

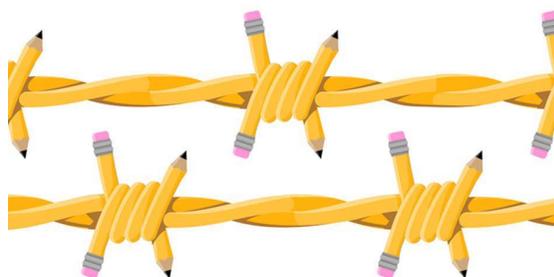
Chap 5
OECD, PISA, AND
GLOBALIZATION:
THE INFLUENCE OF THE
INTERNATIONAL ASSESSMENT
REGIME

Svein Sjøberg University of Oslo,
Norway

EDUCATION POLICY PERILS

TACKLING THE TOUGH ISSUES

Foreword by Fenwick W. English



Edited by **Christopher H. Tienken**
and **Carol A. Mullen**



Tienken, C. H. & Mullen, C. A. (Eds) (2016). Education Policy Perils.
Tackling the Tough Issues. Routledge
<https://www.routledge.com/products/9781138898196>

Reviews

"Enlightening, insightful, practical, and evidence based. A much-needed masterful guide for traversing the bewildering and tough education policy terrain."

--**Yong Zhao**, Ph. D., author of *Who's Afraid of the Big Bad Dragon: Why China Has the Best (and Worst) Education System in the World*

"Our democracy thrives when smart policy replaces bad policy. This book points us in that direction with a variety of chapters, viewpoints, critiques and, best of all, ideas about how to solve some of the pressing education policy issues of our times. A much-needed analysis of how we have been led astray and what we might do about it."

--**David C. Berliner**, Regents' Professor Emeritus of Education, Arizona State University

OECD, PISA, AND GLOBALIZATION: THE INFLUENCE OF THE INTERNATIONAL ASSESSMENT REGIME¹

Svein Sjøberg

UNIVERSITY OF OSLO

EDUCATION
POLICY PERILS
TACKLING THE TOUGH ISSUES
Foreword by Femco H. Engen



Edited by Christopher H. Tienken
and Carol A. Mullen 

Beginning in the mid-1990s, the Organisation for Economic Cooperation and Development (OECD) started creating the Programme for International Student Assessment, now well known as PISA. Since the first publication of PISA rankings in 2001, based on the testing that took place in 2000, the results have become a global gold standard for educational quality. Although the political and educational importance of PISA varies from one country to another, the results often set the scene for public debate on the quality of education for the more than 65 countries and cities that participate in the test. PISA performance tables are widely published in mass media and also used by politicians and education policymakers. In many countries, educational reforms are launched in direct response to PISA results. The testing takes place every three years; including the results from PISA 2012 testing (OECD, 2014), we now have data from five rounds of PISA.

The intentions of PISA are related to the overall political aims of the OECD and the underlying commitment to a competitive global free-market economy. PISA was constructed and intended for the 30+ industrialized and wealthy OECD countries, but a similar number of countries and economies have subsequently joined. When the most recent PISA results were presented, OECD (2014) claimed to have participation from 80% of the global economy. For educators, this may seem a surprising way of reporting student participation; but it indicates the focus of the PISA project: economics. The economic focus might also account for the extreme importance that is now attributed to PISA rankings. It seems common sense that high scores on PISA reading, mathematics, and science are predictors for the country's future economic competitiveness. Bad rankings from PISA are thought to be ominous signals for the economic future of a country.

I will return to the crucial point of interpreting rankings toward the end of this chapter.

Tables of country rankings on PISA scores are often taken at face value, not only in the media, but also by policymakers and politicians. The PISA undertaking is a well-funded multinational techno-scientific machine—undoubtedly the world's largest empirical study of schools and education. Recent estimates suggest that the annual cost, after adding the expense of involving half a million students and tens of thousands of schools and their teachers, is approximately 80 million Euro (Maslen, 2013). Given its size and importance, PISA has to be understood not just as a study of student learning, but also as a social phenomenon in a wider political, social, and cultural context (Lundgren, 2011).

PISA rankings create panic and discomfort among policymakers in practically all countries, including the high-scoring ones (Alexander, 2012). The discomfort produces an urge for politicians and bureaucrats to do *something* to rectify the situation that they believe the results describe. However, because PISA does not tell much about cause and effect, the creative educational reforms that are not at all empirically founded are introduced, often overnight.

This chapter will raise many important questions about the PISA project, focused on two critical arguments with implications for education policy-making. The first argument relates to the PISA project itself and is that basic structural problems are inherent in the PISA undertaking and, hence, cannot be *fixed*. I will argue that it is impossible to construct a test that can be used across countries and cultures to assess the quality of learning in real life situations with authentic texts. Problems arise when the intentions of the PISA framework are translated into concrete test items to be used in a great variety of languages, cultures, and countries. The requirement of “fair testing” implies by necessity that local, current, and topical issues must be excluded if items are to transfer objectively across cultures, languages, and customs. This runs against most current thinking in science education, where “science in context” and “localized curricula” are ideals promoted by UNESCO and many educators, as well as in national curricula.

My second argument relates to some of the rather intriguing results that emerge from analyses of PISA data. It seems that pupils in high-scoring countries also develop the most negative attitudes toward the subjects on which they are tested. It also seems that PISA scores are unrelated to educational resources, funding, class size, and similar factors. PISA scores also seem to be negatively related to the use of active teaching methods, inquiry-based instruction, and computer technology. PISA scores seem to function like a kind of IQ test on school systems. A most complex issue is reduced to simple numbers that may be ranked with high accuracy. But, as with IQ scores, there are serious concerns about the validity of the PISA scores. Whether one believes in the goals and results of PISA, such issues need to be discussed.

One positive aspect of PISA is that it has brought schools and education to the forefront in the media and in political debates internationally, and even more so nationally in many countries. However, the PISA results seem to be accepted at face value, and there are few critical voices. The focus of this chapter, therefore, will be on the especially problematic sides of PISA testing.

What Does PISA Claim to Measure?

The emerging picture of what PISA tests measure is in many ways confusing. In some places, the PISA authors claim that the tests do *not* measure school knowledge or competencies acquired at schools; but in other places, they state that the tests actually do measure the quality of the nations' school systems. Let us consider some details.

The overall aims of PISA had already been stated in 1999, before the first PISA testing took place in 2000:

How well are young adults prepared to meet the challenges of the future?
Are they able to analyse, reason and communicate their ideas effectively?
Do they have the capacity to continue learning throughout life?
Parents, students, the public and those who run education systems need to know.

(OECD, 1999, p. 7)

Those exact words have been repeated in practically all PISA reports from the OECD during the more than 15 years since then. One can hardly object to ambitions like these. It would be great if PISA really did answer these questions, as the authors have claimed. In other reports, however, the authors are more modest. They emphasize that PISA scores do not actually provide measures of the quality of education systems, but rather the collective results of the school, home, and social environments.

PISA is rather explicit that the tests do *not* measure quality according to national school curricula, but measure based on the definitions and framework made by the OECD-appointed PISA experts (OECD, 2006a). The PISA framework clearly states that the knowledge and skills tested by PISA “are defined not primarily in terms of a common denominator of national school curricula but in terms of what skills are deemed to be essential for future life” (OECD, 2006a, p. 11). In essence, the PISA creators are claiming that they have identified the critical skills necessary for future life, for all humans on the planet. The same framework also states that tests exclude elements that are specific to a country's school curriculum.

So, although PISA states that it does not test knowledge acquired in schools and that it does not test according to national curricula, the PISA results are interpreted by OECD officials and policymakers around the globe as valid

measures of the quality of national school systems, and the PISA reports are chock-full of policy recommendations regarding schools (Loveless, 2009).

The Politics of the OECD PISA Project

OECD is an organization for the highly industrialized and economically developed nations. The mandate of the organization lies in the name: Organization for Economic Cooperation and Development. The home site (www.oecd.org) is explicit about the OECD's mission. Its aim is, above all, to promote policies and set standards for economic development in a global, competitive, free-market economy. One should remember that the E in OECD stands for Economic, not Educational. But education is certainly a driving force in economic development and national competitiveness, and has therefore become an important element of the OECD's concerns and policy advice.

The mandate of the OECD also explains why the PISA subjects are reading, mathematics, and science. According to OECD officials, these subjects are seen to be key elements for competitiveness in a world economy driven by science and technological development. But this selection of subjects also carries an implicit message about what is considered to be important in schools and in the development of young people. One should note the domains that are *not* included when PISA measures the quality of schools: the humanities, social sciences, ethics, foreign languages, history, geography, and physical education. One might also note that PISA does not address attributes that are central in many countries' official purposes of education, such as equity, empathy, solidarity, curiosity and engagement, and care for the environment. In public and political debates, statements about the agreed (and legally binding) purposes of the school system are often forgotten or ignored when the quality of the school is based on PISA scores and rankings.

It is interesting to note that in the PISA 2012 testing, a new component was added: "financial literacy" (OECD, 2013a). Of course the addition was as a consequence of the free-market economic mandate and priorities of the OECD. This module was included in the testing by some 40 countries.

The OECD is often very clear about the economic purposes of PISA and the competitive, international nature of the PISA rankings:

In a global economy, the yardstick for success is no longer improvement by national standards alone, but how education systems perform internationally. The OECD has taken up that challenge by developing PISA, the Programme for International Student Assessment, which evaluates the quality, equity and efficiency of school systems in some 70 countries that, together, make up nine-tenths of the world economy.

(OECD, 2010a, p. 3)

There seems to be a contradiction here regarding the structural nature of PISA and what the authors claim to measure. On the one hand, the OECD claims that PISA does not measure quality according to school curricula, or even the knowledge acquired at school. On the other hand, the OECD claims that it does evaluate “the quality, equity and efficiency of school systems.” As mentioned, it is also interesting that the overall importance of PISA is defined in terms of the fraction of the world economy tested, rather than in terms of the fraction of the world’s population, further indicating its structural focus on competitive economics, not comprehensive education.

The competitive economic perspective is also at the forefront when PISA results are presented to the public. At the PISA 2006 Release Conference in Washington, DC, on December 4, 2007, a portion of the invitation read as follows:

Losing Our Edge: Are American Students Unprepared for the Global Economy?

The lessons learned from the PISA results . . . can, and should, be used to inform U.S. education policy so that our students graduate . . . ready to compete, thrive, and lead in the global economy of the twenty-first century.

(Alliance for Excellent Education, 2007)

The political, economic, and indeed normative use of PISA by the OECD is also very clear. The OECD makes regular economic reports to many countries, with advice on future policy. My own country—Norway—is an example. In the Economic Survey report to Norway in 2008, OECD experts gave the following general advice: Norway should increase differences in salaries, reduce public spending, increase the rate of unemployment, reduce the level of sick leave salaries, and reduce pensions for people with disabilities (OECD, 2008). This advice was given just before the financial crisis.

This particular OECD report to Norway had the education system as the focus. With PISA data as input for its calculations, OECD gave advice on how to make Norwegian schools better. The operational definition of a “better school” was a school that is “cost-effective,” that is, could give more PISA points per dollar spent on education. The very definition of a good school thereby ignored the national priorities set for our school system. The OECD educational advice was that Norwegian schools could become better by closing smaller schools, increasing class size, introducing more testing, publishing results at school (and teacher) level, and basing teacher payments on achieved test results. The report ended with a clear warning: “Higher spending on schools will have no effect” (OECD, 2008).

The essence of this “expert advice” was that Norway should become a different kind of country—hardly objective, neutral, scientific advice. In fact, Norway is not the only country to receive this kind of advice. PISA creates country-specific reports for all participating OECD countries.

National Policies Based on PISA: Examples

Though the attention given to PISA results in national media varies between countries, in most countries it is formidable and has increased after several rounds of PISA testing (Breakspear, 2012). In Norway, the results from both PISA 2000 and PISA 2003 made the newspaper headlines. Our then Minister of Education Kristin Clemet (2001–2005), representing the conservative party Høyre, commented on the PISA 2000 results, which were released a few months after she had taken office: “Norway is a school loser, now it is well documented. It is like coming home from the Winter Olympics without a gold medal” (which, of course, for Norway would have been a most unthinkable disaster!). She even added, “And this time we cannot even claim that the Finnish participants have been doped!”—a reference to a recent cross-country championship (*Aftenposten*, January 2001, as translated by author).

The warlike headlines in all the Norwegian newspapers told residents that “Norway is a school loser.” The headlines, however, were misleading: Norway’s scores were actually close to the OECD average in the three test domains in PISA 2000 and PISA 2003. With the media’s help, however, the results from PISA shaped the public image of the quality of our school system, not only in regard to the aspects that had in fact been studied, but also for more or less all other aspects of school. It became common wisdom that Norwegian schools in general were poor quality and that Norwegian classrooms were among the noisiest in the world, although noise levels had barely been studied in PISA. The media presented tabloid-like and oversimplified rankings. Unfortunately, the public as well as politicians accepted these skewed generalities as objective scientific truths about our education system. There was little critical public debate, and the researchers behind the PISA study did nothing to modify the false impression and remind the public about the limitations of the study and the fact that Norway in reality ranked among the middle of the OECD countries. In sum, PISA created a public image of the quality of Norwegian schooling that was unjustified and detrimental to the long-term health of the education system and the country itself. Surveys among Norwegian teachers have since shown that they consider the effects of the PISA project as a serious detriment to their their daily work.

PISA has not only shaped the public image of Norwegian schools, but has also served to legitimize potentially harmful school reforms. Under Minister of Education Clemet, a series of educational reforms was introduced in Norway. Most of those reforms referenced international testing, mainly PISA, as

justification. In 2005, there was a change in government, and shortly afterward Clemet's Secretary of State Helge Ole Bergesen (2006) published a book in which he presented the "inside story" on the reforms made while they were in power. A key feature of the book was the many references to large-scale achievement studies. Bergesen (2006) explicitly stated that these studies provided the main arguments and rationale for curricular as well as other school reforms. Under the heading "The PISA Shock," he confirmed the key role of PISA:

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 the political power in the Ministry of Education and Research, the PISA results provided a "flying start."

(Bergesen, 2006, pp. 41–42, author's translation)

In these memoirs, Bergesen (2006) also described how the Ministry deliberately created an atmosphere of panic and urgency, and how all resistance was successfully characterized as unscientific and ideological.

When the next PISA round showed a small fall in Norway's test scores, the ground was prepared for a major school reform, to be called The Knowledge Promotion. All political parties were on board. Later, in the 2009 parliamentary election campaign, the prime ministerial candidate (who was to be Prime Minister for the same party from 2013) had the following main message, even with a personal signature in the leading Norwegian newspaper:

I, Erna Solberg, herewith guarantee that if we [i.e., Høyre, the moderate/conservative party] form the Government after the election, we can promise more PISA points.

(Aftenposten, March 27, 2009, author's translation)

It is most interesting that this statement was made shortly after the Norwegian Parliament unanimously passed a new law setting out the foundational values for Norwegian schools. During the election campaign, no reference was made to the key words in this law (e.g., equity, solidarity, empathy, concern for the environment). It is also notable that the red/green Labour-dominated Norwegian government that took office in 2005 followed more or less the same policy.

In the many white papers on schools that were presented from the government to the parliament in the years 2001 to 2013, the reference to the PISA project and "OECD experts" steadily increased, and by 2013 they appeared on nearly every page. There is no doubt that the major reforms of Norwegian schools over the last decade have been strongly influenced by the OECD, with PISA as the main instrument. These reforms are also characterized as examples of New

Public Management (Møller & Skedsmo, 2013). Among many reforms, we have established a high-stakes national testing system whose categories are more or less adapted from the PISA framework. The new national curriculum is also strongly influenced by the language and categories in the PISA framework.

In sum, international rankings, in particular PISA, are seen as the ultimate measure of the total quality of the Norwegian school system, and new reforms were introduced in response to the perceived challenge. Most reforms on curriculum, national testing, accountability, and transparency also follow the policy advice that emerged from the PISA studies. More private schooling, free choice of schools, and the growth of private consultancies for testing and reporting also are current trends that respond to the panic in the wake of PISA rankings.

In the autumn of 2014, Norwegian teachers went on strike—not for higher salaries, but in response to more external control of their working hours and other working conditions. The underlying cause of the conflict was the growing demand for more testing, reporting, and control of teachers' work, which had been triggered by the “PISA shock.” Similar developments have occurred in other countries, including our neighbouring countries Denmark and Sweden. It is noteworthy, however, that all these “solutions” to the perceived crisis are more or less the opposite of what our Nordic neighbour and declared “PISA winner,” Finland, is doing.

Germany had a similar PISA shock. The results from the first PISA round placed Germany below the OECD average. This became an important issue in the German election the following year, and the perceived bad results also led to major initiatives to promote the quality of science and mathematics education. The German national institute for science education, IPN (Institut für die Pädagogik der Naturwissenschaften), received large grants to improve the quality of science education. IPN also had the contract to run PISA in Germany. From the perspective of science education, one may say that bad results are good news, much the same way that the Sputnik shock was good news for science and mathematics educators in the Western world. The influence of PISA in Germany went even further, with the country's later introduction of national standards for science education (Steffen & Höble, 2014).

The OECD readily boasts about the influence of PISA. An official working paper reviewed the policy impacts and the normative effects of PISA. With obvious pride, the report stated:

PISA has been adopted as an almost global standard, and is now used in over 65 countries and economies. . . . PISA has become accepted as a reliable instrument for benchmarking student performance worldwide, and [the] PISA results have had an influence on policy reform in the majority of participating countries/economies.

(Breakspear, 2012, pp. 4–5)

This report reviewed literature as well as results from the OECD's own questionnaires, and provided a ranking of the impact that PISA had on all OECD countries. The report noted that even "high-performing countries such as Korea and Japan have enacted reforms in response to a large range of PISA results" (Breakspear, 2012, p. 12).

Interestingly, the United States ranks 28th on the PISA listing based on the scoring of "informing policy-making process," with PISA's influence classified as "moderate" (Breakspear, 2012, p. 14). Rutkowski (2014) argued that PISA's rather limited impact on American schools may be the main reason why the federal government and OECD are eager to introduce PISA for school testing in the United States. That move could get PISA closer to the decision-makers in the U.S. education system and thereby increase its normative power.

PISA, Free-market Thinking, and Globalization

My point so far has been to argue that PISA should be seen and understood in a wider political context. The two key elements here are free-market thinking and globalization. In regard to free-market thinking, the PISA project, organized by the OECD, can be seen as part of a current international policy trend, where concepts and ideas from the market economy are used in the education sector. The term New Public Management is used to describe this market-driven philosophy, which is supposed to make the public sector more efficient. Words like *quality*, *efficiency*, *transparency*, *accountability*, and *value for money* are among those used to describe these policy reforms in many public sectors. Public services such as schools and higher education, culture, and healthcare are all being invaded by market terms.

Other public sectors are experiencing the same trend. Services such as policing, security, postal delivery, transportation, water supply, household garbage handling, and sewage and waste water management all come under attack in the name of efficiency and value. Traditional public services are increasingly subjected to competitive bids from private actors. Outsourcing of key public services, a process that is eased by new regulations on international trade, is increasingly going to multinational companies. Most major international trade agreements now include provisions for privatizing public sectors. This trend toward marketization and privatization characterizes developments in several countries. And the education sector is at the forefront, with the OECD and its PISA project as an efficient tool (Meyer & Benavot, 2013).

The other, and related, political/economical perspective that pervades PISA is that of globalization. The economy is becoming globalized, large multinational corporations are important actors, and the workforce has to be flexible and moveable. Nations and multinational corporations compete in a common market. Hence, the thinking goes, there is a need for common standards in education, common systems for exams, degrees, and qualifications. Such

tendencies to standardize education processes operate within units such as the European Union, an example being the Bologna process and its introduction of a common degree system in higher education. In key areas, the OECD is playing an increasingly important role by developing and monitoring common standards, indicators, and measures (Grek, 2009).

This PISA-inspired process represents a political pressure to standardize, harmonize, and universalize national institutions such as a country's school system, and to promote competition on the global educational scene. While most science educators argue for context-based teaching and localized curricula, the pressure from the PISA project is in the opposite direction. A driving force behind these reforms is often the use of indicators—quantifiable and measurable standards that can be used for calculations (Popkewitz, 2011). PISA test scores and rankings are ideal for this purpose, whether or not that usage was intended by the PISA researchers.

Universally Valid “Real-Life” Indicators?

A fundamental premise for the PISA project is that it is indeed possible to “measure” the quality of a country's education using indicators that are common to all countries despite differences in things that affect school outcomes such as social structure, traditions, culture, natural conditions, ways of living, and access to free public education. As noted, PISA claims to measure how well the young generation is prepared to meet the challenges of tomorrow's world (OECD, 2007). Such an ambition assumes that the challenges of tomorrow's world are already known and more or less identical for young people across countries and cultures. It assumes one universal conception of success and a universal set of necessary skills.

Although life in many countries does have some similarities, one can hardly assume that the 15-year-olds in the United States, Japan, Turkey, Mexico, and Norway are preparing for the same challenges, careers, and economies, or that they need identical life skills and competencies. It is also important to remind ourselves that the PISA academic framework and its test are meant for the relatively rich and modernized OECD countries. When this instrument is used as a benchmark standard in the 30+ non-OECD member countries that take part in PISA, the mismatch between the PISA test and the needs of the youth in those nations becomes even more obvious.

One should also remember that the target population of the PISA testing is the whole age cohort of 15-year-olds. This is, in most countries, toward the end of what in most countries is comprehensive school. The great majority of these young people have to face realities that are local and national. Only a minority of them will operate in a global, international marketplace.

All countries have their own school and education systems based on national decisions set most often by democratically elected governments and institutions.

National traditions and deliberations have resulted in foundational legal statements about the overall purposes of the school, the level or levels of government with jurisdiction and influence over public schooling, and more concrete details such as time allocations for school subjects, aims, objectives and curricula, and exam structure. These traditions are often at the heart of the nation's identity, and the set of such laws and regulations is the mandate that society has given to the schools, the teachers, and all who work to improve the quality of a nation's schools.

OECD officials, however, explicitly claim that the PISA academic framework does not relate to any national school system because what they measure does not fit any one country's school system. In reality, the framework is made up of sets of perceived universal, presumably culture-free, curricula as decided by the OECD and its experts. The rather explicit goal of the OECD with its PISA project is to be an agent of change in the education system in the participating countries. In this respect, one may say that PISA is a success story (Lawn & Grek, 2012). The international test movement, in particular PISA, leads to policy convergence across nations. However, this goal is detrimental to the expression of educational values within countries and to cultural identity and diversity in general.

Steps toward the PISA Test

The process of developing the PISA ambitions into the actual test items the students get has several stages, each with serious obstacles that require many decisions. The first step from intention to test is the selection of the knowledge domains (or school subjects) that should be included. The OECD chose three domains (or "literacies") for the PISA testing: reading (in the mother tongue), mathematics, and science. Though these are important and basic subjects, one should keep in mind that most subject-area domains are not included in PISA, nor are the three domains in their entirety.

Of course, one test, even a test like PISA, cannot assess all possible school subjects; but by selecting some subjects and ignoring others, PISA implicitly sends a message to the public as well as politicians about what is important for schools and future life. The actual choice of reading, science, and mathematics also reflects the basic purpose of OECD: the concern for economic competitiveness in a global, high-tech market economy. As mentioned, when PISA extended its repertoire in 2012, the added domain was "financial literacy" (OECD, 2013a).

The PISA Framework

The next step in the process of developing the actual PISA test is to create a testing framework, in reality a curriculum. Here is where the measurement and

curriculum experts come in. The key institutions that win the competitive bid, and the selected subject-matter specialists, are in charge of a lengthy process to develop the framework. The people selected for this purpose are well-known international experts in their fields, often among the most respected and accomplished in the world. But, of course, they work within the frames decided by PISA as a project, and they must all be fluent in English, which is the dominating language in all deliberations and working documents. In addition to the subject-matter specialists, psychometricians play a key role in the whole process.

Most scholars will probably find the PISA frameworks developed by these expert groups to be most interesting, with ideas, perspectives, and subject-matter detail that is of very high quality (e.g., OECD, 2013a). Rather than models to be copied, these documents could be used as sources for inspiration to make national curricula and to stimulate the debate over educational priorities. The problem is, however, that this framework now serves as a normative international, universal curriculum, and a framework for an international testing regime.

Item Selection and Test Construction

The next step is to “operationalize” the frameworks—that is, to use the frameworks for the development and selection of test items, and for the construction of the PISA tests as a whole. For more detail on the technicalities in this complicated process, readers are encouraged to access the more than 400-page technical reports for each test (e.g., OECD, 2009a, for the PISA 2006 testing). However, some elements in the process are the following. Each PISA country (OECD countries only) is invited to submit test items that fit the framework and are based on authentic texts for real life situations. Through a complicated process involving initial screening and selection, national and international piloting, prefield trials, main field trial round, and psychometric analysis that involve many actors and subcommittees and many meetings for negotiations and debate, the final series of test items is decided. The complexity of this one stage in the process is apparent from the following extract from the technical report:

These analyses . . . included the standard ConQuest® item analysis (item fit, item discrimination, item difficulty, distracter analysis, mean ability, and point-biserial correlations by coding category, item omission rates, and so on), as well as analyses of gender-by-item interactions and item-by-country interactions. On the basis of these critical measurement statistics, about 40 new items were removed from the pool of items that would be considered for the main study.

(OECD 2009a, p. 41)

A logical consequence of wanting to make a fair international test is that an item cannot be used if it behaves in an unfair fashion. While this is a sensible argument from a statistical, psychometric point of view, it also means that items close to the real-life contexts of some countries but not others have to be removed. The principles for exclusions are described as follows:

The main reasons for assessing units as unsuitable were lack of context, inappropriate context, cultural bias, curriculum dependence, just school science and including content that was deemed to be too advanced.

(OECD, 2009a, p. 34)

This section of the technical manual clearly states that units (items) that relate to issues considered “inappropriate” (controversial in a particular country), have a “cultural bias” (positive or negative), or are close to the school curriculum (in some countries but not in others) were excluded. The statement also explicitly states that “just school science” should be excluded. This is, again, a clear statement that PISA does not measure knowledge or issues related to school curricula. Based on these criteria, it seems somewhat strange that such a test is used to judge the quality of science taught at school in each country. In the final science literacy test, for example, Norwegian students will find nothing about the key elements of the Norwegian economy. They will not find questions on topics such as oil and gas in arctic conditions on the continental shelf, aquaculture and fish farming, and hydroelectric power plants. Neither will they find anything about current topical issues and conflicts regarding the conservation of nature, current political conflicts between nature conservation (e.g., wild wolves) and sheep farming, snow, skiing or skating, the Northern Lights (a main focus of research of the university in Tromsø), or the challenges of an arctic climate. Students in other countries, of course, are also not likely to find questions relating to their own culture, nature, history, or current national challenges.

In reality, the test items in the final science test are decontextualized, or the context is contrived or historical. This cleansing of culture and context does not occur because of nefarious intentions built into the testing framework, but because of statistical necessity and concern for “fairness.” The decontextualized and contrived nature of the assessments runs contrary to all recommendations by science educators as well as by many national goals of promoting a science curriculum that is relevant, interesting, and context-based.

Item Texts, Language, and Translations

A further set of complications arises in relation to item texts, language, and translation. Most PISA items are based on rather lengthy texts that constitute the stem, called the “stimulus.” The intention is positive: namely, to present real,

authentic texts and real-life situations. But this contrived realistic format—in particular, the length and complication of the stimulus text—also makes the PISA items different from most tests that are commonly used in mathematics and science (also in TIMSS, the other large-scale study of science and mathematics achievement). This format is, of course, a deliberate choice by PISA specialists, and it reaffirms that PISA does not really test subject-matter school knowledge from school curricula.

It is often claimed that many PISA items are to a large degree testing reading skills rather than science and mathematics competencies. The strong correlations between the test results on the reading, mathematics, and science assessments lend support to such a claim. The fact that PISA items in later PISA versions have become shorter may indicate that this critique has been taken seriously and has led to a reduction in the heavy load on reading skills.

A robust finding in PISA, as well as other kinds of reading tests such as PIRLS (Progress in International Reading Literacy Study), is that girls outperform boys in reading in all countries. More surprising is that the gender difference in the PISA science and mathematics scores favors girls more than most other kinds of tests. This unusual gender pattern may, at least partially, be explained by the heavy reading load in many PISA items and the strong correlation between reading achievement on PISA and achievement on the mathematics and science sections. PISA test scores show a gender pattern in science and mathematics that is different from the TIMSS results in many of the same countries, as well as other tests such as national exams. It is also interesting that the PISA gender pattern differs when the students answer questions on a computer-based questionnaire, as they do in the so-called Computer-Based Assessment in Science (CBAS) version. In the computer-based test, the boys actually outperform the girls in science (OECD, 2010b). This is an indication that the context and the mode of data collection also influence the results to a significant degree, which is important to recognize as PISA migrates to an entirely computer-based format.

The authentic texts that constitute the stimulus in each item originate in a certain situation in one of the OECD countries and in the language of that country. This text is then translated into the two official PISA languages before being submitted for consideration. If accepted, the item is then translated into the language of each of the participating PISA countries. This translation process follows very strict rules and detailed instructions (e.g., OECD, 2009a).

This translation process raises many questions. Thorough analysis of the PISA reading test items has been carried out by Arffman (2007) and discussed in journal papers. Arffman (2010) provided a detailed text-analytical study of the translation from English to Finnish of three PISA items as an example of how both meaning and context change in the translation. Her study reveals in detail many critical dimensions in this process. One of her conclusions, based on translation theory and on a review of results from earlier empirical studies,

is that one can never arrive at what may be called “equivalence of translation.” For example, neither poetry nor good prose can be translated according to a formalized set of rules—a fact that all good translators will acknowledge. Something is always lost, or meaning is modified in translation.

Yet even where the quality rules should have been followed, seemingly nonstandard translations appear. There seems to be a lack of empirical studies looking into this important aspect of PISA (and also TIMSS and PIRLS) testing. The key role played by these texts in PISA makes such scrutiny very important.

A thorough cross-national check of translations requires the cooperation of researchers from many countries with considerable linguistic skills as well as subject-matter knowledge. Some languages, however, lend themselves to rather easy comparisons, even for nonlinguists. The three Scandinavian languages Swedish, Danish, and Norwegian, for example, are very similar—more like dialects, in fact, in part with a common literary tradition. A simple demonstration uses the translation of a single item about cloning from English into these three languages. Since its release in the 2006 test, this item about cloning the sheep Dolly has become a well-known example (OECD, 2009b). The stem text of the original, in English, is reproduced in Figure 5.1.

Based on this English (and the French) original, the three Scandinavian texts (now available from the national PISA website) were translated, presumably according to the detailed rules and instructions given by PISA. The most striking and immediate observation is that the three Scandinavian texts become strange and clumsy. Equally important is the fact that the resulting three versions are quite different from one another, and they have all changed the original meaning in some dramatic ways:

- The Swedish, Danish, and Norwegian texts changed the word “nucleus” to “cell nucleus,” thereby providing the hint that the “small piece” in Question 2 is indeed a cell.
- While the English (and Swedish) texts stated that he removed “the material that would have determined sheep 2 characteristics,” the Danish text stated, “he removed the genetic material,” thereby changing the meaning in the sentence as well as introducing a science concept that does not appear in the original.
- In the Norwegian version, “all material is removed from the egg-cell” makes the sentence more or less meaningless.
- The Danish text altered Question 1 and asked, “Which sheep is Dolly a copy of?” In Danish, the word “identical” was considered problematic, which is indeed true. The Danish version is also more in line with the title of the item: “A copying machine for living things.” This way of talking and writing about cloning is actually never used in any Nordic language, and probably not in other languages either.

S128: Cloning

Read the newspaper article and answer the questions that follow.

A copying machine for living beings?	
<p>Without any doubt, if there had been elections for the animal of the year 1997, Dolly would have been the winner! Dolly is a Scottish sheep that you see in the 5 photo. But Dolly is not just a simple sheep. She is a clone of another sheep. A clone means: a copy. Cloning means copying 'from a single master copy'. Scientists succeeded in creating a sheep (Dolly) that 10 is identical to a sheep that functioned as a 'master copy'.</p> <p>It was the Scottish scientist Ian Wilmut who designed the 'copying machine' for sheep. He took a very small piece from the 15 udder of an adult sheep (sheep 1).</p>	<p>From that small piece he removed the nucleus, then he transferred the nucleus into the egg-cell of another (female) sheep (sheep 2). But first he removed from that 20 egg-cell all the material that would have determined sheep 2 characteristics in a lamb produced from that egg-cell. Ian Wilmut implanted the manipulated egg-cell of sheep 2 into yet another (female) 25 sheep (sheep 3). Sheep 3 became pregnant and had a lamb: Dolly.</p> <p>Some scientists think that within a few years it will be possible to clone people as well. But many governments have already 30 decided to forbid cloning of people by law.</p>

Question 1: CLONING

S128Q01

Which sheep is Dolly identical to?

- A Sheep 1
- B Sheep 2
- C Sheep 3
- D Dolly's father

Question 2: CLONING

S128Q02

In line 14 the part of the udder that was used is described as "a very small piece". From the article text you can work out what is meant by "a very small piece".

That "very small piece" is

- A a cell.
- B a gene.
- C a cell nucleus.
- D a chromosome.

FIGURE 5.1 The original English text and two questions for the item "Cloning," reproduced exactly as they appeared in the student's questionnaire (OECD, 2009, pp. 197–198).

PISA technical reports assert to readers that it uses a top-quality translation process:

As in PISA 2003, one of the most important quality control procedures implemented to ensure high quality standards in the translated assessment materials consisted in having an independent team of expert verifiers, appointed and trained by the consortium, verify each national version against the English and French source versions.

(OECD, 2009a, p. 91)

The procedure for the translation control is then described in detail. The "translation equivalence across PISA countries" is also thoroughly discussed in Grisay, de Jong, Gebhardt, Berezner, and Halleux-Monseur (2007). In light of

this objective, it is rather surprising that big blunders can be discovered with just a cursory look at published items.

Even a hasty reading by nonexperts shows that the translated texts are put in a strange and awkward prose that one cannot find in any Scandinavian publications. Such texts cannot possibly be called “authentic.” Arffman (2010) noted that bad translations also may cause readers to lose interest and motivation to become engaged with the text, and that this may have a severely negative effect on the test results. This effect, I assert, is likely to be greater in countries where students are critical, independent, and unwilling to obey the authority of schools and teachers. This point about students’ motivation and willingness to engage in the whole exercise is elaborated elsewhere (Sjøberg, 2007).

Written Test as “Real Life” Situation?

The basic claims of the OECD as listed in its technical documents are that the PISA test results can provide reliable evidence about (a) how well young people are prepared for future challenges, (b) whether they can analyze, reason, and communicate effectively, (c) whether they have the capacity to continue learning throughout life, and (d) to what extent they have acquired some of the knowledge and skills essential for full participation in society. These ambitions are great, but they are contradicted by the very format of the testing: The PISA test is a traditional pen-and-paper test, where students sit for 2.5 hours to answer written questions, in solitude, and without access to sources of information. How “real life” is that test situation? How does it relate to the challenges that young people may face in their future life as citizens, as participants in tomorrow’s democracy, and as members of a skilled workforce? The fact is that the PISA test situation does not resemble any real life, problem-based situations. The only place where you sit in solitude with a written test is in exams at schools and universities. The only places where students are not allowed to communicate, collaborate, or use modern information technologies are similar contrived test situations.

Real life, in private, at leisure, as well as at the workplace, is more or less the opposite of the PISA test situation. While one should expect that an organization like the OECD should emphasize the competencies needed by the big international actors on a competitive global market, the PISA test situation is different. Therefore, PISA does not even serve the political/economic goals of the OECD.

Test Scores and Economic Prosperity

It does sound like “common sense” that high scores on science and math tests at school are good predictors of future economic prosperity. The assumed strong connection between scores on tests like TIMSS and PISA and the economic competitiveness of the country is a core assumption behind these studies. As

noted earlier in this chapter, bad rankings on PISA are thought to be bad signals for the future of a country. This assumption is probably the main reason for the extreme importance given to PISA results and rankings. PISA is, in fact, presented, marketed, and understood in this perspective, as also noted earlier.

But this commonsensical assumption may now be studied empirically. In January 2013, *New Scientist* published an article titled “West vs Asia education rankings are misleading,” claiming:

For developed nations, there is scant evidence that TIMSS rankings correlate with measures of prosperity or future success. The same holds for a similar test, the Programme for International Student Achievement (PISA). . . . Analysis of 23 countries found a significant negative relationship between 2009 PISA scores and ranking on the Global Entrepreneurship Monitor’s measure of perceived entrepreneurial capabilities.

(Campbell, 2013)

Among the references in this article is a study by Tienken (2008), who used a series of indicators for countries’ economic competitiveness and prosperity, and looked at how these correlate with scores on international tests of TIMSS (using data since the early 1960s) and PISA (since 2000). Such studies are most interesting, because they undermine the fundamental premise behind PISA. Further studies of the connections between national test scores and economic prosperity are likely to shed more light on this important issue. Such statistical studies may be supplemented with studies undertaken by the actors on the market. Two examples follow.

Competencies for the Future: The Voice of Employers

Many sources provide qualified accounts of the skills and competencies that large employers in the high-tech sector require from their workforces. At the Official Bologna Seminar in 2008 on “Employability: The Employers’ Perspective and Its Implications,” Frank Stefan Becker (2008), head of Human Resources at Siemens, gave a presentation of his company’s view regarding the competencies it needs from its employees. He presented the following list:

Professional Competence—Vital Skills for Today’s Employees

- Thorough knowledge of one’s subject;
- Ability to judge analytically, structure one’s work, make “plausibility checks,” carry out research, evaluate information, and identify problems;
- Ability to look beyond one’s own area of competence and take other factors into account;

- Independence, initiative, independent learning, work techniques, discipline, frustration tolerance, ability to set priorities;
- Interpersonal skills: communication, feedback, a feeling for situations, capacity for teamwork, fluent English.

Siemens AG is the largest Europe-based electronics and electrical engineering company. Siemens and its subsidiaries employ approximately 360,000 people across nearly 190 countries. One may easily see that most of the competencies on the preceding list are not addressed on the PISA test.

The second example is an investigation done by Abelia, the Business Association of Norwegian knowledge- and technology-based enterprises. Based on a survey among 500 leaders in the most competitive sector of the Norwegian economy, the association ended up with the following ranking of competencies:

Competencies for Future Leaders and Key Personnel

- Good communication skills;
- Aptitude for strategic thinking;
- Ability to motivate;
- Concern for staff and coworkers;
- Self-confidence and self-efficacy;
- Solid educational background;
- Visionary;
- Understanding numbers and quantities.

(DAMVAD Group, 2013, Author's translation)

As one can readily see, the two examples emphasize similar competencies, but they are quite different from what PISA is testing. Interestingly, the perspectives expressed from high-tech industry in many ways coincide with the purposes and aims of schooling in many modern democracies, and they also are in line with many aspects of “progressive pedagogy.” Advice based on PISA results may, in fact, be counterproductive even for companies that operate in the competitive global market.

Problematic Statistics and Lack of Transparency

The PISA project is a large undertaking that has many of the characteristics of “big science” and “techno-science.” It is costly and involves the cooperation of approximately 70 countries. The logistics of the project are complicated, as specified in piles of documents with detailed instructions to the national groups who are responsible in the participating countries. Hundreds of experts from several fields of expertise are involved. Contracts with subcontractors are given

by bids. Thousands of schools and teachers participate, with nearly half a million students spending 2.5 hours answering the test and the questionnaire. Data are carefully coded by thousands of specially trained markers. And the list goes on.

Some of the many problematic issues in the process from intentions to test items have been raised earlier, but there are more issues that are problematic. The final test consists of items that are selected according to a process previously described, but the actual test booklets and questions that are answered by the students are not identical. A system of “rotation” of items means that the students answer several different booklets. In this way, PISA can include a larger number of items in its test. After the time-consuming and tedious coding and data-entry process, the data undergo complicated statistical analysis. The statistical processes that lead from actual responses to these numbers is based on Item Response Theory (IRT) and Rasch modeling. Moreover, the final overall scores are normalized to provide an international mean score of 500 with a standard deviation of 100 for the OECD as a whole.

The road from the actual responses to the final numbers and rankings in the publicly available tables is long and not very transparent. The methods have been criticized by well-qualified statisticians, even among those who actually work on PISA data. Svend Kreiner, professor of biomedical statistics at Copenhagen University, argued that by simply changing some of the parameters in the complicated statistical analysis, Denmark’s rank can fluctuate from as high as number 2 to as low as number 42. He also noted that the PISA methods of statistical calculations are published only in a general form, making detailed critique and replication difficult (Kreiner & Christensen, 2014).

Problematic and Intriguing Results

The political/economic aspects of PISA and the overall basic weaknesses of the project, as described, are primary concerns. But other serious concerns also should be addressed, especially by those who embrace PISA and believe that PISA provides valid data on educational quality and student achievement. The following is an overview of some of these concerns.

Resources and Finance Have No Influence?

From the beginning of the PISA project, the OECD has produced graphs and indicators that have shown small or negligible correlations between a country’s PISA scores and its spending on education (OECD, 2001). This, of course, has led to the OECD advice that more spending on education will not improve its quality.

In the five Nordic countries, the relationship between public spending and PISA scores is actually strongly negative. Such findings are often used to the detriment of schools and teachers. Finland, a country that ranks high in all areas of PISA, has one of the lowest levels of spending. This example of an inverse correlation between public spending and PISA points is used by political actors in the ongoing debate about return on public investment. Finnish teachers have difficulties asking for higher salaries and more funding because the Finnish scores already rank so high and hence no changes need to be made. Norway, on the other hand, is lower in the PISA rankings, but with higher amounts of public spending on schools. Based on PISA, Norwegian politicians have argued that it has been demonstrated that more spending would not increase the quality of schools. As noted earlier, the OECD (2008) economic report to Norway actually warned Norway not to increase spending on schools, emphasizing that “this will not improve the quality.”

High PISA Science Scores Bring Lower Interest and Negative Attitudes?

PISA scores are often presented as league rankings between countries, with the winners on top and the losers at the bottom. But PISA tests also include a questionnaire with many questions about attitudinal aspects of how young people relate to science and the other tested subjects. This was an important element of the PISA 2006 study, when science was the core subject. The definition of *science literacy* in PISA 2006 actually included “willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen” (OECD, 2006a, p. 22). The indices and constructs that were developed for this broad category, however, were not included in the PISA scores that were used for rankings.

A special issue of the *International Journal of Science Education* (2011) presented several interesting results from analyses based on these data. The simplest and possibly most surprising finding was that many countries with the highest mean PISA science scores were at the bottom of the list rating students’ interest in science (Bybee & McCrae, 2011). Finland and Japan were the prime examples. Both sat atop the PISA rankings for science scores, yet were at the very bottom on constructs like student “interest in science,” “future-oriented motivation to learn science,” as well as on “future science job,” that is, inclination to see themselves as scientists in future studies and careers. In fact, the PISA science score correlated negatively with “future science orientation” ($r = -0.83$) and with “future science job” ($r = -0.53$), as Kjærnsli & Lie (2011) reported. These negative relationships are based on countries as the units of analysis. When individual students within each country are the units, some of the correlations are positive. Making unjust statistical inferences from groups to individuals is labeled “ecological fallacy.”

Such findings are most disturbing for those interested in having more youth pursue careers in science. If the students in top-ranking PISA countries leave compulsory school with strong negative orientations toward science, one needs to step back and think about the reasons for this as well as the possible consequences. Of course, care should be taken not to interpret correlation as cause and effect, but one should at least think twice before using these countries as educational models and ideals to be copied if they are producing negative attitudes toward an important subject like science.

Interestingly, many of the winners in the PISA science test also have the largest gender differences in PISA scores. Finland is again a prime example, where girls outperform boys on all three PISA subjects. In reading literacy, the difference in means is about 50% of a standard deviation. Again, such findings should call for some caution against trying to copy the “PISA winners.” What are these winners actually producing? It seems at least in science, the winners are producing inequity in science achievement with overall negative attitudes toward the subject. Is that something to be celebrated or copied?

Traditional Teaching Equates to Better Results?

The PISA student questionnaire includes a series of questions about family background, home environment, and cultural artifacts. It also contains a series of questions to students about the teaching methods and classroom practices that characterize their school experiences. When looking for possible relationships between these variables and the PISA scores, many of the results are surprising and should be given attention. The most intriguing aspect of the results is that they run contrary to current advice from science educators as well as “accepted wisdom” among policymakers and curriculum specialists on what constitutes good instruction. The following is a brief indication of some problematic results.

A trend in current science education is emphasis on active learning and inquiry. Such teaching methods are supported by panels of science education specialists (e.g., Osborne & Dillon, 2008) as well as OECD (2006b) experts and the influential European Union (European Commission, 2007). Policy statements made in these and other reports have channeled much of the research and development work supported by the European Union’s current Framework Programme FP7 into inquiry-based science education (IBSE).

Reference is often made to the French program *La main à la pâte* organized by the Academy of Sciences. The key person in the project is Pierre Léna, a well-known astrophysicist as well as former Director General of Higher Education and Research at the Ministry of Education. This French program was inspired by a U.S. program with hands-on science initiated by the Nobel Laureate Leon Lederman. Such projects are seen as sources of inspiration by science educators as well as by national policymakers. PISA scores, however,

seem to be negatively correlated with many active teaching methods like “formulating your own problems and finding answers” and doing experimental work (Kjærnsli, Lie, Olsen, & Roe, 2007).

Current science education trends and reforms are reviewed by Jenkins (2009). Key concepts and acronyms in current thinking in science education are well known: science in context, IBSE (inquiry-based science education), hands-on science, active learning, NOS (nature of science), SSI (socio-scientific issues), argumentation, STS (science, technology, and society). There seems to be no evidence from PISA to back up these recommended methods; PISA rather provides counterevidence. This possible contradiction should at least be seen as problematic when interpreting the value of high PISA scores in science.

The Use of ICT Leads to Lower Scores

PISA includes several questions regarding the use of information and communication technologies (ICT) in schools and has made two constructs based on them. One construct or index is related to the use of the Internet in schools; the other is related to the use of software and educational programs. In a detailed study of the five Nordic countries, Kjærnsli et al. (2007) documented a clear negative relationship between the use of ICT and PISA scores. The PISA winner, Finland, has by far the lowest use of ICT of any Nordic country, and is actually below the OECD average. In contrast, Norway is top of the OECD in all indicators on the use of ICT in schools, but has only average PISA scores. Nevertheless, the policy advice in Norway is to increase the use of ICT in schools to improve achievement.

Intriguing PISA Results: Concluding Remarks

Some of the problematic results described here are not difficult to understand. A written test like PISA can hardly measure the skills and knowledge acquired in experimental work in a lab or on an excursion; neither can it capture the kind of interest, curiosity, and enthusiasm that may be the result of argumentation, inquiry, and the search for solutions to questions that students have formulated themselves. If the final test of quality is a written test, it is no surprise that teaching will be more cost-effective if students do not spend time on excursions, experimental work, or discussion of socio-scientific issues.

The use of PISA data for policy recommendations is, at best, very selective. If one believes in the free-market mission of the OECD and the meaningfulness of the PISA results, then one has to take all the results seriously, including those that are counterintuitive and at odds with other research findings and policies recommended by educational experts. PISA enthusiasts cannot selectively pick results that support their political positions while completely ignoring data that disprove their positions.

Critique from Academics

Parallel to the increasing global influence of PISA on educational debate and policy, there has been a growing critique of the PISA project in the academic world. Several anthologies have raised serious questions about the meaningfulness of the results (Hopmann, Brinek, & Retzl, 2007; Meyer & Benavot, 2013; Pereyra, Kotthoff, & Cowan, 2011). The authors raising important questions represent a cross-section of thinkers and researchers who come from many countries and academic fields, and include well-known philosophers, sociologists, economists, and educators.

In May 2014, a group of these and other academics sent an open letter to Andreas Schleicher, head of PISA and Director for Education and Skills at the OECD. In the letter (Meyer & Zahedi, 2014), they voiced a series of concerns about the growing influence of PISA. They argued that PISA is killing the joy of learning and is detrimental to basic values for which schools should strive. This initiative received public attention through coverage in *The Guardian* and other news media worldwide. The open letter has been signed by more than 2,000 academics from about 40 countries. Behind the initiative are leading educators such as Stephen Ball, David Berliner, and Robin Alexander. Noam Chomsky is also behind this initiative; likewise Diane Ravitch, who was previously U.S. Assistant Secretary of Education. Ravitch is now, as a distinguished professor of history and philosophy of education, the most influential critic of market-driven education policies. She is author of several influential books, including *The Death and Life of the Great American School System* (Ravitch, 2011) with the telling subtitle *How Testing and Choice Are Undermining Education*.

It seems fair to say that criticism of the uses and misuses of PISA is now common among most academics concerned about schooling and education. Because PISA has been extending its scope and influence during the past few years, this critique has been fueled. Further analysis of this development lies outside the scope of this chapter, but here are some aspects of the development.

PISA has recently established a close cooperation with Pearson, the owner of the *Financial Times*, *The Economist*, Allyn & Bacon, and Prentice Hall. Pearson has expanded its activities into the education sector and has become the world's largest company for testing and education programs, with 40,000 employees in more than 80 countries. Approximately 80% of Pearson's revenues now come from what Pearson terms the "education sector." Pearson won the bid for important parts of the PISA 2015 testing and has developed strong links with the OECD. Pearson has, of course, a vested interest in creating a market for its services and products. Through its close partnership with the OECD, it created a strong position to expand its market share as well as its global influence. Ravitch (2012) expressed her concern about this influence: "Are we prepared

to hand over our children, our teachers, and our definition of knowledge to Pearson?”

Pearson has an even stronger grip over PISA 2018. The company won the bid for key elements of the entire undertaking. A joint press release (Pearson 2014) from OECD and Pearson proudly announced:

Pearson, the world's leading learning company, today announces that it has won a competitive tender by the Organisation for Economic Co-operation and Development (OECD) to develop the Frameworks for PISA 2018. . . . The frameworks define what will be measured in PISA 2018, how this will be reported and which approach will be chosen for the development of tests and questionnaires.

This key role in PISA does not, of course, mean that Pearson's staff will do the work. But they will organize and administer the process. Pearson will continue to forge personal links with countless academics in key positions and the countless representatives for national educational authorities. This contract is a most valuable investment for Pearson. The cooperation is already in place for several by-products, such as a video series about “Strong Performers and Successful Reformers in Education” (www.oecd.org/pisa/pisaproducts).

Andreas Schleicher is also on Pearson's Advisory Board. He is, among several roles, also heavily involved in the initiative called *The Learning Curve*, where the main “product” is a ranking of the quality of educational systems based on several data sources (e.g., PISA, TIMSS, PIRLS). This ranking list gets a lot of attention from the media as well as from politicians, who often panic when their country is lower than they expect or when they move down on the rankings.

PISA itself is also widening its repertoire. One new product, the PISA-based Test for Schools, is a complementary test used to assess local-level performance. The results of this local test are reported individually to the school or district, in contrast to the aggregate scores reported on the international PISA test. Individual schools or school districts may get “information and analyses . . . comparable to main PISA scales” (www.oecd.org/pisa/aboutpisa/pisa-basedtestforschools.htm). This, of course, may create an enormous market and also bring competition, testing, and rankings even closer to the daily activities in schools. The PISA influence on local education policy will also be stronger.

With its PISA for Development test product, the OECD is widening its influence over education in developing countries. Such a test will provide standardized definitions of worthwhile skills and knowledge that are measurable and common for developing countries, independent of their culture, economic contexts, traditions, natural resources, and local challenges. Such scores will be seen as “benchmarks” and objective measures of quality by both donors and national authorities. The OECD may thereby push aside the influence of

United Nations organizations like UNESCO and UNICEF, which may have different priorities than those of the OECD. PISA for Development has created close links with these organizations as well as with the World Bank.

While PISA has 15-year-olds as its target population, an emerging OECD project—Programme for the International Assessment of Adult Competencies (PIAAC)—has been labeled PISA for Adults. Results from the first full-scale data collection in 24 countries were published in 2013 (OECD, 2013b), and the program is likely to increase its importance for the higher education and training sector as well (www.oecd.org/site/piaac).

Conclusions

This chapter focused on the problematic sides of PISA, but the positive virtues of PISA should not be ignored. The PISA project has led to an increased interest in and concern for education and the competencies that young people need to develop to achieve the different “literacies” that are needed for their future lives as well as for the well-being of their societies. The data bank generated by successive rounds of PISA is tremendous, and is likely the largest and most professional data source in the history of social science and educational research. These data are also well-documented and available for research purposes.

International comparisons in education are important: they can introduce new perspectives, and they can provide inspiration and ideas for educators, researchers, and policymakers. However, international comparisons have a kind of Janus face: They can be understood and used in two opposite ways. Such studies may acknowledge and give cause to celebrate the great variety among youth, nations, and cultures on aspects of education and, as such, serve as a source of inspiration. But they can also be used normatively, creating pressure to oblige and fit to allegedly universal and common standards set by the authority of external specialists.

What we are seeing is a prime example of New Public Management (Møller & Skedsmo, 2013) as well as a kind of global governance and standardization of education (Ball, 2012; Rinne, 2008). As indicated earlier, academics from several disciplines have raised concerns about various aspects of the PISA undertaking and about the OECD acting like a global ministry of education. The open letter to the OECD is also a sign of a growing concern about how PISA is used to overrule national and local priorities in education.

The official intentions of PISA, as cited earlier in this chapter, can easily be endorsed. No one can disagree with the need to ascertain whether young people are developing the knowledge, skills, and competencies needed to face the challenges as citizens of the future. But the underlying economic and political ambitions behind the OECD-driven PISA project are often ignored or undercommunicated. Even researchers in the PISA project seem not to realize (or accept) the overall political/economic aspects of the project.

The inherent difficulties in evaluating what PISA claims to measure are seldom fully understood or considered. The road from ambitious intentions to the actual test instruments and valid data to policy proposals is long and murky. This chapter has pointed to some of the problematic issues in the process. In this chapter and elsewhere, I stated that it is not just problematic to live up to the intentions laid down in the overall statements of PISA: it is, in fact, “mission impossible” (Sjøberg, 2007).

The public, media, and policymakers, however, often take the PISA numbers and rankings as given facts. They trust that the PISA experts know what they are doing and that the numbers are objective, reliable, valid, and neutral measures of education quality. They trust that PISA scores measure the stated intentions.

No test is better than the items of which it consists. The secrecy over most PISA items that appear on tests makes critique and scrutiny from the academic community, and even the public, difficult. Many of the publicly published PISA item samples have faced serious critique for issues related to content, language, and relevance. Translations into the many different languages have only begun to be examined, but it is easy to find flaws and even substantive changes and mistranslations. More research is needed there. The problematic use of statistics and the lack of transparency must also be examined.

Similarly, there seems to be little attention to the fact that many of the results of PISA are at odds with what educators recommend, and with what politicians propose as interventions to improve the quality of schools. Many politicians want to copy the PISA winners, but they don't fully understand the consequences; in order to copy the winners, policymakers often prescribe measures that are the opposite of what these winners actually do. There is a need to seriously address these paradoxical results. If one really believes in PISA, one also has to accept and address the contradictions.

PISA has a profound influence on educational policy in many countries, and this is indeed the intention behind the project. It is, however, obvious that PISA results are used selectively, misused, and even distorted for political purposes in many countries. The reference to PISA to justify and legitimize educational reforms is widespread. This influence ought to be better researched and scrutinized. PISA is, in essence, part of a free-market political project—a perspective that often falls outside the agenda of the educational research community. The recent expansion of PISA into schools and school districts, adult education, and education in developing countries needs to be followed with great concern; likewise the close connection between OECD and global, commercial actors like Pearson.

Extensive resources are used to run the PISA project and to produce reports and publications, but critical research is scarce and not well-funded. A key aspect of the academic ethos is to provide a critical voice, and to question and challenge conventional wisdom. Given the great political and

educational importance of PISA, there is a strong need for critical and independent research.

So What? Implications for School Policy and Leadership

For educators and school leaders, in particular, it is of paramount importance to understand the forces acting on the educational scene—globally, nationally, and locally. With PISA as an instrument, the OECD has become a key actor in education at all these levels. The OECD's perspective is mainly reflected by the E in its acronym: Economic. The preparation of human resources to be active participants in the global economy is the organization's prime concern, and this is also how it presents its results. Worldwide, governments look to PISA results and rankings as objective indicators and predictors of a country's future competitiveness on the global scene. In this chapter, I have cast doubt on the soundness of such "common sense" interpretations of PISA data. I also have pointed to serious concerns about other aspects of the PISA test. School leaders should be aware of all these serious problems and the pitfalls of putting too much weight on PISA results.

The OECD boasts that PISA results have triggered educational policy reforms in nearly all participating countries (Breakspear, 2012). According to this analysis, the U.S. ranking on PISA has not yet triggered reforms. This is likely because of the decentralized structure of power in U.S. schools. However, as U.S. education policy becomes increasingly standardized and centralized, the influence of PISA will probably increase. Rutkowski (2014) considered this issue and asserted that the development of the PISA may strongly increase the normative influence of PISA/OECD on U.S. schools. When local schools are able to compare themselves with international PISA winners such as Shanghai or Finland, one may expect a race to climb on this indicator.

Rutkowski, Rutowski, and Plucker (2014/2015) addressed whether individual U.S. schools should participate in a PISA-like test. The authors provided detailed arguments against such participation, based on technical perspectives of the test as well as educational arguments. A key argument is that the PISA tests explicitly state that they do not address school curricula or school knowledge. As instruments to monitor how schools live up to the mandates from national or local authorities, PISA should by definition be ruled out. Schools are bound by state-mandated curricula in the U.S. I do not claim to know every detail of the inner workings of the U.S. education system, but I endorse the conclusion that the PISA-based Test for Schools program will serve no good purposes in U.S. schools or schools in other countries. The economic argument for not administering this test might also be added, although the main arguments are educational.

Not only in the U.S., but also worldwide, schools and the education arena in general have become open markets for corporations to snag large contracts

and large profits. In recent years, the testing industry has expanded and become global. Large sums of public money float into the pockets of these companies. Competition, privatization, and market-orientation are threatening the values that used to be the cornerstone of public education in many countries. School leaders and principals should take a stance in this battle over priorities and should not become passive recipients of new ideas disguised as a means to increase quality and efficiency.

What about participating in the PISA test itself? This international version of the test is administered in only a sample of schools, to only some students in each school. Whole classes do not participate—only a certain number (about 30) from the whole school, all at the age of 15. This, of course, adds to the logistical complexity for the schools taking part. The participating schools and students do not get anything of educational value in return. The test is anonymous for the students as well as for the schools. The test booklets are meticulously collected and removed when completed. Test items are treated as secret. There is no possibility of discussing the items for the purpose of learning from them. No results are given back to schools.

In some countries, schools have refused to take part in PISA, a decision usually made by the principal. This may also be an option for U.S. principals and school leaders.

Notes

- 1 Based in part on “PISA and global educational governance: A critique of the project, its uses and implications” by Svein Sjøberg. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(1), 111–127. Copyright © 2015 by International Society of Educational Research. Used by permission of the publisher.

References

- Alexander, R. (2012). Moral panic, miracle cures and educational policy: What can we really learn from international comparison? *Scottish Educational Review*, 44(1), 4–21.
- Alliance for Excellent Education. (2007). Losing our edge: Are American students unprepared for the global economy? PISA 2006 Release Conference, Washington, DC. Retrieved from www.all4ed.org/events/losingedge
- Arffman, I. (2007). *The problem of equivalence in translating texts in international reading literacy studies: A text analytic study of three English and Finnish texts used in the PISA 2000 reading test* (Research Reports 21). Jyväskylä, Finland: University of Jyväskylä, Finnish Institute for Educational Research.
- Arffman, I. (2010). Equivalence of translations in international reading literacy studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59.
- Ball, S. J. (2012). *Global education Inc.: New policy networks and the neo-liberal imaginary*. London, England: Routledge.
- Becker, F. S. (2008, November). *What companies need and how they can cooperate with universities to get it*. Presentation at Official Bologna Seminar, Luxembourg.

- Bergesen, H. O. (2006). *Kampen om kunnskapsskolen* [The fight for a knowledge-based school]. Oslo, Norway: Universitetsforlaget.
- Breakspear, S. (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, France: OECD Publications. doi:10.1787/5k9dfqffr28-en
- Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33(1), 7–26.
- Campbell, M. (2013). West vs Asia education rankings are misleading. *New Scientist*, 22–23.
- DAMVAD Group. (2013). *Integrating global talent in Norway: Survey report*. Oslo, Norway: Author.
- European Commission. (2007). *Science education now: A renewed pedagogy for the future of Europe* (Rocard report). Brussels, France: Author.
- Grek, S. (2009). Governing by numbers: The PISA “effect” in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Grisay A., de Jong, J. H., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8(3), 249–66.
- Hopmann, S. T., Brinek, G., & Retzl, M. (Eds.). (2007). *PISA zufolge PISA [PISA according to PISA]*. Berlin, Germany: LIT Verlag.
- International Journal of Science Education*. (2011). Special issue: Visual and spatial modes in science learning. Author: 33(3).
- Jenkins, E. W. (2009). Reforming school science education: A commentary on selected reports and policy documents. *Studies in Science Education*, 45(1), 65–92.
- Kjærnsli, M., & Lie, S. (2011). Students’ preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education*, 33(1), 121–144.
- Kjærnsli, M., Lie, S., Olsen, R. V., & Roe, A. (2007). Tid for tunge løft. Norske elevers kompetanse i naturfag, lesing og matematikk i PISA 2006 [Time for heavy lifting: Norwegian students’ competence in science, reading, and mathematics in PISA 2006]. Oslo, Norway: Universitetsforlaget.
- Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness: A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, 79(2), 210–231.
- Lawn, M., & Grek, S. (2012). *Europeanizing education: Governing a new policy space*. Providence, RI: Symposium Books.
- Loveless, T. (2009). *How well are American students learning?* [The 2008 Brown Center report on American education] (Vol. II, No. 3). Washington, DC: Brookings Institution.
- Lundgren, U. P. (2011). PISA as a political instrument: One history behind the formulating of the PISA programme. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 17–30). Rotterdam, Holland: Sense Publishers.
- Maslen, G. (2013, November 29). Andreas Schleicher and the PISA phenomenon. *The Sydney Morning Herald*.
- Meyer, H.-D., & Benavot, A. (Eds.). (2013). *PISA, power and policy: The emergence of global educational governance*. Oxford, England: Symposium Books.
- Meyer, H.-D., Zahedi, K., et al. (2014). Open letter to Andreas Schleicher, OECD, Paris [Blog post]. Retrieved from <https://oecdpsialetter.wordpress.com>

- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–353.
- Organisation for Economic Co-operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2006a). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2006b). *Evolution of student interest in science and technology studies* (Policy report from the Global Science Forum). Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world* (Vol. 1: Analysis). Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2008). *OECD economic surveys: Norway 2008*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2009a). *PISA 2006 technical report*. Paris, France: OECD Publications. Retrieved from www.oecd.org/pisa/pisaproducts/42025182.pdf
- Organisation for Economic Co-operation and Development. (2009b). Take the test: Sample questions from OECD's PISA assessments. Paris, France: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264050815-en>
- Organisation for Economic Co-operation and Development. (2010a). *PISA 2009 results: What makes a school successful? Resources, policies and practices* (Vol. IV). Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2010b). *PISA computer-based assessment of student skills in science*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2013b). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris, France: OECD Publications.
- Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Paris, France: OECD Publications.
- Osborne, J., & Dillon, J. (Eds.). (2008). *Science education in Europe: Critical reflections* (A report to the Nuffield Foundation). London, England: King's College.
- Pearson. (2014). Pearson to develop PISA 2018 student assessment 21st century frameworks for OECD [Press release]. Retrieved from www.prweb.com/releases/2014/12/prweb12386913.htm
- Pereyra, M. A., Kotthoff, H.-G., & Cowan, R. (Eds.). (2011). *PISA under examination: Changing knowledge, changing tests, and changing schools*. Rotterdam, Holland: Sense Publishers.
- Popkewitz, T. (2011). PISA: Numbers, standardizing conduct, and the alchemy of school subjects. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under*

- examination: Changing knowledge, changing tests, and changing schools* (pp. 31–46). Rotterdam, Holland: Sense Publishers
- Ravitch, D. (2011). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Ravitch, D. (2012, May 7). Pearson's expanding role in education. *The Washington Post*.
- Rinne, R. (2008). The growing supranational impacts of the OECD and the EU on national educational policies, and the case of Finland. *Policy Futures in Education*, 6(6), 665–680.
- Rutkowski, D. (2014). The OECD and the local: PISA-based test for schools in the USA [Online]. *Discourse: Studies in the cultural politics of education*. doi:10.1080/01596306.2014.943157
- Rutkowski, D., Rutkowski, L., & Plucker, J. A. (2014/2015). Should individual U.S. schools participate in PISA? *Phi Delta Kappan*, 96(4), 68–73.
- Sjöberg, S. (2007). PISA and “real life challenges”: Mission impossible? In S.T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA [PISA according to PISA]* (pp. 203–224). Berlin, Germany: LIT Verlag.
- Sjöberg, S. (2015). PISA and global educational governance: A critique of the project, its uses and implications. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(1), 111–127. doi:10.12973/eurasia.2015.1310a
- Steffen B., & Höble, C. (2014). Decision-making competence in biology education: Implementation into German curricula in relation to international approaches. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 343–355.
- Tienken, C. H. (2008). Rankings of international achievement test performance and economic strength: Correlation or conjecture? *International Journal of Education Policy and Leadership*, 3(4), 1–15.