Table 2: Mean Scores in major domains

(Bracketed figures are the standard errors associated with the mean scores)

	Maths 2003	Science 2006	Reading 2009
Canada	532 *(1.8)*	534 *(2.0)*	524 *(1.5)*
France	511 *(2.5)*	495 *(3.4)*	496 *(3.4)*
Norway	495 *(2.4)*	487 *(3.1)*	503 *(2.6)*
Switzerland	527 *(3.4)*	512 *(3.2)*	501 *(2.4)*
United Kingdom		515 *(2.5)*	494 *(2.5)*
England		516 *(2.7)*	495 *(2.8)*
OECD	500 *(0.6)*	500 *(0.7)*	499 *(0.6)*
Shanghai-China			556 *(2.4)*
Total countries and economies	41 (30 OECD)	57 (30 OECD)	65 (34 OECD)

Sources: OECD 2004, 2007, 2010b; Gill *et al.* 2002, Bradshaw *et al.* 2010 for England

In addition to overall domain scores, results are produced for subdomains whenever the domain concerned is a major domain. In mathematics four subdomains were reported in 2003: 'quantity', 'space and shape', 'change and relationships' and 'uncertainty'. Bearing in mind that subdomain scores are calculated on the basis of rather few test items (just 86 items in total represented the whole domain for 'mathematics' in 2003) it was reported that Canada was relatively weak on 'space and shape' and relatively strong on 'uncertainty' compared with its performance in the other subdomains. France was relatively strong on 'change and relationships', while Switzerland was relatively weak on 'uncertainty'. Male students did better than female students on all four subdomains in all countries.

In science the process subdomains are 'explaining phenomena scientifically', 'using scientific evidence' and 'identifying scientific issues'. Both Canada and France did relatively well on 'using scientific evidence' compared with the other subdomains, with France doing relatively less well on 'explaining scientific phenomena'. Otherwise, among our set of countries/economies performance was steady across the science domain. As far as gender differences are concerned, in almost all countries surveyed female students did better than male students when 'identifying scientific issues', less well when 'explaining phenomena scientifically', and similarly when 'using scientific evidence'. Knowledge subdomains were 'living systems', 'physical systems' and 'earth and space'. The UK and France showed relatively better performance on 'living systems' than on the other two subdomains, with France relatively weak on 'earth and space'. Gender differences in favour of male students were strong in most countries for 'physical systems', with Switzerland showing one of the greatest disparities. Male students also generally did better than female students in 'earth and space', though less strongly than in 'physical processes'. Gender differences in 'living systems' were smaller and less consistent.

The reading subdomains are 'access and retrieve', 'integrate and interpret' and 'reflect and evaluate'. Canada and the UK did relatively better on 'retrieve and reflect' than on the other two subdomains, Norway did relatively better and Shanghai-China relatively less well on 'access and retrieve'. France and Switzerland showed stable performances across the subdomains. Female students produced significantly better performances than male students across all aspects of the reading domain in all countries.

Only for reading literacy are we able to explore attainment change with some confidence, since this has been the major domain twice in the survey series – in 2000 and 2009. Table 3 presents the results for our set of countries across the four surveys. A clear feature in the data in Table 3 is the 'bunching' of the four European countries – France, Norway, Switzerland, United Kingdom – with results close to the OECD average (with the exception of the UK, which did better in 2000 – although see the UK country profile for a discussion of sampling problems). Canada is the exception, with markedly higher mean scores than the rest, though with the gap apparently decreasing by 2009, when Canada's score of 534 in 2000 fell to 524 (this difference, however, did not reach statistical significance at the conventional 5% level). As noted previously, Shanghai-China had the highest mean score in the 2009 survey, at 556.

Table 3: Reading performance over the four PISA surveys

(Bracketed figures are the standard errors associated with the mean scale scores, while figures in boldface are countries' positions in the league table)

	2000	*2003**	*2006**	*2009*	*Change 09-00*
Canada	534 *(1.6)* ***2***	528 *(1.7)*	527 *(2.4)*	524 *(1.5)* ***6***	-10 *(5.4)*
France	505 *(2.7)* ***14***	496 *(2.7)*	488 *(4.1)*	496 *(3.4)* ***22***	-9 *(6.6)*
Norway	505 *(2.8)* ***13***	500 *(2.8)*	484 *(3.2)*	503 *(2.6)* ***12***	-2 *(6.2)*
Switzerland	494 *(4.2)* ***17***	499 *(3.3)*	499 *(3.1)*	501 *(2.4)* ***14***	+6 *(7.0)*
United Kingdom	523 *(2.6)* ***7***		495 *(2.3)*	494 *(2.3)* ***25***	
England	523 *(3.0)*		496 *(2.7)*	495 *(2.8)*	-28
OECD average-23	501 *(0.7)*	497 *(0.6)*	495 *(0.7)*	499 *(0.6)*	-2 *(5.0)*
OECD average-26	496 *(0.8)*	m	m	496 *(0.5)*	+1 *(5.0)*
Shanghai-China				556 *(2.4)* ***1***	
Total countries and economies	32 (27 OECD)	41 (30 OECD)	57 (30 OECD)	65 (34 OECD)	

** Given the rather small numbers of test items that have been used to assess domains when they have carried minor status, some caution is required when using the results in time series such as the one shown here.*

Sources: OECD 2001, 2004, 2007, 2010b, 2010f ; Gill *et al.* 2002; Bradshaw *et al.* 2010 for England

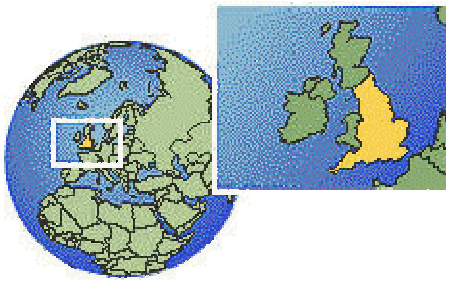
If we look at the published rankings, we see that Canada dropped from 2nd to 6th place, and France from 14th to 22nd. Switzerland, meanwhile, moved up from 17th to 14th, while Norway moved up a notch from 13th to 12th. These are the results that politicians look for first, and that they sometimes use, appropriately or otherwise, to initiate debate about system reforms and on occasion to justify reforms. It is these results that should be treated with particular caution, however. This is because rank positions can quickly shift from one survey to the next, not necessarily, and certainly not only,

in response to any genuine change in student attainment, but also in response to the set of countries and economies that happen to participate.

LIMITATIONS OF THE CURRENT STUDY

The current study examines just six countries/regions using a case study approach. The overall goal is not to generalize the findings, but to profile countries from distinct educational systems, with different assessment and accountability traditions. Moreover, since only reading literacy has been assessed as a 'major' domain more than once to date, any statements about 'trends' and country reactions to these must be confined to this one domain, and treated with caution. That said, the essentially qualitative analysis offered here has served to raise questions for a larger more in-depth analysis of international test experience and policy reaction.

England Country Profile



At first glance, England's PISA results appear to have been in decline: initially, the country was ranked 4th, 7th and 8th (in science, reading and mathematics respectively), but those rankings have dropped in successive studies to 16th, 25th and 27th. However, the number of participating countries has risen since 2000 - from 32 to 65 – and this is at least part of the explanation. OECD's official position is that England's trends should only be reported from 2006, due to sampling problems (outlined below). England's results in 2006 and 2009 are average for reading and mathematics and the change over time is non-significant. For science, England is significantly higher than the OECD average in 2006 and 2009 and there is a non-significant change between those years. The quandary for England is how national test results such as key stage tests and GCSEs have shown rising outcomes whilst comparisons with other countries have not changed. A great deal of debate has focused upon 'teaching to the test' for the national, high-stakes qualifications. Stakes for these qualifications are high for students at GCSE and for schools for key stage tests.

Educational Structure

Categorisation: **Group-adapted system**

Children in England must attend school between the ages of 5 and 16. At age 11, there is a transition to secondary school with all children attending secondary school from age 14 to 16. Pupils of PISA-taking age are therefore typically in their second year of upper secondary school. At this level, across the UK, in 2008 the gross enrolment ratio was 96 with a gender parity of 1.06. Pupils progress with their age cohort through schooling. A large proportion of the cohort chooses to attend school or college until age 18. In total, there were over three million pupils in secondary schools in England in 2010, with an average pupil-teacher ratio of 15.7:1. In 2008, across the UK, 17% of upper secondary pupils were enrolled in private general education. Spending per pupil was £4,320 on average in 2007, with public expenditure on education at 5.4 percentage of GDP in the UK in 2007. There were 221,300 qualified teachers working in secondary schools in 2010. OECD figures show that England's spending on education in the 14-18 phase is similar to the OECD and EU average.

Just under one million students aged 16 – 18 attended further education colleges in 2009. Vocational education is typically delivered in further education colleges once students are over 16-years-old. Across the UK, in 2008, there were 717,002 pupils enrolled in technical or vocational education at the upper secondary level.

Despite a number of recent initiatives, teaching is not considered to be as well paid or respected a profession as in some other countries: the average salary is approximately £35,000, which is lower

than the average for workers with degrees (OECD, 2011a, p.13). Most teachers hold a degree level qualification (83%) but there is no requirement for a higher degree. Under a quarter of teachers in secondary schools are aged 50 or over.

Assessment Structures and Formats

- There is a great deal of angst about over-assessment in the English education system, with the phrase 'assessment as learning' having become associated with the fact that students are being taught to the test and are learning test materials.
- There are a plethora of qualifications, with over 175 regulated examining bodies.
- Teacher assessment forms part of most of the qualifications and is usually moderated.
- Assessments are often short-answer or extended answer format external tests.
- High-stakes external tests are conducted in English and mathematics at age 11, in a range of subjects at age 16 (GCSE, etc.) and at age 18 (A-level, etc.)
- Compared with the Key Stage tests and GCSE, the assessment formats of PISA have a lot of reading demand for science and mathematics and there is more of an emphasis upon non-continuous text and non-fiction (Bradshaw *et al.,* 2010, p.14).

Responses to PISA

- There are relatively few media articles in England about the PISA results and those that are published are sometimes written by academics (e.g. Mortimore, 2008). Some articles have featured a comparison of devolved countries within the UK, especially of England with Wales. Educational policy has diverged in the devolved constituencies and the dropping of league tables in Wales has received particular attention.

  - 'England needs a rocket under them to improve their PISA scores', says Minister for Education.
  - England's PISA results are average and stagnant whilst national assessments show a rising trend.

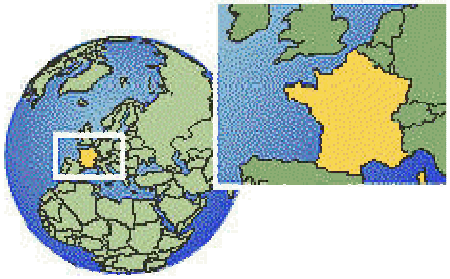
- The UK did well in the first two rounds of PISA and there was relatively little commentary in the press, or influence upon policy (Grek, 2009). The Labour Government was largely pleased with the findings and policy-makers claimed that PISA had little influence over policy. However, controversy surrounded the sampling for the 2000 and 2003 rounds of data collection and a report prepared for the Department for Education and Skills concluded that biased sampling had resulted in an over-estimation of England's performance (Micklewright and Schnepf, 2006, p.iv). Given this controversy, it is perhaps unsurprising that the 2005 White Paper on *14-19 Education and Skills* (DCSF p.14) referred to TIMSS and PIRLS improvements, rather than performances on PISA.
- In 2006 and 2009, the UK's rankings and performance were average. Interpretation of the results by policy-makers and Government need to be contextualized by the political cycle. From 2010, the government changed to a coalition between the Conservative and Liberal Democrats and a prevailing rhetoric has been fixing the 'mess' left by the previous government. The Minister for Education, Michael Gove claimed that the previously published White Paper, 'The

Importance of Teaching' was almost a bespoke response to conclusions from the 2009 survey and he draws comparison with the PISA performances and findings in other countries (e.g. Gove, 2011).

- Certainly, the 'Importance of Teaching' White Paper addressed issues associated with teacher qualifications and school autonomy within an accountability framework. In a *Times Education Supplement* article (Gove, 2010), the Minister pointed to the fact that a fifth of 15-year-olds reached unacceptable levels of literacy and numeracy. Countries such as Chile, Mexico and Brazil that improved their literacy results, often did so by reducing the proportion of low achievers (OECD, 2010f, p.13) and there is no direct strategy for doing this presented in the White Paper. Equally, the likely effectiveness of the Government's policy selections has been questioned, with Andreas Schleicher arguing that investment in the professional development of the current workforce would produce quicker results (Stewart, *Times Education Supplement*, 2011, p.18). Furthermore, there are many policies contained in the White Paper that have no obvious relationship with the PISA findings. Thus, use of the word 'bespoke' was at best a rhetorical flourish as a description of the match between recent policy in England and the PISA 2009 findings.
- The coalition government has signalled in the White Paper and elsewhere that it values international comparisons. The government has required the examinations regulator (Ofqual) to conduct research on the comparability of English examinations with those abroad, to ensure that they meet international standards.
- Ofqual has argued that PISA rankings have to be interpreted cautiously, citing the problems raised in academic articles.

France Country Profile

France has performed close to the OECD average in all subjects in PISA, though with a flatter than average scale score distribution, reflecting higher than average proportions of very high scoring and very low scoring students. Performance in reading and science has not significantly changed over testing cycles, while for mathematics there has been a small statistically significant decline in scores.



Educational Structure

Categorisation: **Unified**

France is a country of over 62 million people, with French the predominant language. Its education system has 11 years of compulsory education until students are 16 years of age. The majority of students of PISA-taking age are in the first year of upper secondary education or final year of lower secondary education. At secondary level in 2008, 78% of the 5.3 million students were enrolled in state schools, with a further 22% in private schools funded largely by the government (teachers' salaries), leaving just fewer than 1% in independent private schools. In state schools the average class size is just over 24 students for lower secondary level and 27 for upper Secondary level; the student-teacher ratio is 14:6. Upon progression to upper secondary school, two-thirds of the students continue an academic programme in a *lycée général*, the remaining one-third embarking on vocational/technical programmes in a *lycée professionnel*.

Public expenditure on education as a percentage of GDP is 5.5% with a per pupil spend as a percentage of GDP per capita of 26% for the secondary level. Eighty percent of secondary level teachers have completed at least a three-year postsecondary degree (*licence*), and most of them also have a teaching certificate. In 2009, however, the government decided that all teachers would be required to undertake university education to Masters level.

Assessment Structures and Formats

Throughout lower secondary schooling student assessment takes the form of continuous assessment, and there are no final-year examinations. Tests are produced and marked by the students' own teachers, and homework is also taken into account in student evaluation. In the upper secondary school examinations for the high-stakes *Baccalauréat* take place at the end of the final year (*terminale*). The examinations are nationally produced and externally marked.

In addition to assessment for student certification, France also has a system of national assessment in operation in the primary sector (in CE1, 2nd year, and CM2, end of primary), with the principal

aim of system monitoring. School participation is compulsory, and teachers in participating classes are paid for the extra workload. The programme began in 2009, and focuses on literacy (reading *and* writing) and mathematics. Aggregated results are reported at national, regional and departmental levels. At this point in time school league tables are not published or planned. This is therefore low stakes assessment for students and teachers. Nevertheless, there has been an important degree of teacher opposition to the new programme. One reason is that it has replaced a long-running test-based census that took place at the beginning of the school year (in CE2, 3rd year in primary, and 6e, 1st year in lower secondary) that provided information for teachers to use in programme planning. Teachers have lost this 'formative' assessment support, but they see no gain to justify the loss. At CM2 in particular, with testing taking place in January, the assessment comes too late to inform programme planning and too early to provide a valid picture of achievement by the end of the school year.

Another summative assessment programme was launched in 2003: *Cedre*. The aim of this sample-based programme of annual surveys, in which around 10,000 students are tested each time, is to assess student achievement against the national curriculum at the end of lower secondary schooling. Each year a different subject is assessed (French, mathematics, science, history-geography, foreign languages), on a six-year cycle (for example, 'written and oral comprehension' was the focus in 2003 and 2009).

Teacher assessment, whether formative or summative, by judgement or test-based, is popular in this country. External summative assessment, other than for the *Baccalauréat,* is mistrusted, because of its potential accountability threat to schools and teachers themselves. The *Baccalauréat*, for its part, is almost revered by society at large, including teachers. It has the same 'gold standard' status that the A-level has in England, and is beyond criticism.

There is systematic continuous assessment, regular tests and a school report over the final two years of lower secondary education that contributes to the attribution of one or other of two national school certificates (generally at age 15) – the *Diplôme National du Brevet*, or simply the *Brevet*, and the *Certificat Formation Générale (CFG). Brevet* certification depends on academic performance in those final two years in the *collège*, and also since 2006-07, on the student's behaviour, attendance and general participation in school life – as observed throughout the period of continuous assessment and recorded in a *note de la vie scolaire.* When introduced, this innovation caused some controversy, not everyone in the country considered it appropriate for a child's academic record to be 'tainted' by marks for behaviour.

A professional/technical *brevet* is available for students studying agriculture. The CFG is intended for students with learning difficulties who are deemed not to benefit from regular schooling; most move into vocational education, into an apprenticeship, or directly into work when they leave the *collège.* These end-of-*collège* qualifications are in some sense high-stakes for pupils from disadvantaged groups, because they represent evidence of a successful conclusion to their lower secondary education. In practice, however, *brevet* certification is low stakes, and has little influence on the future educational path of most students, 80% of whom achieved the *brevet* in 2009.

In upper secondary education there are essentially two paths available: continuing academic education leading (mostly by age 18) to the *Baccalauréat* (BAC), and vocational education leading to a *Certificat d'Aptitude Professionnelle* (CAP, with 200 different specialisms in the industrial, commercial and service sectors) or, increasingly, to a vocational baccalaureate. In reforms intended

to increase the number of students leaving school with a baccalaureate qualification, the range of options has been broadened over recent years. The traditional three academic options – scientific, social sciences, literature and creative arts – are now accompanied by vocational, technological and professional. In an age-group more than 70% of students are candidates for a baccalaureate, and since 86% succeed this means that two thirds of the entire age-group achieved the baccalaureate in one or other of its forms. The Government's target for BAC completion is 80% of the age population, and its other target is not leaving upper secondary school without some kind of diploma.

Responses to PISA

There is always media reaction when international survey results are published. But reaction to the first PISA results was less strong than it might have been, given that France had already suffered not 'PISA shock' but 'IALS shock' several years earlier, when the results of the International Adult Literacy Survey were announced. France, along with the UK incidentally, was identified as among those countries with very high rates of adult illiteracy (more than 40%), so high that no-one in the country believed the finding could be true (IALS was repeated in 2002 in parallel with a field trial for France's own adult literacy survey (IVQ) and rates were later revised downwards to 15%).

- 'IALS shock' in France was so great that it overshadowed the possibility of 'PISA shock'.
- The international surveys have been used in support of a wide-ranging review and reform of education in France.
- Government ministers have been guilty of over-interpreting and exaggerating survey results, portraying France in a more negative light than is justified.

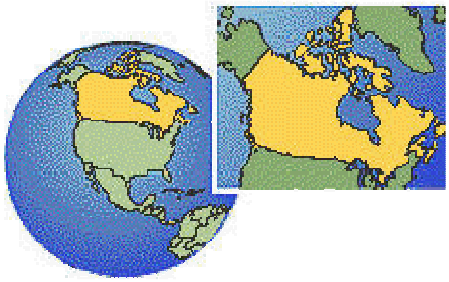
In PISA France's reading literacy performance is consistently around the OECD average, as is its performance in maths and science. Despite this, there has been a PISA-triggered debate about the educational system, with the practice of requiring weak students to repeat years (*le redoublement*) coming in for particular attention (Mons & Pons, 2010). The Government (e.g. Xavier Darcos, former Minister for Schools - *ministre délégué à l'enseignement scolaire*) has been criticized for using international survey results from TIMSS, PIRLS and PISA to justify its system reforms in recent years, with exaggerated statements to the media about France's standing in the surveys, such as:

> Today, when we look at international comparisons, even though France spends much more than its partners she is classed among the very last in all the international league tables
>
> (France 3, March 2008), and
>
> We find ourselves right at the bottom of the league table
>
> (France-Inter, April 2008).

France has produced stable performances in reading literacy in PISA surveys, with scale scores little different from the average for the OECD countries and well within margins of error. Nevertheless, a

finding that has concerned policy makers and politicians is the wide spread of pupil scores: the score distribution for reading literacy in 2009 was even flatter than in 2000. This means that while there appears now to be a higher proportion of pupils in the top band of performance, which is good news, there is equally a higher proportion in the lowest band, which is bad news. In response, according to a press conference on PISA 2009 given in December 2010 by Luc Chatel (French Education Minister), in a first plan of action the Government is to 'refocus on fundamentals', by reforming the primary school curriculum, introducing a strategy to fight illiteracy, and launching an initiative in science. A second plan of action aims to introduce personalised learning assistance throughout the system, to help students in difficulty. The third plan of action is to 'personalise' educational resourcing, such as making schools autonomous managers of their own budgets.

Canada Country Profile



Canada scored significantly above average in the 2009 PISA survey in reading, mathematics and science, and this has not changed significantly over the survey series.

Educational Structure

Categorisation: **group-adapted comprehensive educational system (setting)**

Students of PISA-taking age in Canada are in their first year of upper secondary education. The student population at this school level is 1,788,479 with a gross enrolment ratio of 104, a gender parity index of .98 and pupil teacher ratio of 18:4. At the upper secondary level, 5.8% of students are enrolled in private general programmes. The public expenditure on education as a percentage of GDP is 4.9%.

In Canada, the second largest country in the world in geographical terms, there is no integrated national education system. The ten provinces and three territories assume responsibility for education. Thus, there are systemic differences in the organisation of school systems and overall educational structure across jurisdictions. Education is compulsory in most of the country until age 16, with publicly-funded education available from Grades 1–12; however, two provinces (Ontario and New Brunswick) have instituted compulsory education until age 18. Most jurisdictions have middle school (junior high), with transition years generally at grades 6 or 7 and grades 9 or 10. Notably, the predominantly French-speaking province of Quebec has a different educational system than the rest of Canada, with compulsory education until grade 11. Thereafter, students may attend a two-year pre-university programme or a three year technical programme at a publicly-subsidised or private college.

In 2004/5 the number of teachers in Canada totalled 310,000 with the number of females far exceeding males in all education levels, and most full-time teachers being in the 30–59 age group (Canadian Statistics Education Council, 2007). Teachers are accorded some social prestige in Canadian society, although generally not as high as in Europe (Wilson & Lam, 2004). Overall, Canadian teachers need a postsecondary degree and a 1–2 year teacher education programme to obtain teacher certification. Ontario is currently the only province with a qualifying exam for new teachers.

Some provinces have academic and vocational streaming in the last two years of high school (Wilson & Lam, 2004). Overall, streaming practices tend to advantage academically-oriented students over those with practical or applied interests (Schuetze & Sweet, 2003).

Assessment Structures and Formats

The federal government is constitutionally prohibited from imposing a national education policy. Thus, the provinces and territories vary in their curriculum coverage, assessments, and accountability practices, reflecting a diversity of approaches. At the same time, the trend toward increasing large-scale assessments brings about pressures of standardisation, uniformity, and the provision of national statistics to support evidence-based decision-making (Volante & Ben Jaafer, 2008).

Although provincial and territorial achievement tests are administered at different grade levels with different subject coverage for different stated purposes (e.g. monitoring student achievement, gatekeeping), they tend to be criterion-referenced. Teachers are often involved in item development and scoring (Klinger *et al.*, 2008). This reflects some degree of alignment between provincial assessments, curricular goals, and classroom practice.

Provincial and territorial assessments can have high-stakes consequences for secondary students in particular (e.g.30–50% of graduating students' final grades). Consequences are lower-stakes for teachers and administrators, who are neither officially rewarded nor sanctioned for their performance. However, in some contexts, school, and district rankings are published in the media, which raises the stakes.

The Pan-Canada Assessment, conducted by a council composed of provincial and territorial representatives, tests a random sample of 13-year-olds across the country in reading, math, and science. As with PISA, one major domain and two minor domains are designated per administration (once every three years).

Students from all 10 provinces participate in PISA in either English or French (according to the language of the school system), with Canada's four most populous provinces (Alberta, Quebec, Ontario, and British Colombia) out-performing the others. Some provinces (Alberta, Quebec, and Ontario in 2011) will additionally participate in other international comparison assessments (PIRLS, grade 4; TIMSS, grades 4 and 8). Canada's sparsely populated territories have not participated in any international comparisons tests to date.

Responses to PISA

Two major headlines have dominated Canadian national media coverage (e.g. Parkin, 2010; Simpson, 2011). One is that educators should 'take a bow', since Canada performs well on PISA relative to other OECD members. The narrow achievement gap between students from different social background feeds into the national meta-narrative of the positive effects of Canadian values of inclusiveness and diversity on PISA performance. On the other hand, media reports have also emphasized the need to 'pull up our socks' due to slippage in scores over time in some provinces.

> The pervasive use of large-scale assessments in Canada is fostering a climate of increased standardisation procedures across the diverse provincial and territorial jurisdictions.

As the international media has focused on the West losing its competitive edge, coverage of Canada has been mostly cursory. However, a few provinces have been singled out as models of success (e.g. Alberta and Ontario in the British and German press, respectively; see Evans, 2010; von Törne, 2008).

Provincial responses to PISA have been mediated by the strength of their performance relative to that of other provinces, with high-performers tending to use the results to validate the effectiveness of their educational policies (e.g. EQAO, 2010). The lowest performer, Prince Edward Island, implemented provincial achievement tests in response to task-force recommendations conducted as a result of relatively low PISA results (Kurial, 2005). New Brunswick instituted 10 year targets to improve its PISA ranking within Canada from 10th in 2000 to among the top three by 2013 (New Brunswick Department of Education, 2009).

A federal-provincial report highlighting how both the country and its constituent provinces are 'measuring up' on the international stage emphasises the importance of PISA results for evaluating returns on government investment in education and in gauging the competitiveness of Canada's emerging workforce (Knighton *et al*., 2010). Echoing media coverage, the overall message was that Canada has performed well but should not be complacent due to declining performance over time or, in cases where performance is stable, outperformance by other countries.

In 2009, the Pan-Canadian Education Indicators Program introduced a set of harmonized indicators for provinces/territories to report annually to the OECD, including a PISA-based 'excellence in student achievement' indicator (Statistics Canada & CMEC, 2010).

Norway Country Profile

Norway performed significantly above average in 2009 for reading, but had average scores for mathematics and science. Changes over testing series were not statistically significant for reading or mathematics, but there has been a significant improvement in science scores.

Educational Structure

Categorisation: **integrated**

Students of PISA-taking age in Norway are in their final year of lower secondary education. This school level is comprised of 184,876 students with a gross enrolment ratio of 96, a gender parity index of .99, and a student teacher ratio of 15:1. At this level, 3.1% of students are enrolled in private general education. Public expenditure on education as a percentage of GDP is 6.7% with a per pupil spend as a percentage of GDP per capita at 26% for the secondary level.

The Norwegian education system has 11 years of compulsory education until students are 17 years of age. Vocational and technical programmes are offered at the upper secondary level of which there are nine types. These include, for example, building and construction, and health and social care. In 2008, 131,874 students were enrolled in such vocational education programmes.

The teacher workforce is 24,820 strong at the lower secondary level, with 59% of teachers above 45 years of age and 45% above age 55. Forty-nine per cent of teachers are female. Teacher training and/or a higher degree are generally required to teach, though 8.2% of teachers currently hold neither qualification.

Assessment Structures and Formats

The assessment system in Norway is relatively underdeveloped. The 2011 OECD report commented,

> It is well documented that the Norwegian school system needs to develop its assessment practice and assessment culture.

Assessments are mainly teacher-based and research has shown wide variation in how overall marks are determined between subjects, schools and regions. Marks and grades are not given to students before Year 8. National tests are low stakes for pupils as they are only reported at authority level. While there is increasing central involvement this is largely in terms of guidance and funded projects to improve teachers' assessment skills, both summative and formative.

The devolved school system leads to a wide range of assessment practices. The government's policy initiatives emphasise both summative and formative approaches. Testing and examinations become more frequent in upper secondary where they play a central role in the end-of-school assessment of achievement. These are largely teacher set and marked. Marking is on a 1-6 scale with 6 the highest. The marks should be based on achievement but there is still a legacy of including effort and attitude in the assessment – though this should now be recorded and communicated separately under an 'order and conduct' mark. In the vocational streams, in which over half the upper school students are enrolled, assessment will include occupational competencies and may lead to a craft/journeyman's certificate after two years full time in school and two years as an apprentice. These tests involve a 2-5 day practical activity involving planning, implementation and documentation. The pass rate is around 92 per cent.

There is also an annual student survey for pupils from Year 5 onwards. This focuses on perceptions of well-being, school democracy, motivation and the guidance and support received in learning subjects. It is viewed as an important indicator of school climate.

National tests have been introduced, with considerable initial opposition and boycotts, in Years 5, 8 and 9. These tests are in reading, mathematics and English, though they are intended to be a test of basic skills and be cross-curricular in their range rather than subject specific. These are not used as a national accountability measure for schools (there is no publication of school league tables). Analysis is at municipal and county level.

Diagnostic 'mapping' tests have recently been introduced to offer more detailed data on where individual pupils are in their learning. This was partly in response to criticism that schools received only limited individual feedback from the national tests.

In the national examinations in Year 10, students are randomly drawn for a centrally given written examination in one subject (Norwegian, mathematics or English) and a locally given oral examination in one subject. For the remainder of their subjects the overall achievement mark for the course is recorded by teachers on the Lower Secondary school leaving certificate.

- The PISA results of 2000 and 2003 left Norway in a state of 'PISA Shock' as its well-resourced education system generated results below the OECD average and below its Scandinavian neighbours.
- The incoming government was able to use this as a lever to introduce reforms to both curriculum and assessment in a highly devolved system.

Responses to PISA

The 2000 and 2003 PISA results led to what is widely referred to in Norway as 'PISA Shock'. This was the result of Norway being ranked below the OECD average and below other Scandinavian countries, despite generous spending on education and teachers reporting high levels of satisfaction with their self-efficacy (OECD, 2009b). This led to newspaper headlines in the Norwegian press such as 'Norway is a loser'. The incoming Minister of Education commented on the 2000 results that were issued as she took office,

> Norway is a school loser, now it is well documented. It is like coming home from the winter Olympics without a gold medal … and this time we claim the Finnish participants have been doped. (Aftenposten National Daily, Jan. 2001, in Sjøberg 2007, p.2).

In comparative terms Norway remains a devolved system with only limited direct central intervention. At present, attempted improvements are often introduced through large-scale pilots of assessment and/or curriculum initiatives (e.g. the Assessment for Learning initiative). This Minister of Education proceeded to introduce a series of educational reforms which were legitimized by reference to international testing, mainly to PISA. The Secretary of State under this Minister subsequently produced a book on the 'inside story' of the reforms. Under the heading *The PISA Shock* he reported:

> With the [publication of the] PISA results, the scene was set for a national battle over knowledge in our schools…for those of us who had just taken over political power…the PISA results offered a 'flying start.' (Bergesen in Sjøberg, 2007, p.4).

Whilst there has been a gradual improvement in results, with the 2009 results placing Norway above average on the reading scale and average on the science and mathematics scales, this again has led to political rhetoric about the need for continued improvement and educational reform.

The PISA Shock led to a series of reforms of both curriculum and assessment. A national quality assessment system (NKVS) was introduced in 2004 which included national tests and a web-based portal *(Skoleporten)* for presentation of data for school evaluation. 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 *Assessment for Learning*.

The OECD *Country Background Report on Norway* (2011b) acknowledges the influence of international studies such as PISA which 'showed Norwegian students had poorer basic skills than the national authorities expected. A key objective in developing the national tests was to give the national authorities a tool to follow how the Norwegian school is succeeding in developing the students' basic skills' (p.28).

What does not appear to have been affected are the highly devolved school system and the upper school structure with its general and vocational tracks.

SHANGHAI PROFILE

Shanghai is not representative of China as a whole, because it is probably better resourced than other regions. However, its significant achievement, in topping the league table in PISA 2009, can have far-reaching effects nationwide on China's educational policies with regard to curriculum reform and educational equality. As a 'first' timer and 'first' performer in PISA assessments, Shanghai has taken a cautious learning approach towards its participation in PISA and its significant achievement in 2009. The confirmation of lower school-level variance is particularly welcomed as evidence of educational equality that the city and the country are striving to achieve and demonstrate to the citizens and to the world.

EDUCATIONAL STRUCTURE

Categorisation: **separated (regular/comprehensive/academic vs. vocational education) in senior secondary education**

As Shanghai is the only region of China included as a case, we report below the total population in Shanghai and some basic statistics of its secondary education in terms of number of students and teachers, students/teacher ratio, and number of regular/comprehensive and vocational schools.

Population: Shanghai is China's largest city. At the end of 2009, there were 19.2 million permanent residents in Shanghai, including 13.8 million with registered permanent residency (excluding those permanent residents who are away from Shanghai for more than 6 months) and 5.4 million residents without registered permanent residency. According to the 6th National Census (1st November 2010), Shanghai's population is now over 23 million. It seems that its population is ever increasing.

Education system organization: By the mid-1990s China had achieved nine-year compulsory education (i.e. 6 years of primary education plus 3 years of junior secondary education). Shanghai has some slight variations, its formal education is organized at three levels: primary (5 years), secondary including junior (3 years) and senior secondary (3 years), and tertiary education. Nearly 99.9% of junior secondary school students in Shanghai progress on to senior secondary education. At senior secondary education level, there are two separated streams: regular/comprehensive/academic vs. vocational education which includes three main types of school, i.e. specialized polytechnics (*zhong zhuan*), vocational (*zhi ye zhong xue*) and technical/crafts (*ji gong xue xiao*) secondary schools.

Secondary school student population: In 2009, there were around 771,000 enrolled students in schools of secondary education level, including 604,000 in general/comprehensive/academic secondary schools (*pu tong zhong xue* in Chinese), 426,000 in junior secondary schools and 178,000 in senior secondary schools. The rest (around 22%) are mainly in specialized polytechnics, vocational, and technical/crafts schools– the three main types of schools providing vocational training at senior

secondary level. Overall, the number of enrolled students at the secondary level has decreased substantially (about 27% since 2000), from 1,053,000 in 2000, 803,000 in 2008, to 771,000 in 2009; and the rate of decrease in the number of students in vocational training (about 35% since 2000) was much larger than that of comprehensive secondary education (around 24% since 2000).

Table 4: Secondary school student population

(x1,000 students)

Year	2000	2008	2009
(a) Specialized polytechnics, vocational and technical/crafts schools at senior secondary level	257.4	182.0	167.0
(b) General/Comprehensive/Academic secondary schools including:	795.4	617.7	603.7
Senior secondary	239.4	192.6	177.6
Junior secondary	556.0	425.1	426.1
Total at secondary level education	1052.8	802.9	770.7

There are 762 regular/comprehensive secondary schools: including 140 secondary schools which offer both junior and senior secondary education, 133 senior secondary education only, 345 junior secondary education only, and 144 schools offering both primary and secondary education.

There are 109 secondary schools providing vocational training: including 70 polytechnics, 26 vocational and 13 technical/crafts schools. The number of secondary level schools providing vocational training has also decreased.

Table 5: Schools focusing on vocational training at secondary level

Year	2000	2008	2009
Specialized polytechnics	84	73	70
Vocational schools	60	28	26
Technical/crafts schools	115	28	13
Total number of schools focusing on vocational training at secondary level education	259	129	109

Teachers and support staff in regular secondary schools: 67,600 teachers and support staff with 50,500 teachers (16,900 teachers in senior secondary schools, and 33,600 in junior secondary schools). On average, the student and teacher ratio in secondary education was about 13:1 in 2009. The decline in student population has provided Shanghai with opportunities for small class teaching. All teachers in secondary education are degree holders with professional certification; and many teachers hold Master's degrees. In addition, all teachers are required to take in-service professional development courses of 240 contact hours within a 5 year term.

Assessment Structures and Formats

Education and competitive summative examinations and credentials have been valued highly by the Chinese, ever since the formal establishment of the Civil Examination system in 603 AD, or even much earlier. Confucian's learning philosophy has a profound influence in every aspect of Chinese education and its value system. Education has been considered by many as the essential or only ladder for social mobility. Although Shanghai has been a forerunner in educational reform in China in areas such as curriculum change, reducing teacher and student workloads, small class teaching, compulsory teacher professional development, implementation of neighbourhood/catchment area admission policy in primary and junior secondary education thus removing the pressure of end of primary education examinations which are still prevalent in some provinces and regions, setting its own university entrance examinations independent from the national system since 1985, just to name a few. However, it is still fundamentally a typical Chinese education authority with a strong emphasis on using examinations as a lever or baton for curriculum change and social and political agenda setting. Although the number of places in higher education institutions has increased drastically nationwide as well as in Shanghai, with very generous admissions quotas in Shanghai universities and colleges (N=66 in 2009) reserved for Shanghai local students, competition through examinations is still fierce in schools. The pressure from and the trust and confidence in formal summative examinations as a fair gauge persist; particularly in the so-called 3+1+x core subjects (Chinese, Mathematics and English, plus one interdisciplinary, and one other (either politics, history, physics, chemistry, biology or geography) chosen by students themselves). Therefore, it can be safely argued that the 15-year-olds from Shanghai that took the PISA examinations are test takers who have substantial experience in taking formal examinations. Typically, the 15-year-olds in the Shanghai school system could be at the end of their junior secondary education or at the beginning of their senior secondary education phase.

Responses to PISA

In April 2009, 5,115 students from 152 secondary schools representing 112,000 enrolled students of 15-years-old in Shanghai took the PISA assessments (see Table A2.1 of PISA 2009 Report Volume 1: *What Students know and can do: student performance in reading, mathematics and science*). Shanghai is the first region in mainland China to have participated in PISA, and the test results ranked Shanghai firmly as the top performer (Note: 12 provinces participated in PISA, but only the Shanghai results were released). (See *Mervis, Science 10 December 2010: Vol. 330 no. 6010 p. 1461*). As it was Shanghai's debut in PISA assessments, it is probably too early to talk about the real long-term impact of the PISA results on its educational policies. However, analysis of the official documents released by the Shanghai PISA Co-ordinating and Research Centre (SHPISA) and the media coverage in China and around the world (in particular, in the USA) provides some interesting perspectives to

> As a "first" timer and "first" performer in PISA assessments, Shanghai has taken a cautious learning approach towards this significant achievement. The confirmation of lower school-level variance is particularly welcomed as evidence of educational equality.

understand the (potential) PISA effects on educational policies in Shanghai and perhaps other provinces and regions in China.

The response from international media, especially in the USA, e.g. New York Times, The Washington Post, was swift, sharp, and even sometimes hysterical. It was considered a 'Sputnik' moment, a 'Pearl Harbor Day', and a 'massive wake-up call' for American education (Secretary of Education, Arne Duncan). In sharp contrast with the international media, the reaction to Shanghai's performance in PISA assessments within China was strikingly different from the normal Chinese discourse of gold medal winning in the Olympics or successful rocket launching as escalated as a substantive national pride and triumph. At the release of the test results by OECD, the Chinese national and Shanghai local media coverage demonstrated some humble and reflective thinking about this significant achievement.

As a 'first-timer', Shanghai took 'learning' as well as a census approach toward its participation in PISA 2009. As described in the promotional document released by the Shanghai Education Commission, participation in PISA was considered full of high potential and value for the government to promote further reform and development of basic education, inform and guide the whole society to change educational beliefs and philosophy, establish scientific understanding about educational quality, explore the use of internationally acknowledged standards of educational quality and scientific assessment methods to gauge the achievement of primary and secondary school students, with the ultimate aim to meet the demands of educational modernization of the city. In addition, for the participating schools and teachers, PISA was considered as a tool to provide evidence for the schools about their standing in the world. Although this kind of discourse is typically politically motivated and quite common in China, it does signal Shanghai's intention to learn from its own performance and from the international community of educational assessment and quality monitoring. When the results were released by OECD, SHPISA produced a summary of the test results in Chinese on the 8th December 2010. This document reiterated the purposes of the participation: to measure students' achievement in compulsory education and understand their learning ability through international comparative assessment; to improve educational decision making, promote educational equality and further curriculum reform; to learn from advanced educational assessment theories and technology in order to improve the assessment and evaluation system of basic education in Shanghai. The alliteration of '*zi xin*' (self-confidence) and '*zi xing*' (self-reflection) used by Prof. Zhang Minxuan, Deputy Director-General of Shanghai Education Commission and Co-ordinator of SHPISA, when commenting on the significant achievement of the Shanghai students, echoes this 'learning' discourse. By *Zi-xin*, he meant self-confidence in the course that the educational reform in Shanghai has taken. *Zi-xing* refers to self-reflection on the global trends in education, learn from the lessons and experiences of other countries, being aware of the shortcomings, gaps and blind spots that Shanghai education may have, and develop new educational theories and methods. Participation in PISA assessments has also provided useful information for the writing of the *Outline for Medium and Long-term Development and Reform of Education 2010-2020 (Government of China, 2010).*

These two official documents (PISA Shanghai promotional, and executive summary of Shanghai's performance) and Prof Zhang's comments more or less represent and determine the overall tone of Shanghai's official interpretations of the test results for future decision making with regard to two main areas in education: (a) curriculum reform, e.g. addressing gender disparity in reading proficiency, improving students' reading comprehension of non-continuous discourse such as graphs

and tables, developing students' higher order learning skills such as summarization and self-monitoring to reduce student workload in repetitive memorization and drills, and (b) social and political agenda setting to promote balanced educational development to ensure equality in both educational processes and outcomes. Educational equality has been on the top agenda of educational planning at all levels of governments in China, especially for the next decade (2010-2020); it has also become inevitably the cornerstone of Chinese media coverage on Shanghai PISA success. The lower contribution of school-level factors (and to some extent district-level as well) towards students' performance in Shanghai's PISA results and the lower impact of students' socio-economic conditions on their reading performance, compared to the average of the OECD, are politically very useful information to support the argument that educational disparity or equality in Shanghai is better than other OECD countries or regions. In fact, such indicators of educational equality, rather than the top performance *per se*, have been highlighted by Shanghai Education Commission as useful confirmation about its progressive reform in basic education, summarized as 'Shanghai Experience' and 'World Wealth' on the front-page of *Zhongguo Jiaoyu Bao*, the key national newspaper of Chinese Ministry of Education(18th February 2011, http://paper.jyb.cn/zgjyb/html/2011-02/18/node_2.htm). Shanghai's success stories were also made available in a video series called *Strong performers and successful reformers in education* (Pearson Foundation, 2011). *Zhongguo Jiaoyu Bao* devoted four pages a month later (17th March 2011) to discuss PISA reading tests (http://paper.jyb.cn/zgjyb/html/2011-03/17/node_6.htm). Interestingly, several participating schools in Shanghai uploaded these documents to their websites, possibly as teachers' professional development material. This clearly demonstrates the impact of Shanghai's participation in PISA 2009 even at school level.

In addition to official and semi-official publications at government and school levels, there are ongoing discussions by Chinese educational researchers on a range of topics arising from PISA, e.g. evidence-based policy making, the washback of PISA tests on teaching and learning (e.g. whether PISA tests can promote creativity and innovation).

Switzerland Country Profile

Switzerland's students performed close to the OECD average in reading, and above average in science and mathematics, in PISA surveys. Changes over the assessment cycles have not been significant, so this is a static profile.

Educational Structure

Categorisation: **Separated, Group-adapted and Uniform systems in different cantons**

Switzerland is a confederation of 4 linguistic regions and 26 cantons, with a total population of just over 7.5 million. The majority of the population (over 65%) resides in the German-speaking region, followed by the French-speaking region with over 20%, the Italian-speaking region at under 10%, and the tiny Romansh region with under 1%; a relatively high proportion of people, at almost 10%, have a home language that is not one of the four national languages. Education has traditionally been highly decentralized, and there is as yet no federal Minister of Education. Educational decision-making resides with the individual cantons, which provide most of the funding for primary and lower secondary education; responsibility for post-compulsory education is shared with the Confederation. Cantons currently work together through the Swiss Conference of Cantonal Ministers of Education on issues that have national implications, such as mutual recognition of the different cantonal qualifications.

The education system as a whole (involving 26 different cantonal systems) is currently in a state of flux, as a result of several reform initiatives, in part inspired by PISA. There are for the moment nine years of compulsory primary/lower secondary education, from age 6 to 15. The division between primary and lower secondary education has traditionally varied across cantons from 6-3 (the predominant pattern) to 5-4 or 4-5 (see Grossenbacher, 2010 for details), but is now in the process of harmonization towards the 6-3 pattern. There are two, or sometimes three, different types of school in the lower secondary sector, again depending on the canton; they differ in lower and higher requirements. Three structural models exist among the three types of school: streamed, cooperative and integrative. After nine years of compulsory education around two-thirds of students eventually follow a vocational training programme in the upper secondary school while about 25% enter a baccalaureate school.

Most students of PISA-taking age are in their final year of lower secondary education. At this level, over 92% of students are enrolled in state schools, just under 3% in government-dependent private schools and 5% in independent private schools. In state schools the average class size is 19 students, and the student-teacher ratio is 12:1. Public expenditure on education as a percentage of GDP is 5.2%.

The education system can be classified as predominately separated (more than half the cantons), but both of the other models also exist, the choice depending on the canton.

Assessment Structures and Formats

Teachers in Switzerland engage in a combination of formative and summative assessment, generally using teacher-made tests. On transition from the primary school to lower secondary education as well as from lower secondary education to an academic or vocational upper secondary school, academic achievement is an important, but not over-riding, selection criterion. Decisions for the first transition (primary to lower secondary) are made on the basis of a combination of different types of information, including continuous teacher assessment, informal in-school test results, teachers' judgements about relevant aspects of behaviour and the kind of future education that would best suit the child. All relevant information is discussed with the student and the student's family. At the transition from lower to upper secondary education, test-based assessment takes on a greater role in the selection process in the academic secondary school, the ultimate goal being achievement of the high-status examination-based school leaving certificate taken at the end of upper secondary education (the Maturité/Matura/Maturità). Assessment in dual-system vocational education is partly through test-based teacher assessment of 'general subjects' studied in the upper secondary school and partly through employers' workplace-based judgements of proficiency.

A French-style Baccalaureate is the principal school-leaving qualification in Switzerland for academic schools. This is known as the Matura in the German-speaking part of the country, the Maturité in the French-speaking part, and the Maturità in the Italian-speaking part, and is taken at the end of upper secondary education. There are a variety of different options, including general academic (with specialisms in science, social science and arts). Assessment is through classroom tests in nine different subjects, seven of which are compulsory and two are optional choices, plus continuous assessment marks in the student's school report. From the age of 15 students also work on their baccalaureate essay, and this is assessed as one of the nine subjects.

Responses to PISA

After Germany, Switzerland 'had the second largest national press coverage of all countries when the PISA 2000 results were released,' (Ramseier, 2008, p.36). One possible reason for this high level of interest was due to Switzerland being a federated system without any national

- There has been huge press interest in the published findings, especially the lacklustre findings from PISA 2000.
- Within political circles the impact of PISA, and TIMSS to some extent, was to accelerate a long-awaited process of harmonisation in educational structures, practices and curriculum in this federated country, with its history of cantonal autonomy in education and consequent variety of provision.
- The culmination of this process is fully documented in the first-ever *national* report on the Swiss educational system(s), which outlines details of a strategic plan for reform and for PISA-style system evaluation and monitoring.

assessment programme. This was the first time that the country was able to benefit from empirical information about its national performance at the end of obligatory schooling. Boosted samples (see Introduction) gave the country the further opportunity to compare the performance of students in the different cantons within the major linguistic regions.

While not uniquely triggered by the PISA experience, it can be said that participation in PISA, and indeed in the IEA surveys, served to accelerate a number of initiatives that had been brewing for decades. These were aimed at:

a) harmonising educational structures, curricula and standards across the country, or at least within linguistic regions

b) providing information about the education system in its entirety, and

c) monitoring the education system regionally and nationally into the future.

Action towards harmonisation began with the Harmos concordat of 2006, which established age 4 as the national starting age for obligatory education, and 11 years as the duration of that education. Harmos also proposed national educational standards and the creation of common curricula within each linguistic region (Delamadeleine, 2008). A common curriculum is even now being introduced into the French-speaking region, while the German-speaking cantons will follow suit in 2014 with their own version. There were debates as well about the need for a programme of sample-based system monitoring to be set up, following the PISA model (Behrens, 2008; Ramseier, 2008). Such a system is now in the design stage.

Details of the decisions taken towards harmonising education, improving quality and efficiency, and monitoring the system over time, are contained in a 300-page document published in 2010 (*L'éducation en Suisse rapport 2010*; SKBF/CSRE, 2010) – an English translation is available (SKBF/CSRE, 2011) . The report covers every level of education, with each level evaluated in terms of its effectiveness, efficiency, and equity. Both national and international attainment results are included to support decisions. See also Fuentes (2011) for an overview of reforms.

References and Sources

Behrens, M. (2008) Quelles évaluations dans le cadre d'HarmoS et de la Convention scolaire romande ? Bulletin CIIP, no.22.

Bourny, G., Braxmeyer, N., Dupé, C., Rémond, M., Robin, I. & Rocher, T. (2002) *Les compétences des élèves français à l'épreuve d'une évaluation internationale. Premiers résultats de l'enquête PISA 2000*. Paris: Ministry of Youth, Education and Research.

Bradshaw, J., Ager, R., Burge, B. and Wheater, R. (2010) Programme for International Student Assessment 2009: achievement of 15-year-olds in England. DFE-RB069. December. ISBN 978-1-906792-75-6.

Canadian Statistics Education Council, (2007) *Education Indicators in Canada: Report of the Pan-Canadian Education Indicators Program*, Toronto: Canadian Statistics Education Council. Available at: http://www.statcan.gc.ca/cgi-bin/af-fdr.cgi?l=eng&loc=/pub/81-582-x/81-582-x2007001-eng.pdf [Accessed April 20, 2011].

Delamadeleine, Y. (2008) Le PER, Plan d'études romand. *Bulletin CIIP*, no.22.

Department for Children, Schools and Families, (2005) *14-19 Education and Skills White Paper*. 23 Feburary.https://www.education.gov.uk/publications/standard/publicationdetail/page1/Cm%206476

Educational Quality and Accountability Office. (EQAO, 2010) *Programme for International Student Assessment (PISA), 2009: Highlights of Ontario results*. Available at: http://www.eqao.com/pdf_e/10/2009_PISA_Highlights_en.pdf [Accessed April 20, 2011].

Evans, R. (2010) Is the Canadian model right for UK schools? *The Guardian*, January 4, 2011. Available at: http://www.guardian.co.uk/education/2011/jan/04/education-policy-canadian-model
[Accessed April 28, 2011].

Fuentes, A. (2011) *Raising Education Outcomes in Switzerland*, OECD Economics Department Working Papers, no. 838, Paris: OECD Publishing.

Gill, B., Dunn, M & Goddard, E. (2002) Student Achievement in England: Results in Reading, Mathematical and Scientific Literacy Among 15-Year-olds from OECD PISA 2000 Study. London: The Stationery Office.

Goldstein, H. (2004a) International comparison of student attainment: some issues arising from the PISA study. *Assessment in Education*, 11, 3, pp319-330.

Goldstein, H. (2004b) Education for all: the globalization of learning targets. *Comparative Education*, 40, 1, 7-14.

Gove, M. (2010) Pisa slip should put a rocket under our world-class ambitions and drive us to win the education space race. *Times Education Supplement*, 17 December. http://www.tes.co.uk/article.aspx?storycode=6066185

Gove, M. (2011) Education for Economic Success. Education World Forum conference presentation, 11 January. http://www.education.gov.uk/inthenews/speeches/a0072274/michael-gove-to-the-education-world-forum

Government of China (2010) *Outline of China's National Plan for Medium and Long-term Education Reform and Development (2010-2020*), State Council of the Peoples' Republic of China, Beijing, 29 July.

Grek, S. (2009) Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy*, 24, 1, pp. 23-37.

Grossenbacher, S. (2010) Ecole obligatoire: harmonization, integration, développement de la qualité. *Educateur,* 04.10 : pp16-17.

Hilton, M. (2006) 'Measuring standards in primary English: issues of validity and accountability with respect to PIRLS and National Curriculum test scores', *British Educational Research Journal*, 32, pp. 817-37.

Hopmann, S.T., Brinek, G. & Retzl, M. (2007). PISA zufolge PISA – PISA according to PISA. Berlin and Vienna: LIT Verlag.

Hutchison, D. & Schagen, I. (2007) Comparison between PISA and TIMSS – are we the man with two watches? In T. Loveless (ed.), *Lessons learned: what international assessments tell us about math achievement?* Washington DC: Brookings Institute Press.

Klinger, D. A., DeLuca, C. and Miller, T. (2008) The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76. Available at: https://www.umanitoba.ca/publications/cjeap/articles/klinger.html [Accessed April 20, 2011].

Knighton, T., Brochu, P. and Gluszynski, T. (2010) *Measuring up: Canadian results of the OECD PISA study*. Ottawa, ON: Human Resources and Skill Development Canada, Council of Ministers of Education, Canada, and Statistics Canada. Available at: http://www.statcan.gc.ca/pub/81-590-x/81-590-x2004001-eng.pdf [Accessed April 20, 2011].

Kurial, R. (2005) *Excellence in education: A challenge for Prince Edward Island*. Charlottetown, PE: PEI task force on student achievement. Available at: http://www.edu.pe.ca/cslf/Final_report.pdf [Accessed April 20, 2011].

Mervis, J. (2010) *Shanghai Students Lead Global Results on PISA*, Science 10, December 2010: Vol. 330 no. 6010 p. 1461. http://www.sciencemag.org/content/330/6010/1461.short

Micklewright, J. and Schnepf, S.V. (2006) Response bias in England in PISA 2000 and 2003. Research Report 771. ISBN 1 84478 764 8. https://www.education.gov.uk/publications/standard/publicationDetail/Page1/RR771

Mortimore, P. (2008) A league to worry us all. *The Guardian*, 8 January. http://www.guardian.co.uk/education/2008/jan/08/schools.uk1

Mons, N. (2007) *Les nouvelles politiques éducatives*. Presses Universitaires de France

Mons, N. & Pons, X., with van Zanten, A. & Pouille, J. (2010) *La réception de Pisa en France. Connaissances et régulation du système éducatif*. European Union. [Available in English translation]

Mullis, I.V.S., Martin, M.O., Kennedy, A.M. and Foy, P. (2007) *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*. Chestnut Hill, MA: Boston College, TIMSS and PIRLS International Study Center.

Mullis, I.V.S., Martin, M.O. & Foy, P. with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A. and Galia, J. (2008a) *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D. & Stanco, G.M. (Eds). (2008b) *TIMSS 2007 Encyclopedia: a Guide to Mathematics and Science Education Around the World*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Nardi, E. (2008) Cultural biases: a non-Anglophone perspective. *Assessment in Education*, 15, pp. 259-267.

New Brunswick Department of Education. (2009) *An accountability report on the targets of 'When kids come first,' 2008*. Fredericton, NB: Department of Education. Available at: http://www.gnb.ca/0000/publications/comm/report_targets_wkcf_08.pdf [Accessed April 20, 2011].

OECD, (2001) *Knowledge and Skills for Life. First results from PISA*. Paris: OECD.

OECD, (2004) *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD Publishing.

OECD, (2007) *PISA 2006: Science Competencies for Tomorrow's World*. Paris: OECD Publishing.

OECD, (2009a) *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD Publishing.

OECD, (2009b) *Creative Effective Teaching and Learning Environments. First results from TALIS,* Paris, OECD.

OECD, (2010a) *Education at a Glance 2010. OECD Indicators*. Paris: OECD.

OECD, (2010b) *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*. Paris: OECD Publishing.

OECD, (2010c) *PISA 2009 Results: Overcoming Social Background: Equity in Learning, Opportunities and Outcomes (Volume II)*. Paris: OECD Publishing.

OECD, (2010d) *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices (Volume III).* Paris: OECD Publishing.

OECD, (2010e) *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV).* Paris: OECD Publishing.

OECD, (2010f) *Learning Trends: Changes in Student Performance since 2000 (Volume V)*, Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264091580-en

OECD, (2011a) Building a high quality teaching profession. Lessons from around the world. http://www.oecd.org/topic/0,3699,en_2649_37455_1_1_1_1_37455,00.html

OECD, (2011b) OECD *Country Background Report on Norway,* (Draft version) Paris, OECD.

OECD (2011c). *PISA 2009 Technical Report (preliminary version)*, Online at http://www.oecd.org/document/19/0,3746,en_2649_35845621_48577747_1_1_1_1,00.html.

Parkin, A. (2010) The Canadian schools pass international test. *Toronto Star*, 28 April 2011. Available at: http://www.thestar.com/article/904868 [Accessed April 28, 2011].

Pearson Foundation (2011) Series: Strong Performers and Successful Reformers in Education, Shanghai, China. online video http://www.pearsonfoundation.org/oecd/china.html

Prais, S.J. (2003) Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education,* 29, pp. 139-163.

Prais, S.J. (2007) Two recent (2003) international surveys of schooling attainments in mathematics: England's problems. *Oxford Review of Education*, 33, pp.33-46.

Ramseier, E. (2008) Validation of competence models for developing education standards: Methodological choices and their consequences. *Mesure et Evaluation en Education*, 32, pp. 35-54

Ruddock, G., Clausen-May, T., Purple, C. & Ager, R. (2006) *Validation Study of the PISA 2000, PISA 2003 and TIMSS-2003 International Studies of Pupil Attainment (DfES Research Report 772).* Slough: NFER.

Schuetze, H. G. and Sweet, R. (2003) Integrating school and workplace learning in Canada: an introduction to alternation education concepts and issues. In H. G. Schuetze & R. Sweet (eds.) *Integrating school and workplace learning in Canada: principles and practices of alternation education and training*. Montreal, QC, McGill-Queen's University Press, pp. 3–22.

Sjøberg, S. (2007) PISA and 'Real Life Challenges': Mission Impossible?, http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007 , Accessed 22 April 2011.

Simpson, J. (2011) The good, the bad and the ugly education facts. *The Globe and Mail*, January 12, 2011. Available at: http://www.socialsciences.uottawa.ca/irpe-epri/eng/documents/Globe_Jeffrey_Simpson_12JAN2011.pdf [Accessed April 28, 2011].

SKBF/CSRE, (2010) Swiss Education Report 2010.. Bern: SKBF/CSRE.

Statistics Canada and Council of Ministers of Education, Canada (CMEC 2010). *Education indicators in Canada: An international perspective*. Ottawa, ON, Canadian Educational Statistics Council. Available at: http://www.cmec.ca/Publications/Lists/Publications/Attachments/253/indicators-perspective-int-spet2010.pdf [Accessed April 20, 2011].

Stewart, W. (2011) Forget 2:2 degrees, Pisa rankings guru urges Gove. *Times Education Supplement*, 15 April, p18.

Thomas, S.M. & Goldstein, H. (2008) Educational comparative studies of achievement – re-examining the issues and impact. *Assessment in Education*, 15, pp 211-214.

von Törne, V. L. (2008) Kulturrevolution im Klassenzimmer. Der Tagesspiegel, August 24, 2008. Available at: http://www.tagesspiegel.de/meinung/kommentare/kulturrevolution-im-klassenzimmer/1308246.html [Accessed April 28, 2011].

Välijärvi, J., P. Linnakylä, P. Kupari, P. Reinikainen, and I. Arffman (2002) The Finnish success in PISA – And some reasons behind it. Jyväskylä: Institute for Educational Research, University of Jyväskylä.

Volante, L. and Ben Jaafar, S. (2008) Educational assessment in Canada. *Assessment in Education: principles, policy, and practice* 15 (2), pp. 201–210.

Wagemaker, H. (2008) Choices and trade-offs. A reply to McGaw. *Assessment in Education*, 15, pp. 267-279.

Wilson, D. N. and Lam, T. C. M. (2004) 'Canada' In H. Döbert, E. Klieme & W. Sroka (eds.) *Conditions of school performance in seven countries: a quest for understanding the international variation of PISA results*. Münster, Germany, Waxmann Verlag, pp. 15–61.