International Educational Indicators and Assessments: Issues for Teachers

Anna K. Chmielewski, Karen Mundy, and Joseph P. Farrell

Pre-print version—November 8, 2016

Book available for purchase at: https://www.canadianscholars.ca/books/comparative-and-international-education-2nd-edition

Full citation:

Chmielewski, Anna K., Karen Mundy, and Joseph P. Farrell. 2017. "International Indicators and Assessments: Issues for Teachers." Pp. 361-391 (Ch. 13) in *Comparative and International Education: Issues for Teachers, 2nd Edition*, edited by Kathy Bickmore, Ruth Hayhoe, Carly Manion, Karen Mundy, and Robyn Read. Toronto: Canadian Scholars' Press Inc.

International Educational Indicators and Assessments: Issues for Teachers

Anna K. Chmielewski, Karen Mundy, and Joseph P. Farrell

Introduction

Statistics about educational achievement and other social issues increasingly influence and drive debates over public policy. Both political actors and media commentators make claims based on such statistics: we have all read headlines about the growth of income disparities and other forms of economic inequality, the growth of secondary dropout rates, or the poor literacy skills of youth. Educational statistics are also regularly used in comparing nations (and in federal nations such as Canada and the United States, to compare provinces/states). You may have noticed media reports on how a country ranks internationally, describing, for example, Canadian or American student achievement in math, science, or literacy in relation to students in other countries around the world. The tests on which these rankings are based are different from those used at the district and the school level to assess educational performance. Yet both types of tests may be thought of as part of the same global movement to gather more educational data for comparison.

Why are statistics increasingly perceived as useful? At least part of the reason stems from their ability to place disparate students, schools, provinces/states, or countries onto a single external scale for comparison. Statistics can alert teachers and local administrators to blind spots, such as areas of learning that may be neglected or groups who may be disadvantaged. But placing all students on a common scale inherently means simplifying and reducing complex information. Statistics cannot (and are not intended to) capture the nuanced, subjective knowledge of practitioners. Furthermore, it is important to remember that while statistics strive

for objectivity, they are ultimately collected and interpreted by humans. Although there have been enormous advances in survey technology in recent decades, flaws in methods of data collection and survey design can still limit accuracy and representativeness. More importantly, misconceptions or even political agendas can bias interpretation, reporting and decision-making based on educational statistics. Thus, it is crucial to read statistics with a critical eye toward "fine print" describing how the data were collected and exactly what they measure.

As citizens, students, and professional educators, there is a pressing need for us to better understand these numbers. This chapter aims to demystify some widely used international educational data, which can and do have an important influence on the careers and day-to-day practices of educators. We begin with a history of various efforts to develop statistical indicators for educational systems. We look first at international indicators that provide cross-national (or cross-provincial/state) comparisons of various core aspects of an educational system. Then we turn to international large-scale assessments (ILSAs), that is, cross-national achievement tests. ILSAs are sometimes classified as part of the set of international indicators, but the issues involved in their creation and use are sufficiently different from other indicators that they need separate treatment. We also look at how indicators and ILSAs can be combined to provide a glance or snapshot of different aspects of a nation's educational system, highlighting in particular what a comparative snapshot can tell us about Canada. Throughout this chapter, we emphasize the importance of looking critically at cross-national data on education. As a source of externally standardized information, statistics are an invaluable resource that should not be ignored by educational policy-makers and practitioners. But neither should statistics be our only source of evidence for policy-making.

International Indicators

Indicators are statistics used to measure and monitor systems over time and to compare across jurisdictions, such as provinces, states, or countries. International educational indicators have been with us for quite a long time. The first formal intergovernmental efforts to assemble systematic international comparisons of educational systems was undertaken by the International Bureau of Education (IBE) in the early 1930s. Beginning in 1933, the IBE collected basic data about the structure of its members' educational systems, as well as information about specific policies or issues, which was published in an annual *Education Yearbook*.[1] After the Second World War and the formation of the United Nations, UNESCO (the United Nations Educational, Scientific, and Cultural Organization) assumed responsibility for such cross-national data collection and reporting. UNESCO's first questionnaire-based survey of education received responses from 57 member states in 1950. This seems a small number now, but at the time covered almost all of the independent nation-states in the world. The UNESCO survey collected data on school enrolments by level, public expenditure in education, literacy, and a variety of other features of the educational system. The resulting information was published in a *Statistical Yearbook* (which continues to be published under the title *Global Education Digest* (see UNESCO Institute of Statistics, 2012).

To comparative educators in the 1960s, it quickly became apparent that the indicators being reported by UNESCO were, for purposes of comparison, often quite misleading. Beyond questions of accuracy, it was not clear what was actually being counted in any given nation compared to the same statistical indicator in some other nation, since the educational systems of various nations were actually quite different. For example, how would one compare data on various aspects of primary schooling, when some systems end primary schooling after five years

and other systems go up through Grade 8? This concern eventually led to the development in the early 1970s of the International Standard Classification of Education (ISCED), illustrated in Table 13.1, which provided standards for all nations as to what should be counted under which indicator. This classification was updated in 1997 and again in 2011 to incorporate early childhood education for children under three years old as well as more fine-grained categories of higher education ("short-cycle" or community college, bachelor's, master's, and doctoral degrees).

Table 13.1

Original International Standard Classification of Education (ISCED)

Level	Age Range	Stage	Examples
4	22–25	6	• Postgraduate study
3	21–22	5	• Professional schools • Higher stage of university study • Teacher training
	18–19	4	• Advanced technical schools • Lower stage of university study • Teacher training
2	14–15	3	• Full- and part-time vocational schools • Upper section of high schools • Grammar schools • Gymnasiums • Teacher training
	10–11	2	Upper section of elementary schools • Lower section of high schools • Grammar schools • Gymnasiums
1	5–7	1	• First six years of primary school

Compulsory School Begins	0	• Nursery and kindergarten

Note: The stages are illustrated by typical examples; ages stated are also illustrative.
Source: Adapted from Holmes and Robinsohn (1963, p. 57); see also UNESCO, 1997a; 2011.

UNESCO was intended to be the main gatherer of international educational statistics, and in many respects still is so. However, as budgetary crises and leadership problems developed within the United Nations in the 1980s, the agency's ability to develop and improve such indices declined (Puryear, 1995). Other international agencies began developing their own indicators, often using UNESCO indicators, but adding other sources. For example, UNICEF (United Nations Children's Fund) began publishing an annual report titled *The State of the World's Children*, which includes a wide variety of indicators regarding children, including educational data. The World Bank also publishes an annual *World Development Report*, which includes tables listing a wide variety of economic, social, and educational data from most nations in the world. UNESCO began to regain some of its prominence in this area with the establishment in 1999 of the UNESCO Institute for Statistics (UIS), located at the University of Montreal, whose task is to gather quality statistical information from and for member states, and to report on the global situation of education.

Most education-related indicators are essentially head counts. These include total enrolment by level of schooling (or in some cases grade level), retention or dropout rates by level, enrolment ratios by level (the number of students enrolled compared to the number in the population who are age-eligible for that level), number of teachers and teacher/student ratios, government expenditures on education, teacher/faculty average salaries, adult literacy rates, and so forth. They are primarily derived from the administrative information that Ministries or Departments of Education routinely have to collect and assemble for their own administrative

and management activities and obligations, or from censuses that usually occur once every decade.

Several examples of the use of such statistical indicators for comparative purposes can be found in this book. In Chapter Three, for example, they are used to describe the general status and condition of the three nations, Bangladesh, Colombia, and Mexico, and to compare them to Canada. There, in Table 3.2, we find Gross National Income (GNI) per capita, which is a rough measure of the amount of wealth available per person in the nation; a measure of income distribution, which roughly measures the percentage of that wealth available to the poorest 20 percent of that population; and the adult literacy rate. In the accompanying text there is an indication of how those wealth and distribution indicators translate into the actual value, in current international dollars, of the annual income available to the poorest 20 percent of the country's population, per capita. These are approximate figures, with no claim to precise accuracy, but they do provide a general snapshot that can help the readers locate these nations quickly in their own mental maps. Similarly, statistics are used in Chapter Eleven to compare basic education enrolment rates in Tanzania and Kenya and in Chapter Seven to demonstrate gender disparities in education.

Until the 1990s, international indicators have rarely incorporated detailed layers of research (such as on students' family backgrounds, socio-economic status, racial/ethnic identity, family private expenditures on schooling, or the quality of teaching and learning), which would provide context for the indicators surveyed. They have also been limited by the data that governments report. Because governments self-report, some of the international statistical series are of questionable or suspicious reliability and accuracy. One part of the problem is that the quality of the information provided to international agencies depends on the resources available

to governments to collect such information. Richer nations have the resources available to collect and analyze reasonably accurate information on all sorts of government concerns—but collecting such information is expensive. Thus statistics from poorer nations are often simply the best guess of government officials. It is not uncommon, for example, to find that Ministry of Education officials in poorer nations have only the vaguest idea of how many students and teachers are in their formal education system. Elaborate procedures and requirements for data collection are in place on paper but the resources to accurately gather this data are not available. Moreover, governments often deliberately misreport, for domestic and/or international political reasons, such as to exaggerate their own accomplishments or minimize the accomplishments of a previous regime.

Beyond these problems, it is difficult to know what some often-cited international indicators actually refer to, and how they are measured. Adult literacy rates are a good example of this. At one level it seems clear what literacy refers to: the ability to read text, and in some cases to write as well. Literacy statistics, however, are much more complicated than that. Scholars of literacy often distinguish among levels or types of literacy, ranging from basic or functional literacy (usually thought of as the ability to read with understanding fairly simple texts, such as local newspapers or instructions for medicines or farming/gardening chemicals), to much more complex forms, such as the ability to read complicated texts—for example, the plays of Shakespeare. We also now speak about new kinds of literacy, such as computer or mathematical literacy. So it is important to understand what level and type of literacy is being referred to.

Furthermore, it is generally difficult and expensive to actually test people's level of literacy, especially for large population groups. So proxies are used. For example, in many

international statistical series, literacy is taken as the proportion of the adult population who have completed primary school, on the assumption that it normally takes at least five or six years of primary schooling to become literate. In other cases, literacy rates are based on self-reporting from censuses, in which people often exaggerate their level of schooling and/or literacy (Farrell, 2007). For instance, Latin America is generally considered to be one of the most- schooled and literate regions of the developing world, with primary enrolment ratios for the most part well over 90 percent. However, a study done early in the current millennium, covering a large sample of adults from the region, found that of the 63 percent who reported completing primary schooling and being literate, only about 50 percent could actually read with understanding a short paragraph taken from the front page of a local popular newspaper (Schiefelbein, 2006).

In recent years, advances in information technology and the growth in funding for the collection of international data has led to an ever-wider range of statistical information, and to the development of more robust efforts to clarify relationships among statistical indicators. Detailed information on attitudes and lived experiences is often collected alongside administrative data through methods such as household surveys and surveys of students and teachers. As we shall see in the next section, in education such efforts have led to a sometimes bewildering proliferation of cross-national comparative data, in forms that critics argue contribute to both greater homogenization of educational systems, and greater surveillance and control. On the other hand, such data is increasingly sensitive to questions of inequality and may allow us to see whether formally agreed universal entitlements, such as the right to education, are in fact realized in the distribution of educational opportunities around the world. UNICEF's annual *Report Card* on child poverty and well-being in rich countries (UNICEF, 2013), and

UNESCO's *Global Education Monitoring Report* (previously the *Education for All Global Monitoring Report*) (UNESCO, 2015) exemplify this trend.

The Birth of International Large-Scale Assessments

The term *international large-scale assessments* (ILSAs) is commonly used to refer to tests of educational achievement carried out in more than one nation using the same tests and testing methodology. The first ILSA was the First International Mathematics Study (FIMS), which was conducted in 1964, under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). The IEA had been founded in 1958, growing out of a meeting at the UNESCO Institute for Education of an international group of educational psychologists, and curricular and measurement specialists from a variety of Western countries. The founding chair of the IEA and the head of the FIMS study was eminent Swedish professor of educational psychology Torsten Husén. Husén and the other founders of the IEA viewed the world as a "natural educational laboratory," encompassing far more variation than could be observed in a single national setting, and thus prime for educational research (IEA, n.d.). FIMS consisted of multiple-choice and fill-in-the-blank math problems translated into eight different languages and administered in 12 different countries. The countries were all relatively high income and located primarily in Europe, plus Australia, Israel, Japan and the United States. (Canada did not participate in FIMS.)

In each country, tests were administered to randomly-chosen, anonymized samples of about four thousand Grade 8 students and four thousand students in the pre-university stream of the final year of high school. In addition, students, teachers and principals completed questionnaires to provide context on family and educational background, attitudes toward

learning, resources and teaching practices. The scope and ambition of FIMS was ground-breaking for its time. After the success of FIMS, the IEA went on to undertake studies in six more academic subjects in the 1970s (science, reading comprehension, literature, English and French as foreign languages, and civic education). In the 1980s and 1990s, the IEA conducted follow-up studies in many of these same subjects in order to expand on earlier findings and to study changes in education over time. With each successive study, new countries joined, including some Canadian provinces in 1980, dozens of other industrialized countries, and a small number of middle-income and developing countries, including Chile, Iran, Nigeria, Swaziland, Thailand, and Zimbabwe.

Aside from refining techniques for educational measurement, early ILSAs produced a number of interesting findings. One important issue was how to compare achievement in the final year of secondary school across countries with very different "retentivity" rates, meaning the share of students who had not left school or entered vocational training by this grade. While in comprehensive systems like Sweden's, nearly all youth were still in school and following a general stream by the end of secondary school, in selective systems like Germany's, vocational stream students graduated after grade nine or ten, meaning that only an elite few remained until the end of secondary school. Thus, a more appropriate comparison was at Grade 8, when virtually all students were still enrolled in school, at least in the wealthy participating countries (Husén, 1967b).

A second finding from the early ILSAs is perhaps the most important and well-known: the concept of "opportunity to learn" (OTL). OTL originated as a way to explain and validate cross-national differences in achievement based on the fact that students' content exposure was not equally well aligned to the test in every country. Even though the tests were designed to

represent an “international consensus” curriculum of agreed-upon topics, it appeared that teachers implemented curriculum differently across countries. OTL was conceptualized as the second of three levels at which the curriculum operates: (1) the “intended curriculum,” which was the official curriculum of the country (or province/state), (2) the “implemented curriculum,” or OTL, meaning the content that teachers actually taught in their classrooms and (3) the “attained curriculum”, or the content that students learned, as evidenced by their performance on the test. By collecting extensive curricular information from government officials and from teachers, the scholars of the early ILSAs found large discrepancies between the intended and the implemented curriculum in some countries. Further, they found that the implemented curriculum, or OTL, went a long way toward explaining cross-national differences in achievement. Finally, ILSA results drew attention to the unequal distribution of OTL within some countries, particularly those practicing curricular tracking and streaming between or within schools (McDonnell, 1995).

Throughout the 1980s and 1990s, the IEA implemented many other innovative research ideas. For example, the Second International Mathematics Study (SIMS 1980) included longitudinal (pre/post-test) designs in some countries, in order to observe how much students learn in one year. Other IEA studies went beyond collecting data only from tests and surveys. The Civic Education Study (CIVED 1999) included impressive in-depth qualitative case studies of civic education in 24 countries (Torney-Purta, Schwille, & Amadeo, 1999). The Third International Mathematics and Science Study (TIMSS 1995 and 1999) collected classroom video data in eight countries to allow for even greater analysis of instructional practices. These videos are now freely available online, meaning teachers themselves can access and review them. (A link is provided in the audio-visual resources at the end of this chapter.) Reflecting the attitude of

seeing the world as an educational laboratory, an article based on the TIMSS video study concluded, “The opportunities to see the familiar in new light might offer many opportunities for teachers to rethink the taken-for-granted practices and see them as choices rather than inevitabilities” (Givven et al., 2005).

The founders of the IEA were academic researchers, primarily curricular and measurement specialists. They intended for ILSAs to contribute to a global body of knowledge on teaching and learning, but they did not promote educational policy changes on the basis of their findings (Pizmony-Levy, 2014). Most particularly—and perhaps most naively—they were opposed to using the test results to rank countries by performance. As Torsten Husén argued in the FIMS (1964) report, “The IEA study was not designed to compare countries; needless to say, it is not to be considered as an ‘international contest’” (Husén, 1967b, p. 288). Relatedly, Husén emphasized that ILSAs could not be used to make causal arguments and claims about why certain countries achieved the results they did, or which policy changes might improve a country’s results (Husén, 1967a, p. 31). ILSAs merely provided a snapshot of a country’s achievement and instructional practices at a single point in time; as the popular mantra goes, “correlation does not imply causation.”

However, the basic design and intention of international assessment studies did not stop their results from being used both as direct evidence for educational policy decisions and as fodder for competition among nations. The early ILSA reports listed countries’ average scores in alphabetical order rather than ranked from highest to lowest, in a table buried in the middle of the reports after chapters of background material—or avoided publishing country averages altogether. Yet it was easy for politicians and journalists to reconstruct the rankings themselves based on the information in the reports. What were those rankings? In the early math and science

assessments, the consistent top performer was Japan. The lowest performers tended to be the developing countries. Among the higher-income countries, the lowest performers were often the US, Sweden and Finland. Canada's performance was generally above average. For national policy-makers, it did not go unnoticed that top-scoring Japan was also experiencing rapid economic growth in the 1980s, and this success was assumed to be the product of a superior school system. The simplistic use of rankings linked to economic competitiveness is best illustrated by the 1983 US Department of Education report *A Nation at Risk,* which stated that across all ILSAs conducted to date, "American students were never first or second and, in comparison with other industrialized nations, were last seven times" (National Commission on Excellence in Education , 1983, p. 8) and equated this low performance to "unilateral educational disarmament" (ibid., p. 5).

The IEA's explanations for Japan's high scores were generally curriculum-focused. They noted that Japanese students experienced a high level of OTL in math. The TIMSS video study revealed that Japanese math teachers devoted more time to introducing new content and less time to reviewing old lessons than teachers in other countries (Givven et al., 2005). Yet the US standards-based reform movement that grew out of *A Nation at Risk* did not directly target improvements to OTL and instruction but instead aimed to increase standards indirectly through accountability, high-stakes testing, and decentralization of management. In this early example of educational reform justified though ILSA results, the wealth of information from the world's "educational laboratory" was reduced to simple rankings and used to support a pre-existing reform agenda. Unfortunately, the policy effects of ILSAs in later decades and in other countries have sometimes followed a similar pattern.

Globalization, Large Scale Assessment, and the Politics of League Tables

In the 1990s, as the Cold War faded from view, governments around the world focused their policies even more intensely on the challenges of globalization and international economic competition. Human capital—particularly in terms of skills in science, math, and literacy—was increasingly seen as central to national efforts to maintain economic advantage within the world economy (see Chapters One and Eleven for further discussions on human capital). The information available to measure those skills has dramatically proliferated. After 1999, the IEA changed the name of TIMSS from "Third" to "Trends" in International Mathematics and Science Study, and began conducting the study every four years at Grades 4 and 8. The IEA also runs assessments of reading skills every five years, civic education every seven years, and computer literacy every five years, and is developing a test of kindergarten early literacy (see Table 13.2 for a full listing of current ILSAs). These other subject tests have lower participation rates and receive less media attention than TIMSS, most likely because in the policy discourse, they are less linked to economic growth than are math and science.

Around 50 countries now participate in each cycle of TIMSS—yet the composition of this set of countries has changed over time. The number of Western countries participating in TIMSS has declined markedly since 1995, while the number of Middle Eastern and African countries has increased (and the number of East and Southeast Asian countries has remained high). Germany, France, most Canadian provinces, and many other countries have left TIMSS. This may seem surprising given the growing worldwide policy focus on human capital and economic growth, but the likely explanation is the advent of a new ILSA: the Programme for International Student Assessment (PISA).

PISA is run by the Organization for Economic Co-operation and Development (OECD), an intergovernmental economic and policy organization founded in 1961 that includes most of the world's wealthiest countries. The OECD is based in Paris, France and describes itself as committed to the market economy and democracy and working to promote economic progress, world trade and policy best practices (OECD, n.d.). One of the main functions of the OECD has been collecting and publishing its member countries' economic indicators, such as gross domestic product. The OECD also collected educational indicators, such as numbers of students enrolled in primary, secondary, and post-secondary education and government expenditures on education, releasing these in an annual publication called *Education at a Glance*. But starting in the mid-1990s, the OECD expanded its mission to measuring the skills necessary for a productive national workforce. The OECD hired a German statistician from the IEA named Andreas Schleicher to help develop a new assessment in which all OECD countries would regularly participate. PISA was conducted for the first time in 2000 and is repeated every three years. In addition to all 34 OECD member countries, the number of non-member "partner" (mostly middle-income) countries participating in PISA has quickly grown and surpassed OECD countries, making PISA the largest ILSA ever conducted. The ability to compare academic performance to all of one's major economic competitors (with the notable exceptions of China and India) is likely one reason that many wealthy countries choose to participate in PISA rather than TIMSS. Another reason is that PISA directly caters to the desire of policy-makers to draw policy lessons from ILSAs. Unlike the IEA, which historically was an organization of academics pursuing research questions that interested them, the OECD has always been an organization of governments seeking policy solutions. In recent years, the IEA has also begun to see heavier government involvement and more policy focus (Pizmony-Levy, 2014).

PISA shares many similar technical aspects with the IEA assessments, including randomly, anonymously sampling 4,000 to 4,500 students in 150 schools per country, administering student and principal contextual questionnaires, and using similar statistical methods for calculating test scores. Both PISA and the IEA assessments are nationally representative samples that include both public and private school students (Martin & Mullis, 2013; OECD, 2014a). But the differences between PISA and TIMSS illustrate the OECD's focus on future workforce skills. First, while TIMSS is a curriculum-based test, PISA tests mathematical, scientific and reading "literacy," with a focus on problem-solving and real-world application. While TIMSS test questions are mostly multiple-choice and fill-in-the-blank, PISA has fewer multiple-choice and more open-ended questions (ibid.). (The types of questions used in PISA and TIMSS can be compared by accessing the publicly released items for each test, which are listed among the resources at the end of this chapter.) Second, PISA's contextual surveys are less focused on curriculum. Its school principal surveys ask more questions about management practices (such as decentralization of decision-making, accountability, and school choice), and PISA does not collect teacher surveys at all. Third, rather than sampling Grade 8 or Grade 4 students, PISA samples 15-year-old students, regardless of which grade they are in (OECD, 2014a). Age 15 was chosen to correspond to the end of compulsory schooling in many member countries (at least in the late 1990s when the test was being developed; compulsory schooling has been lengthened in many countries since then) (Baird et al., 2011). Thus, PISA was intended to measure the "yield" of national educational systems for the broadest cross-section of students possible—in effect it purports to measure the economic competitiveness and skills of a nation's future labour force.

Despite its shorter history, PISA appears to have gained more international recognition than the IEA studies among policy-makers, the media, and the public. This may be due to the appeal of the OECD's explicit focus on policy relevance as well as a more intensive dissemination strategy that does not shy away from rankings. PISA releases its official reports every three years in December in a major media event, and those reports display country rankings in their first few pages. (In the 1990s, IEA studies also began publishing country rankings.) Many observers refer to these rankings by a new name: "league tables"—a term meant to invoke the relative standing of teams as reported in the sports pages of newspapers. In some countries, the release of PISA results triggers over 100 articles in the main national newspaper alone (Martins & Niemann, 2013). PISA is perceived as having a reputable "brand" among policy-makers and the press (Grek, 2009). Andreas Schleicher, the director of PISA, has become a well-known public figure, invited to give lectures to education ministries around the world and a TED Talk on the results of PISA. At least 18 national Ministries of Education have begun to set performance targets specifically benchmarked to PISA scores or rankings (Breakspear, 2012).

Unlike the IEA, the OECD attempts to identify policies and characteristics of successful systems, although it claims not to make explicit policy recommendations based on PISA. Further, the content of all OECD publications and recommendations must be approved by all OECD member states (though not the non-members), and these recommendations are not binding; countries may choose to implement them voluntarily (Bieber & Martens, 2011). Descriptions of policy best practices appear in the main PISA reports, in country-specific policy advice that the OECD produces on request, and in the video series co-produced with educational company Pearson called "Strong Performers and Successful Reformers" that profiles top-scoring systems.

One of the most consistent messages of PISA concerns educational equity. Even in its design, PISA has always had a greater emphasis on equity than do the IEA studies. PISA's practice of sampling students by age rather than by grade means that countries' results suffer if they have high rates of grade retention. PISA's student surveys also collect more information on family socio-economic context than do the IEA studies, which allows the OECD to conduct many within-country analyses of inequality between students of different socio-economic statuses, including the correlation between socio-economic status and achievement as well as the level of socio-economic segregation between schools (OECD, 2013a). But it is in the publicity surrounding some of the top-performing countries in PISA where the equity focus is most striking. Two of the surprise top-scoring countries in the first round of PISA in 2000 were Finland and Canada, both known for relatively equitable educational and social policies. Along with South Korea, the other top performer, all three countries had among the lowest levels of socio-economic segregation between schools and smallest impact of socio-economic status on achievement. In addition, Finland had among the smallest gender differences in achievement, and Canada had some of the smallest achievement gaps between immigrant and native-born students (OECD, 2004). These findings led to a major conclusion of PISA: that educational equity was compatible with educational excellence—and perhaps even led to excellence. Although both Finland's and Canada's scores declined somewhat in later years, equity remains a strong message of PISA. OECD publications urge extra support for low-achieving, socio-economically disadvantaged, immigrant and language learner students, and are critical of policies that tend to increase social segregation between schools, such as selective school admissions and between-school academic and vocational tracking (OECD, 2013a).

Other policies that the OECD has identified in top-scoring countries include high teacher quality, positive school climate, greater autonomy for local schools, rigorous academic standards, and accountability (OECD, 2004; 2013b). Here, the emphasis on standards, accountability and particularly the role of assessments merits further discussion, as this has been one of the most controversial impacts of PISA and is also an area where the evidence gathered by the OECD remains inconclusive. While some of the OECD's early case studies profiled in a positive light the development of different test-based accountability systems, such as those of Ontario, Brazil, and Germany (OECD, 2004; 2010), there were also many prominent examples of successful systems operating under different models, such as Finland, which has very little standardized testing, and Korea and Japan, which have high-stakes university entrance exams but limited use of standardized testing in early grades. More recently, in the PISA 2012 results, the OECD reported that countries where more schools' achievement data were tracked by administrative authorities in fact had *lower* average PISA scores (OECD, 2013b, p. 59). Despite this somewhat inconsistent picture, what is clear from the PISA principal surveys is that the amount of test-based accountability has increased quite dramatically in many countries since the start of PISA in 2000 (ibid).

What has been the impact of PISA on educational policy? Among OECD member countries, the response has varied greatly, from large reforms in many European and East Asian countries to little recognition of PISA in many English-speaking countries. The amount of national media coverage of PISA follows a similar pattern, with high coverage in Spain, Germany, Mexico, and Finland, and low coverage in the United States, the United Kingdom, Canada, and New Zealand. Media coverage appears unrelated to how well or poorly countries perform in PISA (Martens & Niemann, 2013). A number of countries experienced "PISA Shock"

following lower-than-expected results, prompting large reforms. Germany, after disappointing results in 2000, enacted national standards, assessments in each federal state, and greater support for disadvantaged students, particularly immigrants, among other reforms (Breakspear, 2012). Perhaps the most unexpected outcome of PISA has been the "Finnish Miracle". Finnish educators themselves were quite surprised at their system's top performance in PISA 2000 and at the massive international attention it drew to the small country (Grek, 2009). Ministry officials, researchers and journalists travelled to Finland to discover the secrets to its success, and Finnish ministry official Pasi Sahlberg's book *Finnish Lessons* became a best-seller. The Finnish craze even reached Japan, the country that had received the most international attention in the previous decade for its TIMSS results, but that had experienced a "PISA Shock" of its own when scores fell in 2003 and 2006 (although they were still far above the OECD average) (Takayama, 2009). Across Japan and other countries, scholars have observed that education reformers from both the right and the left used the external Finnish example to lend greater legitimacy to their preferred agendas (Dobbins & Martens, 2012). On the other hand, the English-speaking world has only recently begun to take notice of PISA. In the United Kingdom, PISA received little attention until after 2006 when performance appeared to fall (possibly due to correction of sampling problems in earlier waves) and a new government was interested in criticizing old policy (Baird et al., 2011). In the United States, PISA went relatively unnoticed until 2009, when the Chinese city of Shanghai participated for the first time and topped the rankings, prompting a massive response in the national media and government, which often interpreted the results for the single city of Shanghai as representing the entire country of China (Baird et al., 2011; Martens & Niemann, 2013). Overall, it appears that the policy effects of PISA are increasing over time, and

that some of the most common policy responses are implementing national standards and assessments, as well as aligning these standards and assessments with PISA (Breakspear, 2012).

PISA and other ILSAs are not themselves used for accountability purposes. As ILSAs are administered only to small, randomly-selected, anonymized samples of students in each country, they cannot have high stakes for individual participating students, teachers, or schools, as domestic assessments can. However, ILSAs and domestic assessments do share some similarities: Both measure achievement on a standardized scale for purposes of comparison, whether comparison of countries or of schools. Both often result in visible public rankings that may end up "naming and shaming" low-performing systems. Thus, it could be argued that, although ILSAs are not high-stakes for students, they can have stakes for national actors, such as ministers of education. Finally, as described above, many countries have begun aligning their national assessments with PISA's concepts of literacy and problem-solving, creating some convergence in the content of the tests themselves. (See Chapter Five by Anderson and Sivasubramaniam for more discussion of testing and assessment programs.)

PISA (as well as other ILSAs and indicator projects) also have far-reaching policy impact within the developing world. As early as 1990, the international community linked better assessment to the achievement of education as a universal right (as described by Mundy and Read in Chapter Eleven). Article 4 of the *World Declaration on Education for All* (adopted in Jomtien, Thailand, in 1990 and ratified in Dakar, Senegal, in 2000), states, "It is necessary to define acceptable levels of learning acquisition for educational programmes and to improve and apply systems of assessing learning achievement" (World Conference on Education for All, 1990, p. 36). Many developing countries produce shockingly low levels of learning acquisition: basic levels of literacy and numeracy are often not acquired during the full primary cycle,

causing mounting international concern with educational quality (an issue that is sometimes neglected in the push for greater access). Beginning in the 1990s, many governments introduced national assessment programs; in several countries whose governments did not initiate assessments, citizen-led assessments emerged (e.g., in India, Pakistan, Kenya, and others) (Results for Development, 2015). According to a recent count, 65 percent of developing countries now have national assessments (Benavot & Köseleci, 2015). Regional indicator and assessment programs were undertaken in Southern Africa and Latin America—most notably the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) and the Latin American Laboratory for Evaluating the Quality of Education (LLECE).[2] Many other countries simply joined in the IEA's TIMSS or the OECD's PISA program: non-OECD countries represent more than half the current participants of both assessments. The relatively demanding TIMSS and PISA tests generally draw middle-income countries, and some developing countries have stopped participating in ILSAs after disappointing results (Wiseman, 2013). Thus far, developing countries appear to favour using their own national assessments over participation in regional or international assessments (Kamens & Benavot, 2011). Participation in all of these types of assessments is heavily supported by the World Bank, which increasingly sees assessments as essential to both the efficient allocation of scarce educational resources, and as providing a key way of mobilizing policy-makers around educational reform programs. There is some evidence that developing countries that participate in ILSAs also subsequently receive more foreign aid to education (Kijima, 2010). The new version of Article 4 in the UN's post-2015 Sustainable Development Goals for the first time explicitly mentions using international, regional, and national assessments to monitor educational quality. In response, both the OECD and the IEA are implementing new assessments aimed at monitoring more basic skills in

developing countries (PISA for Development, TIMSS Numeracy, and PIRLS Literacy). Such efforts have raised criticisms of increasing pressure to participate in ILSAs and the creation of a de facto global curriculum. However, UNESCO officials argue that universal participation in ILSAs is unnecessary; instead, common scales can be developed for comparing disparate national assessments (Benavot & Köseleci, 2015; Rose 2015).

Canada and the International Indicators

What implications and issues are raised by the rapid growth of international indicator programs for educators? In this section, we answer this question by first describing Canadian involvement in international assessments, and then looking at some of the findings from these assessments and the policy debates they have stimulated. Our goal here is to provide educators with a practical guide to international indicators, and to encourage greater critical engagement with them, using the Canadian experience as an illustration.

Table 13.2

A Guide to Current International Assessments in Canada

Title (Organization)	Goals/ Method	Years	Canadian Participation
PISA Programme for International Student Assessment (OECD)	• Reading, math and science • 15-year-olds (regardless of grade) • Tests literacy, with emphasis on problem solving and application of knowledge (does not test mastery of a curriculum) • Student and school contextual questionnaires	Every 3 years starting in 2000	All years, all provinces

TIMSS Trends in International Mathematics and Science Study (IEA)	• Science and math • Grades 4 and 8 • Curriculum-based tests (based on international consensus curriculum) • Student, teacher and school contextual questionnaires; national context survey	Every 4 years starting in 1995	1995 & 1999 - all provinces; since 2003 - ON & QC in all years, BC & AB in some years
PIRLS Progress in Reading Literacy Study (IEA)	• Reading • Grade 4 • Tests literacy in both literary texts and informational documents • Student, teacher, school and parent contextual questionnaires; national context survey	Every 5 years starting in 2001	2001 - QC & ON; 2006 - AB, BC, NS, QC, ON; 2011 - All provinces
Civics Education Studies: **CIVED (1999)** Civic Education Study **ICCS (from 2009)** International Civics and Citizenship Study (IEA)	• Civic and citizenship knowledge and attitudes • Grade 8 • Achievement test assesses civic knowledge and interpretation of democracy/citizenship, identity/intl. relations, social cohesion/diversity • Student, teacher and school contextual questionnaires • CIVED included qualitative country case studies	CIVED 1999; ICCS every 7 years starting in 2009	Canada participated in CIVED qualitative case study only, not assessment; did not participate in ICCS
Surveys of Adult Skills: **IALS (1994)** International Adult Literacy Survey (StatsCan) **ALL (2003)** Adult Literacy and Lifeskills Survey (StatsCan) **PIAAC (from 2011)** Programme for the International Assessment of Adult Competencies (OECD)	• Literacy and numeracy skills • Adults ages 16-65 • Tests literacy in both prose and informational documents; numeracy applied to real-life situations (not curriculum-based) • Test was computer-based starting in 2011 • Participant contextual questionnaires	IALS 1994; ALL 2003; PIAAC every 10 years starting in 2011	All years, all provinces

Computer Studies: **COMPED (1992)** Computers in Education Study **SITES (2006)** Second Information Technology in Education Study **ICILS (2013)** International Computer and Information Literacy Study (IEA)	• Computer and information literacy • Grade 8 • Computer-based test with questions testing knowledge and tasks using simulated software applications • Student, teacher and school contextual questionnaires; national context survey • SITES included qualitative country case studies	COMPED 1992; SITES 2006; ICILS every 5 years starting in 2013	COMPED - BC only SITES - all provinces participated in qualitative case study; only AB & ON participated in assessment; ICILS - NL & ON only

Canada (through the national Council of Ministers of Education Canada [CMEC], Human Resources and Skills Development Canada [HRSDC], Statistics Canada [StatsCan], and the provincial Ministries of Education) has been quite an active participant in many of the major international assessment exercises since relatively early. Canada's first participation in an ILSA came in 1980 when British Columbia and Ontario took part in the Second International Mathematics Study (SIMS). As education in Canada is a provincially mandated responsibility, in most of the early ILSAs, only some provincial ministries of education made the decision to participate. In TIMSS 1995 and 1999, for the first time all ten provinces participated and were reported as a single country. Throughout the IEA assessments of the 1980s and 1990s, Canada generally achieved above-average results compared to other countries. Canada has also played a leadership role in ILSA design and analysis. SIMS was headed by David Robitaille, professor of mathematics education at the University of British Columbia. Robitaille was also involved in the initial planning for TIMSS 1995. Statistics Canada led the administration of the first two surveys of literacy skills for adults ages 16-65, IALS 1994, and ALL 2003.

Table 13.2 lists the various international assessments in which Canada currently participates. Since 2003, only Ontario and Quebec have consistently participated in every year of TIMSS. Both provinces generally score above the international average, with Quebec outperforming Ontario in math and both provinces performing similarly in science (Mullis et al., 2012). Different provinces have participated in each year of the IEA's Grade 4 reading test (PIRLS) and Grade 8 computer skills test (ICILS). No Canadian provinces currently participate in the IEA's Grade 8 civics test (ICCS). PISA constitutes a marked change for Canadian participation in ILSAs because all 10 provinces (though no territories) have participated in every wave of the study. Unlike in the IEA studies, which are generally funded by the participating provincial Ministries of Education, direct costs for PISA are paid for by the federal ministry, HRSDC, with provinces, CMEC, and StatsCan as collaborating partners. Moreover, Canada is one of a handful of federal countries that chooses to select an especially large PISA sample in order to obtain reliable results for individual provinces or states. Rather than the typical PISA sample of 4,500 students, Canada selects over 20,000 students from across the 10 provinces in each wave of PISA. This enables comparison of results for all provinces, as well as for English and French systems where applicable.

As described in the previous section, Canada has received a great deal of international attention for its high and equitable performance in PISA. In particular, recent educational reforms in Ontario have been profiled in the OECD's reports on "Strong Performers and Successful Reformers" and the OECD/Pearson video series of the same name, as well as reports by McKinsey & Company and the National Centre on Education and the Economy (Tucker, 2011; OECD, 2010; Mourshed, Chijioke, & Barber, 2010). Alberta also was covered in the *Economist* (2006) for its top performance and its education policies emphasizing school choice,

competition, and accountability. Inside Canada, the picture is different. As in the United Kingdom and United States, initial media attention to PISA was rather low (Grek, 2009; Martens & Niemann, 2013). As coverage began to grow, it was primarily focused on ranking the provinces rather than international comparisons (Stack, 2006). Generally, British Columbia and Alberta are the highest scoring provinces, although Quebec performs well in math and Ontario in reading; the Atlantic provinces tend to have the lowest average scores (Brochu et al., 2013). Provincial league tables may have been fascinating to many Canadians because they were relatively new. When the first results of PISA 2000 were released, CMEC's federal testing program, then known as the School Achievement Indicators Program (SAIP), was less than 10 years old (CMEC, n.d.). As with international results, it is difficult to prove whether particular policies in each province are the cause for these results. Ontario's curricular reforms profiled in the OECD reports above began in 2003, after the province had already achieved relatively high scores in the first wave of PISA (Stack, 2006).

In comparison to other countries, the policy impact of PISA within Canada has been medium (Breakspear, 2012). The largest impact at the federal level has been on the federal testing program, which in 2007 was revised to be more aligned with PISA (at that time, it was also renamed from SAIP to the Pan-Canadian Assessment Program (PCAP) (CMEC, n.d.)). Similarly to PISA and other ILSAs, PCAP also tests a randomly-selected sample of students from within each province and each of the English and French systems, in order to perform comparisons. Some have commented that PISA may provide external legitimacy to unified federal indicators that otherwise would have been difficult for provinces to agree on (Smith & Baker, 2001). Others have criticized PISA for increasing pressure toward standardization and uniformity across the provinces, and HRSDC's involvement in PISA as marking a new trend in

federal intervention in education. PISA has also had policy effects for individual provinces. For example, Ontario has used PISA results to monitor and validate the effectiveness of its recent reforms, Prince Edward Island has implemented provincial assessments partly in response to its low initial PISA performance, and New Brunswick has set targets to improve its ranking within Canada (Baird et al., 2011). Again, these policy effects are primarily focused on domestic rather than international comparisons.

More recently, Canada's relatively small but statistically significant declines in scores in PISA 2009 and 2012 have caused alarm among policy-makers and researchers and prompted calls for curricular reforms (Alphonso, 2013; The Canadian Press 2013). Falling scores call into question the success of educational reforms implemented in many provinces over the past decade. However, it is important to recognize that these declines are relatively small—in fact, Finland's performance has fallen much more precipitously during the same period. Additionally, Canada's achievement remains among the most equitably distributed in the world, particularly for immigrant students. Yet both ILSA results and the experiences of Canadian educators remind us that socio-economic segregation and achievement disparities are far from zero. This serves to highlight the persistent severity of educational inequality throughout the developed and developing world.

Conclusion: International Indicators—What Teachers Need To Know

In May 2014, a group of nearly 100 professors and educators wrote an open letter to Andreas Schleicher that was published in the *Guardian*. They called for a moratorium on PISA testing until the OECD addresses a number of criticisms, including an overreliance on league tables; an overly narrow focus on economically relevant skills and neglect of students' civic,

moral, and artistic development and well-being, leading to a narrowing of school curricula; partnerships with for-profit educational companies (Pearson co-produced the "Strong Performers and Successful Reformers" video series and had recently won the contract to develop the next round of PISA); and inadequate involvement of a wide range of stakeholders, such as teachers, parents, school administrators, and other international organizations such as the United Nations (Meyer and et al., 2014). The OECD responded with a public statement saying that "less than 1% of the PISA reporting is devoted to league tables"; that PISA contextual surveys collect a wide range of information including student attitudes, motivation and socio-economic factors; that PISA contractors, whether for- or not-for-profit, win contracts through open competition; and that all OECD member countries (though not non-members) have equal representation in PISA governance (OECD, 2014b). Later, Pasi Sahlberg and Boston College education professor Andy Hargreaves wrote a blog post in the *Washington Post* arguing that PISA is flawed but should be saved. While they also had misgivings about for-profit contractors, they pointed out that PISA's equity emphasis had had a positive impact on education policy, drawing global attention toward the success of relatively equitable countries such as Finland and Canada and away from market-based reform trends in the United States and United Kingdom (Sahlberg & Hargreaves, 2015).

ILSAs are clearly highly controversial. On the one hand, they have highlighted countries where all students have high OTL and equitable access to education; and they have exposed other countries' low and unequal distribution of opportunity, shaking them out of their parochialism and false assumptions that their schools are the best in the world. On the other hand, ILSAs are widely used for simplistic league table comparisons and mischaracterizations of successful systems like Finland's and Japan's, as well as for justification of test-based accountability and other reforms with limited evidence of effectiveness. Some of the

responsibility for these effects is borne not by the OECD and IEA themselves but by policy-makers and journalists. Both the IEA and the OECD publish each ILSA along with thousands of pages of documentation and analyses of national contexts, much of which gets ignored. When one encounters rhetoric about ILSAs, it is important to keep in mind:

- Rankings can be misleading. They can exaggerate small differences: The country in first place and the country in fifth place may have average scores that differ by only 10 points on a 1000-point scale. When examining changes in performance over time, it is preferable to look at changes in scores rather than changes in rankings, as rankings can be affected by new countries joining. Some news and political outlets publish truncated league tables showing only those countries that score higher than the country in question, giving the false impression that that country ranks in last place.
- It is important to look beyond country averages to the distribution of scores within countries. Every ILSA to date has found more variation within countries than between them (i.e., not every student in Korea outscores every student in Spain). Variation within countries can draw attention to unequal opportunities to learn.
- Correlation is not causation. ILSAs give us a snapshot of how countries perform at a single moment in time, but cannot tell us which policies caused these results. It is particularly faulty logic to attribute a country's performance to educational reforms that are currently underway, rather than recognizing that 15-year-old students are the products of 15 years of experiences both in and out of school—and that countries are the products of hundreds or thousands of years of historical development.

Questions for Reflection and Discussion

1. Why have governments become so interested in participating in efforts to compare their educational performance?

2. What kinds of limits should we be aware of when assessing information from international assessments and indicators?

3. How does the collection of international assessments and indicators differ in developed and developing countries? What practical, political, and/or ethical issues does this pose?

4. In your view, do international assessments and indicators offer valuable information for educators? Explain.

Suggested Audio-Visual Resources

Instead of selecting a film for this chapter, we suggest that instructors explore with students the activities and audio-visual resources provided on international assessment websites. Holding this class in a computer lab would be ideal.

a) Explore TIMSS questions: timssandpirls.bc.edu/timss2011/international-released-items.html

b) Explore PISA questions: www.oecd.org/pisa/test

c) Watch footage of math and science classrooms around the world from the TIMSS 1999 Video Study: www.timssvideo.com

d) Compare and contrast the videos above with the OECD/Pearson video series "Strong Performers and Successful Reformers" www.oecd.org/pisa/pisaproducts

Suggestions for Further Reading

- Council of Ministers of Education Canada. (2013). *Measuring Up: Canadian Results of the OECD PISA Study: 2012 First Results for Canadians Aged 15*. Toronto: Council of Ministers of Education Canada.
- Meyer, Heinz-Dieter & Aaron Benavot. (2013). *PISA, Power, and Policy: The Emergence of Global Educational Governance*. Southampton: Symposium Books.
- OECD. (2014). *Education at a Glance 2014: OECD Indicators*. Paris: OECD.
- Pizmony-Levy, Oren et al. (2014). "On the merits of, and myths about, international assessments. [Moderated discussion]." *Quality Assurance in Education* 22(4):319-338.
- Smith, Thomas A. & David P. Baker. (2001). "Worldwide Growth and Institutionalization of Statistical Indicators for Educational Policy Making." *Peabody Journal of Education* 76, no. 3–4: 141–52.
- UNESCO. (2015). *Education for All 2000-2015: Achievements and Challenges*. Paris: UNESCO.
- UNICEF. Child Poverty in Perspective: An Overview of Child- Well-being in Rich Countries. *Innocenti Report Card 7*. Florence: UNICEF Innocenti Research Centre, 2007. www.unicef-irc.org/ publications/pdf/rc7_eng.pdf.

Endnotes

1. For an overview of the IBE's *International Education Yearbook*, see UNESCO, "Unesco 50 Years for Education," (UNESCO, 1997b)

2. Information about the SACMEQ and the LLECE programs can be found online at: www.sacmeq.org and www.llece.org.

References

Alphonso, Caroline. (2013, December 3). Canada's Fall in Math-Education Ranking Sets Off Alarm Bells. *Globe and Mail*. Retrieved from http://www.theglobeandmail.com/news/national/education/canadas-fall-in-math-education-ranking-sets-off-red-flags/article15730663/.

Baird, Jo-Anne., Isaccs, Talia., Johnson, Sandra., Stobart, Gordon., Yu, Guoxing., Sprague, Terra., and Daugherty, Richard. (2011). *Policy Effects of PISA.* Oxford: Oxford University Centre for Educational Assessment.

Benavot, Aaron, and Köseleci, Nihan. (2015). Seeking Quality in Education: The Growth of National Learning Assessments, 1990-2013. Background paper prepared for the Education for All Global Monitoring Report 2015. *Education for All 2000-2015: Achievements and Challenges.* ED/EFA/MRT/2015/PI/53.

Bieber, Tonia., and Martens, Kerstin. (2011). The OECD PISA Study as a Soft Power in Education? Lessons from Switzerland and the US. *European Journal of Education, 46*(1): 101-16.

Breakspear, Simon. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*. OECD Education Working Papers, No. 71, Paris: OECD.

Brochu, Pierre., Deussing, Marie-Anne., Houme, Koffi., and Chuy, Maria. (2013). *Measuring Up: Canadian Results of the OECD PISA Study*. Toronto: Council of Ministers of Education.

CMEC. (no date). *Overview: Pan-Canadian Assessment Program (PCAP).* Retreived from http://www.cmec.ca/240/Programs-and-Initiatives/Assessment/Pan-Canadian-Assessment-Program-(PCAP)/Overview/index.html.

Dobbins, Michael., and Martens, Kerstin. (2012). Towards an Education Approach À La Finlandaise? French Education Policy after PISA. *Journal of Education Policy, 27*(1): 23-43.

Farrell, Joseph P. (2007). Literacy and International Development: Education and Literacy as Basic Human Rights. In David R. Olsen and Nancy Torrance (Eds.), *Handbook of Literacy*. Cambridge, UK: Cambridge University Press, pp. 518-534.

Givven, Karen Bogard., Hiebert, James., Jacobs, Jennifer K., Hollingsworth, Hilary., and Gallimore, Ronald. (2005). Are There National Patterns of Teaching? Evidence from the Timss 1999 Video Study. *Comparative Education Review, 49*(3): 311-43.

Grek, Sotiria. (2009). Governing by Numbers: The PISA 'Effect' in Europe. *Journal of Education Policy, 24*(1): 23-37.

Holmes, Brian., and Robinsohn, Saul. (1963). *Relevant Data in Comparative Education*. Hamburg, Germany: UNESCO Institute for Education.

Husén, Torsten. (Ed.). (1967a). *International Study of Achievement in Mathematics* (Vol. I.). Hamburg: Evaluation of Educational Achievement.

Husén, Torsten. (Ed.). (1967b). *International Study of Achievement in Mathematics* (Vol. II.). Hamburg: Evaluation of Educational Achievement.

IEA. (no date). *Brief History of IEA: 55 Years of Educational Research*. Retreived from http://www.iea.nl/brief_history.html.

Kamens, David H., and Benavot, Aaron. (2011). National, Regional and International Learning Assessments: Trends among Developing Countries, 1960–2009. *Globalisation, societies and education, 9*(2): 285-300.

Kijima, Rie. (2010). Why Participate? Cross-National Assessments and Foreign Aid to Education. In Alexander W. Wiseman (Ed.), *The Impact of International Achievement Studies on National Education Policymaking*, Bradford, West Yorkshire: Emerald Group Publishing, 35-61.

Martens, Kerstin., and Niemann, Dennis. (2013). When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US. *German Politics, 22*(3): 314-332.

Martin, Michael O., and Mullis, Ina V. S. (2013). Methods and Procedures in TIMSS and PIRLS 2011. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA.

McDonnell, Lorraine M. (1995). Opportunity to Learn as a Research Concept and a Policy Instrument. *Educational Evaluation and Policy Analysis, 17*(3): 305-22.

Meyer, Heinz-Dieter, et al. (2014, May 6). OECD and PISA Tests Are Damaging Education Worldwide: Academics. *The Guardian*. Retreived from http://www.theguardian.com/education/2014/may/06/oecd-PISA-tests-damaging-education-academics.

Mourshed, Mona., Chijioke, Chinezi., and Barber, Michael. (2010). *How the World's Most Improved School Systems Keep Getting Better*. London: McKinsey & Company.

Mullis, Ina V. S., Martin, Michael O., Foy, Pierre., and Arora, Alka. (2012). *TIMSS 2011 International Results in Mathematics.* Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA.

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. In *A Report to the Nation and the Secretary of Education by the National Commission on Excellence in Education*: US Department of Education.

OECD. (2004). *What Makes School Systems Perform? Seeing School Systems through the Prism of PISA.* Paris: OECD.

OECD. (2010). *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*. Paris: OECD.

OECD. (2013a). *PISA 2012 Results: Excellence through Equity: Giving Every Student the Chance to Succeed* (Vol. II). Paris: OECD Publishing, 2013.

OECD. (2013b). *PISA 2012 Results: What Makes Schools Successful?* (Vol. IV). Paris: OECD, 2013b.

OECD. (2014a). *PISA 2012 Technical Report*. Paris: OECD.

OECD. (2014b). *Response to Points Raised in Heinz Dieter-Meyer "Open Letter."* Paris: OECD. Retreieved from http://www.oecd.org/PISA/aboutPISA/OECD-response-to-Heinz-Dieter-Meyer-Open-Letter.pdf, 2014.

OECD. (no date) *About the OECD*. Retrieved from http://www.oecd.org/about/.

Pizmony-Levy, Oren. (2014). Back to the Future in International Assessments. In Madhabi Chatterji and James Harvey (Eds.) *EdWeek Blog: Assessing the Assessments: K-12 Measurement and Accountability in the 21st Century*, Bethesda, MD: Education Week. Retrieved from http://blogs.edweek.org/edweek/assessing_the_assessments/2014/04/back_to_the_future_how_international_large-scale_assessments_came_about.html.

Puryear, Jeffrey. (1995). International Education Statistics and Research: Status and Problems. *International Journal of Educational Development, 15*(1): 79-91.

Results for Development. (2015, June). *Bringing Learning To Light: The Role of Citizen Led Assessments in Shifting the Education Agenda*. Washington DC: R4D and the Hewlett Foundation. Retrieved from http://www.hewlett.org/sites/default/files/R4D%20-%20Bringing%20Learning%20to%20Light%20-%20June%202015.pdf

Rose, Pauline. (2015). Is a Global System of International Large-Scale Assessments Necessary for Tracking Progress of a Post-2015 Learning Target? *Compare, 45*(3): 486-490.

Sahlberg, Pasi., and Hargreaves, Andy. (2015, March 24). The Tower of PISA Is Badly Leaning. An Argument for Why It Should Be Saved. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/24/the-tower-of-PISA-is-badly-leaning-an-argument-for-why-it-should-be-saved/

Schiefelbein, Ernesto. (2006). *School Performance Problems in Latin America: The Potential Role of the Escuela Nueva System*. Paper presented at the Second International New Schools Congress. Medellin, Colombia.

Smith, Thomas M., and Baker, David P. (2001). Worldwide Growth and Institutionalization of Statistical Indicators for Education Policy-Making. *Peabody Journal of Education, 76*(3&4): 141-52.

Stack, Michelle. (2006). Testing, Testing, Read All About It: Canadian Press Coverage of the PISA Results. *Canadian Journal of Education, 29*(1): 49-69.

Takayama, Keita. (2009). Politics of Externalization in Reflexive Times: Reinventing Japanese Education Reform Discourses through "Finnish PISA Success". *Comparative Education Review, 54*(1): 51-75.

The Canadian Press. "Canada's Students Slipping in Math and Science, OECD Finds." *CBC News*, December 3 2013.

The Economist. (2006, September 21). *Clever Red-Necks: It's Not Just the Economy That Is Booming; Schools Are Too*. Retrieved from http://www.economist.com/node/7945805.

Torney-Purta, Judith., Schwille, John., and Amadeo, Jo-Ann. (1999). *Civic Education across Countries: Twenty-Four National Case Studies from the IEA Civic Education Project*. Amsterdam, Netherlands: International Association for the Evaluation of Educational Achievement.

Tucker, Marc S. (2011). *Standing on the Shoulders of Giants: An American Agenda for Education Reform*. Washington, DC: National Center on Education and the Economy, 2011.

UNESCO. (1997a). *International Standard Classification of Education ISCED97*. Paris: UNESCO. Retrieved from http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

UNESCO (1997b). UNESCO 50 Years for Education. Paris: UNESCO. Retreived from http://unesdoc.unesco.org/images/0011/001102/110264eb.pdf

UNESCO. (2011). *International Standard Classification of Education ISCED 2011*. Paris: UNESCO.

UNESCO. (2015). Education for All 2000-2015: Achievements and Challenges. *Education for All Global Monitoring Report*, Paris: UNESCO.

UNESCO Institute of Statistics. (2012). *Global Education Digest: Opportunities Lost: The Impact of Grade Repetition and Early School Leaving*. Montreal, QC: UNESCO Institute of Statistics.

UNICEF. (2013). Child Well-Being in Rich Countries: A Comparative Overview. *Innocenti Report Card 11*. Florence, Italy: UNICEF Innocenti Research Centre.

Wiseman, Alexander W. (2013). Policy Responses to PISA in Comparative Perspective. In Heinz-Dieter Meyer and Aaron Benavot (Eds.), *PISA, Power, and Policy: The*

Emergence of Global Educational Governance, Southampton: Symposium Books, 303-322.

World Conference on Education for All. (1990). *Meeting Basic Learning Needs: A Vision for the 1990s*. New York: Inter-Agency Commission for World Conference on Education for All.