

PISA and Global Educational Governance – A Critique of the Project, its Uses and Implications

Svein Sjøberg
University of Oslo, NORWAY

Received 16 October 2014; accepted 21 December 2014

The PISA project has steadily increased its influence on the educational discourse and educational policies in the now 70 participating countries. The educational debate has become global, and the race to improve PISA-rankings has become high priority in many countries. For governments the PISA-test is a high-stake test. Governments are blamed for low scores, and governments are quick to take the honour when results are improving. National curricula, values and priorities are pushed aside. The influence of PISA is of high importance also for science educators; the battle to improve test results may conflict with our work to make science relevant, interesting and motivating for the learners. This article tries to unpack some of the challenges and to raise some of the problems caused by PISA as a global measure of quality.

Keywords: comparative studies, educational policy, globalization, OECD, PISA-test, standardization

INTRODUCTION

From the mid 1990's, the OECD started the planning of Program for International Student Assessment, now well known as PISA. Since the first publication of PISA results in 2001, based on the testing in 2000, the results have become a kind of 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 debates on the quality of education. PISA league tables are widely published in mass media, and also used by politicians and educational authorities. In many countries, educational reforms are launched as direct responses to the PISA results.

The testing takes place every three years, and when results from PISA2012-testing were published in December 2013 (OECD, 2013b), we now have data

from five rounds of PISA. When the 2015-testing takes place, "Science Literacy" will for the second time be the core theme. Results will be released at the end of 2016.

The intentions of PISA are, not surprisingly, related to the overall political aims 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 has later been joined by a similar number of countries and "economies". When PISA is presented, its importance is stated like this: "PISA has participation from 90% of the global economy" (OECD, 2013b). For educators, this may seem a surprising way of counting pupils, but it indicates the focus of the PISA-project: economy. This may also account for the extreme importance that is now attributed PISA rankings: it seems "common sense" that high scores on reading, mathematics and science are predictors for the country's future economic competitiveness. Bad rankings on PISA are thought to be bad signals for the future of the country. We will return to this crucial point towards the end of this article.

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"

Correspondence to: : Svein Sjøberg, Science Education, Department of Teacher Education and School Research (ILS), University of Oslo, NORWAY.
Email: svein.sjoberg@ils.uio.no
doi: 10.12973/eurasia.2015.1310a

State of the literature

- The PISA test is well funded and is the largest study of education that has ever existed. The PISA reports are numerous and dominate national and international educational debates and policy.
- OECD has, with PISA as an instrument, become the most potent actor on the global educational scene. The underlying belief is that competition always promotes quality, and that markets and free choice are inherently good.
- The official OECD reports, reviews and advice based on PISA are influential and exert a pressure towards standardization of curricula, teaching and learning. Independent, critical voices do exist, but are often neglected and ignored.

Contribution of this paper to the literature

- It is argued that PISA is mainly a social, political and ideological project, not just as an educational project. The ideological and political commitments underlying PISA is described and criticized.
- The article outlines the process of the development of the PISA test, from intentions, through framework to the actual test items, and the problems encountered in this process.
- Dilemmas and paradoxes emerging from PISA results are outlined, in particular for science education. It is argued that the race to improve the PISA-scores and rankings is in conflict with basic views of what constitutes ideals for good science teaching and learning.

machinery, undoubtedly the world's largest empirical study of schools and education. Recent estimates suggest that the annual cost is around 80 million Euro (PISA-leader Schleicher in interview, Sydney Morning Herald, Nov 29. 2013). This sum does, however, *not* include the costs of involving half a million of students, tens of thousands of schools and their teachers. Given the size and importance, PISA has to be understood not just as a study of student learning, it has to be understood as a "social phenomenon" in a wider political, social and cultural context. (Lundgren, 2011).

PISA rankings create panic and discomfort in practically all countries, also in high-scoring countries. (Alexander, 2012). This produces an urge for politicians and bureaucrats to do "something" to rectify the situation. But since PISA does not tell much about cause and effect, the creativity blooms, and educational

reforms that are not at all empirically founded are introduced, often overnight.

The present article will present critical points of two categories.

The first category relates to the PISA project as such. These problems are inherent in the PISA undertaking, and hence cannot be "fixed". It will be argued that it is impossible to construct a test that in a fair and objective way can be used across countries and cultures to assess the quality of learning in "real-life" situations with "authentic texts". Problems also arise when the brave intentions of the PISA framework are translated to 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. This runs against most current thinking in e.g. science education, where "science in context" and "localized curricula" are ideals promoted by e.g. UNESCO, science educators as well as in national curricula.

The second category of critical points relates to some of the rather intriguing results that emerge from analysis of PISA data: It seems that pupils in high-scoring countries also develop the most negative attitudes to the subject. It also seems that PISA scores are unrelated to educational resources, funding, class size etc. PISA scores also seem to be negatively related to the use of active teaching methods, inquiry based instruction and the use of ICT. Whether one believes in PISA or not, such intriguing results need to be discussed.

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

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 for IQ-scores, there are serious concerns about the validity of the PISA-scores. What does PISA claim to measure and how does it live up these claims? The first point, then, is to look into what PISA claims to measure.

WHAT DOES PISA CLAIM TO MEASURE?

The emerging picture is in many ways confusing. In some places PISA claim that they do not measure school knowledge or competencies acquired at schools, in other places they state that they actually do measure the quality the nations' school system. Let us consider some details. The overall aims of PISA were stated

already 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)

These exact words have been repeated in practically all PISA reports from OECD during the 15 years since then. One can hardly object ambitions like this. It would be great if PISA really “answers these questions and more”, as they also state at the same place.

In other parts of their reports, they are a little more modest. They stress that PISA scores do not actually provide measures the quality of education systems, but the collective results of school, home and social environment. In the main report from PISA2006 this is expressed clearly:

If a country's scale scores in reading, scientific or mathematical literacy are significantly higher than those in another country, it cannot automatically be inferred that the schools or particular parts of the education system in the first country are more effective than those in the second. However, one can legitimately conclude that the cumulative impact of learning experiences in the first country, starting in early childhood and up to the age of 15 and embracing experiences both in school and at home, have resulted in higher outcomes in the literacy domains that PISA measures. (OECD, 2007a)

PISA is rather explicit that they do not measure according to national school curricula, but based the definitions and the framework made by the OECD-appointed PISA experts (OECD, 2006). The PISA Technical Reports clearly state that the knowledge and skills tested on 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, 2009). The same report also states that items that are close to the curriculum and items with “school science” are excluded.

So, although PISA states that it does not test school knowledge, and that it does not test according to national curricula or testing school knowledge, the PISA results are interpreted, also in OECD reports, as valid measures of the quality of national schools systems, and the PISA reports are full of policy recommendations regarding schools (Loveless, 2009). More on this later.

THE POLITICS OF THE OECD PISA PROJECT

OECD is the organization for the highly industrialized and economically developed nations, and the mandate of the organization lies in the name: Organization for Economic Cooperation and Development. The home site (<http://www.oecd.org/>)

is explicit about the mission of OECD. 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 Economy, not Education. 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. These subjects are seen to be key elements for the 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, physical education etc. One might also note that the PISA does not address aspects that are central in many countries' official purposes of education, like equity, empathy, solidarity, curiosity and engagement, care for the environment, etc. In the public and political debates, these statements about the agreed (and legally binding) purposes of the school system are often forgotten or ignored when discussions about 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 also a consequence of the mandate and priorities of the OECD. This module was included in the testing by some 40 countries.

OECD is often very clear about the economical purpose 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 90 percent of the world economy. (OECD, 2010a, Foreword)

There seems to be a contradiction here. On the one hand OECD/PISA state that they do not measure according to school curricula, and not even the knowledge acquired at school. On the other hand they claim that they do evaluate “the quality, equity and efficiency of school systems.” As mentioned, it is also interesting to note that the importance of PISA is defined in terms of the fraction of the world economy, not in terms of the fraction of the world's population.

The competitive economical perspective is also at the forefront when PISA results are presented to the

public. At the PISA2006 Release Conference in Washington DC on, December 4, 2007, the invitation reads as follows.

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

The lessons learned from 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. (Extracts only, full quotes and videos on <http://www.all4ed.org/events/losingedge>)

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 ought to increase differences in salaries, reduce public spending, increase the rate of unemployment, reduce the level of sick leave salaries and reduce pensions for 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 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", i.e. could give more PISA points per dollar spent on education. The very definition of a good school is thereby ignoring the national priorities set for our school system. The OECD educational advice was that Norwegian schools can 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" is in fact that Norway should become a different kind of country, hardly an objective, neutral, "scientific" advice.

NATIONAL POLICIES BASED ON PISA: AN EXAMPLE

The attention given to PISA results in national media varies between countries, but in most countries it is formidable, and it has increased after several rounds of PISA testing.

In Norway, the results from PISA2000 as well as from PISA2003 provided war-like headings in most national newspapers. Our then Minister of Education (2001-2005), Kristin Clemet (representing Høyre, the Conservative party), commented on the PISA 2000 results, released a few months after she had taken office, following a Labour government: "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!" (This was the case in the recent cross country was championship.) (Aftenposten, January 2001).

The headlines in all the newspapers told us, with war-like headlines, that "Norway is a school loser". In fact, even the headings were misleading: Norway was close to the OECD average in the three test domains in PISA2000 and PISA2003, but this was translated to be "a loser".

The results from PISA shaped the public image of the quality of our school system, not only for the aspects that had in fact been studied, but for more or less all other aspects of school. It became "common wisdom" that Norwegian schools in general had low quality, and that Norwegian classrooms are among the noisiest in the world (although this had barely been studied in PISA). The media presented tabloid-like and oversimplified rankings. It seemed that the public as well as politicians accepted these versions 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 this false impression and remind the public about the limitations of the study and the fact that Norway was in fact among in the middle of the OECD countries. In sum; PISA created a public image of the quality of the Norwegian school that is not justified, and that may be seen to be detrimental. Surveys among Norwegian teachers have shown that they consider the effects of the PISA project as a serious cause of trouble in their daily work.

But PISA has not only shaped the public image of Norwegian schools, it also served as legitimization of school reforms. Under Clemet as Minister of Education (2001-2005), a series of educational reforms were introduced in Norway. Most of these reforms were legitimized by reference to international testing, mainly to PISA. In 2005, there was a change in government, and Clemet's Secretary of State, Helge Ole Bergesen, shortly afterwards 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 is the many references to large-scale achievement studies. He explicitly states that these studies provided the main arguments and rationale for curricular as well as other school reforms. Under the heading "The PISA Shock", he confirms 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: 41-42. Author's translation)

In these memoirs, Bergesen (2006) also describes 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 the test-scores of Norway, the ground was prepared for a major school reform, to be called The Knowledge Promotion. The other political parties followed suit.

Later, in the parliamentary election campaign in 2009, the prime minister candidate, and from 2013 our Prime Minister for the same party, 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, authors' translation)

It is most interesting that this statement was made shortly after the Norwegian Parliament unanimously had passed a new law stating the value foundation for Norwegian schools. No reference was made to the key words in this law (like equity, solidarity, empathy, concern for environment etc.) in the elections campaign. It is also notable that the red/green Labour-dominated Norwegian government that took office in 2005 has followed more or less the same policy.

In the many White papers on schools that have been presented from the government to the Parliament in the years 2001 to 2013, the reference to the PISA project and "OECD experts" steadily increase, and by 2013 they appear on nearly every page. There is no doubt that the major reforms of Norwegian schools the last decade has 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). We have, among many reforms, established a high-stakes national testing system, where the 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 schools system, and new reforms were introduced as attempts to meet the perceived challenge. Most reforms on curriculum, national testing, accountability, transparency etc. also follow the policy advice that emerges from the PISA-study: More private schooling, free choice of schools and the growth of private consultancies for testing and reporting are also current trends that follow in the panic created in wake of the PISA rankings.

In the autumn of 2014 Norwegian teachers went on strike. Not for higher salaries, but because of 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. It is evident that this development had been triggered by the "PISA-shock".

Similar developments have occurred in other countries, also our neighbouring countries Denmark and Sweden, but limited place makes this impossible here. It is interesting to note, however, that all these "solutions" to solve the perceived crisis are more or less the opposite of what our Nordic neighbour and declared "PISA-winner", Finland, is doing.

One may also note that other countries had similar PISA-shocks. Germany was one of them. The results from the first PISA round placed Germany below the middle. 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", very much the same way that the "Sputnik-shock" was "good news" for science and mathematics educators in the Western world when the Soviet Union launched their Sputnik in 1957.

The influence of PISA in Germany went even further, as noted also in this journal: "Germany finally introduced national standards for science education as one reaction following the results of the PISA studies" (Steffen & Höbke, 2014).

PISA, MARKET THINKING AND GLOBALIZATION

The point so far in the article is to argue that PISA should be seen and understood in a wider political context. The two key elements here are market thinking and globalization.

As for the 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. Terms like quality, efficiency, transparency, accountability and "value for money" are among the (often positively laden) terms that are used in these policy reforms in many public sectors. Public services like schools and higher education, culture, health and care are all being invaded by market terms. Other

(previously) public sectors experience the same trend: police, security, postal services, transport, water supply, handling of household garbage, sewage and waste water cleaning etc. Traditional public services are increasingly subjected to competitive bids where they compete with private actors. Out-sourcing of key public services is increasing internationally, often to multinational companies. This trend towards marketization and privatization characterize the development in several countries. And the education sector is in forefront, with OECD and its PISA project as an efficient tool. (Meyer & Benavot, 2013).

The other, and related, political/economical perspective is that of globalization. The economy is getting globalized, large multinationals are important, and the workforce has to be flexible and moveable. Nations and multinationals compete on a common market. Hence, there is a need for common standards in education, common systems for exams, degrees and qualifications. Such tendencies operate within units like the European Union, where an example is 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 harmonize and universalize national institutions like 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 are often the use of indicators, quantifiable and measurable standards that can be used for calculations (Popkewitch, 2011). PISA test scores and rankings are ideal for this purpose, whether the researchers behind the projects like it or not.

UNIVERSALLY VALID "REAL LIFE" INDICATORS?

The basic problem in PISA lays its own statements about what they want PISA to be, as documented above. A fundamental premise for the PISA project is that it is indeed possible to "measure" the quality of a country's education by indicators that are common, i.e. universal, independent of school systems, social structure, traditions, culture, natural conditions, ways of living, modes of production etc.

As noted, PISA claims that they measure "how well the young generation is prepared to meet the challenges of tomorrow's world". 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. Although life in many countries do have some similar traits, one can hardly assume that the 15-year olds in e.g. USA, Japan, Turkey, Mexico and Norway are preparing for the same challenges and that they need identical life skills and competencies. It is also important to remind ourselves that the PISA 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 countries that take part in PISA, the mismatch of the PISA test with the needs of the nation and its youth may become 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, towards 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 these young people will operate in a global, international market.

All countries have their own school and education systems based on national decisions, 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 as well as more concrete details like time allocations for school subjects, aims, objectives and curricula, exam structure etc. 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 school.

PISA, however, makes it explicit that they do not do not relate to any national school system, what they measure does not fit any country's school. In reality, it is kind of universal, presumably culture-free, curriculum as decided by the OECD and its experts. The rather explicit goal of 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. This is detrimental to the expression of educational values within countries and to cultural identity and diversity in general.

STEPS TOWARDS THE PISA TEST

The process from the PISA ambitions to the actual tests that the students get has several stages, each of them with serious obstacles where many decisions have to be taken. The first step from intention to test is of course the selection of the knowledge domains (or school subjects) that should be included. OECD chose

three domains ("literacies") for the PISA testing: reading (in mother tongue), mathematics and science. These are important and basic subjects, of course, but one should keep in mind that most domains are not included, as alluded to earlier in the article.

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

The PISA framework

The next step in the process towards the actual PISA test is to make a testing framework, in reality a curriculum. Here the experts come in. The key institutions (who win the bid) and the selected subject matter specialists are in charge of a lengthy process to develop this framework. The persons selected for this purpose are well known international experts in their fields, often among the most respected and merited 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 dominating language in all deliberations and working documents. In addition to the subject matter specialists, the 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 (see, e.g. OECD, 2013a). These documents could be used, not as models to be copied, but 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 framework, i.e. to use this framework for the development and selection of test items, and for the construction of the PISA test as a whole. There is no place here to go in detail on the technicalities in this complicated process, which is well described in the more than 400 page technical report. (see, e.g. OECD, 2009 for the PISA2006 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 with initial screening and selection, national and international piloting, pre-field trials, main field trial round and psychometric analysis that involve many actors and subcommittees and many meeting for negotiations and debate, the final series of test items is decided. The complication of (just one of the many stages 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 2009, 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 that are too close to real life contexts of some countries, but not in 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 2009, p 34)

This clearly states that units (items) that relate to issues that are considered "inappropriate" (controversial in a particular country), has a "cultural bias" (be it positive or negative), or is 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 school knowledge or issues related to school curricula. From the above it seems somewhat strange that such a test is used to judge the quality of science taught at school in each country.

For example, in the final test, Norwegian students will find nothing about the key elements of the Norwegian economy. They will not find questions about oil and gas in arctic conditions on the continental shelf, aqua-culture and fish farming, hydroelectric power plants etc. Neither will they find anything about current topical issues and conflicts regarding conservation of nature, nothing about current political conflicts between nature conservation (i.e. wild wolves) and sheep farming, nothing about snow, skiing or skating, nothing about the Northern light (which was the foundation of the University in Tromsø) or about the challenges of an arctic climate etc. Students in other countries are, of

course, not likely to find questions relating to their own culture, nature, history or current national challenges.

In reality, the test items in the final test are decontextualized, or the context is contrived or historical. Not by the intentions in the testing framework, but from statistical necessity and concern for “fairness”. This runs contrary to all recommendations by science educators as well as by many national curricula of promoting a science curriculum that is relevant, interesting and context-based, at least for the compulsory school level.

Item texts, language and translations

A further set of complications arise relating to item texts, language and translation. Most PISA items are based on rather lengthy texts that constitute the stem, called “stimulus”. The intention is positive, namely to present real, authentic texts and real-life situations. But this format, in particular the length and complication of the stimulus text, also make the PISA items rather 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 is, of course, a deliberate choice by PISA specialists, and it also underlines that PISA does not really test subject matter school knowledge.

It is often claimed that many PISA items to a large degree are testing reading skills rather than science and mathematics competencies. The strong correlations between the test result on the reading, mathematics and science lends 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 in order to reduce the heavy load on reading skills.

A robust finding in PISA (as well as other kinds of reading tests, like PIRLS (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 are more in favour of girls than in most other kinds of tests. This unusual gender pattern may, at least partially, be explained by the heavy reading load in many PISA items. PISA tests scores show a gender pattern in science and mathematics that is rather different from e.g. TIMSS results in many of the same countries, as well as other tests, like national exams. It is also interesting to note that the PISA gender pattern becomes rather different 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 this test, the boys actually

outperform the girls in science (OECD, 2010b). This is an indication that also the context and the mode of data collection influence the results to a significant degree.

The “authentic texts” which constitute the stimulus in each item have originated in a certain situation in one of the OECD countries, and, of course, in the language of that country. This text is then, as mentioned, translated into the two official PISA languages before submission to PISA. The item is then translated into the language of each of the participating PISA countries. This translation process follows very strict rules that are laid down in detailed instructions. (see, e.g. OECD, 2009).

This translation raises many questions. Thorough work on the PISA reading test items has been done by Arffman, in her PhD (2007) as well as in journal papers (2010). She provides a detailed text-analytical study of the translation from English to Finnish of three PISA items. From a linguistic, translation theoretical perspective, her study reveals in detail many critical dimensions in this process. One of her conclusions based translation theory and on a review of empirical studies, is that one can never arrive at what may be called “equivalence of translation”. She also notes the scarcity of research on this most important issue. Neither poetry nor good prose can be translated according to a formalised set of rules, a fact that all good translators will acknowledge.

But even where the quality rules should have been followed, strange translations do appear. There seems to be a lack of empirical studies to look into this very important aspect of PISA (and TIMSS, PIRLS etc.) testing. The key role played by these texts in PISA makes such a scrutiny very important. A thorough cross-national check of translation requires a cooperation of researchers from many countries, and with considerable linguistic skills as well subject matter knowledge.

But some languages lend themselves to rather easy comparisons, even for non-linguists. The three Scandinavian languages provide good examples. Swedish, Danish and Norwegian are very similar languages, in fact more like dialects, in part with a common literary tradition. Below follows a simple comparison, based on one single item.

Item translations: an example

The item about the cloning of the sheep Dolly is probably the best known PISA item, since it was released in 2006.

Based on this English (and the French) original, the three Scandinavian texts (now available from the national PISA web sites) 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 that fact that the resulting

3 versions are rather different from each other, and they have all changed the original in rather dramatic ways. In brief:

- The Swedish, Danish and Norwegian texts changed the word “nucleus” to become “cell nucleus”, and thereby providing the hint to that the “small piece” in question 2

S128: Cloning

Read the newspaper article and answer the questions that follow.

A copying machine for living beings?

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’.

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).

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.

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.

Question 1: CLONING

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 1. The English original text and two questions for the item “Cloning”. Reproduced exactly as it appeared in the student’s questionnaire. (PISA Document: ReleasedPISAItems_Science.doc, retrieved Oct 9, 2014)

is indeed a cell.

- While the English (and Swedish) texts states that he removed “the material that would have determined sheep 2 characteristics...”, the Danish texts states that “he removes 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”, which makes the sentence more or less meaningless.
- The Danish text altered Question 1, and asks “Which sheep in Dolly a copy of?” (Probably because they find the word identical problematic, which is, indeed true.) The Danish version is also more in line with the title of heading in the text. “A copying machine for living things”. (This way of talking and writing about cloning is actually never used in any Nordic language, probably not in other languages either?)

PISA reports assert the readers that it has top quality in translation processes as well as in all other aspects of its work:

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 2009, p 91)

The procedures for this translation control is then described in detail. The “translation equivalence across PISA countries” is also thoroughly discussed in Grisay et al (2007). In the light of this, it is rather surprising that big blunders can be discovered by just a cursory look at published items.

Even a hasty reading by non-experts show that the translated texts are put in a strange and awkward prose that one cannot find published in any Scandinavian publications. Such texts cannot possibly be called “authentic”. Arffman (2010) notes that bad translations may also cause the readers to lose interest and motivation to get engaged with the text, and that this may severely have a negative effect on the tests results. This effect, I assert, is likely to be higher in countries where students are critical, independent and unwilling to obey the authority of schools and the 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?

As noted, the basic claims of PISA is that they test how well young people are “prepared for future challenges”, whether “they can analyze, reason and communicate effectively”, whether they have “the capacity to continue learning throughout life? And to what extent they have acquired some of the knowledge and skills essential for full participation in society.”

These ambitions are great, but are directly contradicted by the very format of the testing: The PISA test is a pen-and-paper test, where students sit for 2½ hours to answer written questions, in solitude and without access to sources of information. How “real life” is this 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 skilled workforce? Put in this form, the questions are rhetorical: the PISA test situation does not resemble any real life situations. The only place where you sit in solitude with a written test is in fact in exams at schools and universities. The only places where you are not allowed to communicate or allowed to use modern information technologies are similar test situations.

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

TEST SCORES AND ECONOMIC PROSPERITY

It does sound "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 article, bad rankings on PISA are thought to be bad signals for the future of the country. This assumption is probably the main reason for the extreme importance that is given to PISA results and rankings. PISA is in fact, "sold", presented and understood in this perspective, as also noted earlier.

But this "common sense" assumption may now be studied empirically. In January 2013 New Scientist brought an article with the heading West vs Asia education rankings are misleading.

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 Program for International Student Achievement (PISA). [...]

An 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. (MacGregor Campbell, *New Scientist Magazine* 07 January 2013)

Among the references in this article is a study by Chris Tienken (2008), who has used a series of indicators for a country's economic competitiveness and prosperity and seen how these correlate with scores on international tests of TIMSS (data since early 1960's) and PISA (since 2000).

Such studies are most interesting, since 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 are given below.

COMPETENCIES FOR THE FUTURE: THE VOICE OF EMPLOYERS

There are many sources that provide qualified accounts of the skills and competencies that large employers in the high-tech sector require from its workforce. At the Official Bologna Seminar in 2008 on "Employability: the Employers' Perspective and its Implications" Dr. Frank Stephan Becker, head of Human Resources in Siemens, gave a presentation of his company's view on what competences they need from their 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 above list are not addressed by the PISA test.

The second example is an investigation done by the 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, they ended up with the following ranking of competencies for future leaders and key personnel.

Competencies for future leaders and key personnel.

Good communication skills

Sense for Strategic thinking

Ability to motivate

Concern for staff and co-workers

Self-confidence and self-efficacy

Solid educational background

Visionary

Understanding numbers and quantities

(Author's translation)

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

PROBLEMATIC STATISTICS AND LACK OF TRANSPARENCY

The PISA project is a large undertaking. It has many of the characteristics of what is called "Big science" and "techno-science": It is costly; it involves the cooperation of around 70 countries. The logistics of the project is complicated, and there are 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, nearly half a million of students spend 2½ hours answering the test and the questionnaire, data are carefully coded by thousands of specially trained markers etc. etc.

Some of the many problematic issues in the process from intentions to test items have been raised in above. But there are more issues that are problematic. The final test consists of items that are selected in the process previously described. But the booklets 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 their test. After the time-consuming and tedious coding and data entry process, the data undergo complicated statistical analysis. The statistical process 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 in the publicly available tables is long and not very transparent, even for statisticians and other well-informed readers. The methods have, however, been criticized by well-qualified statisticians, also among those who actually work on PISA data. Svend Kreiner, professor of biomedical statistics at Copenhagen University, argue that he can get Denmark up from as PISA rank number 2 down to 42 by changing some of the parameters in the complicated statistical analysis. He also notes that the PISA methods of statistical calculations only are published in a very general form, making detailed critique difficult (Kreiner & Christensen, 2013).

PROBLEMATIC AND INTRIGUING RESULTS

The main concern in this article, as apparent from the above, is about the political/economic aspects of PISA, and the overall, basic weaknesses with the project. But there are also serious concerns that should be addressed, especially by those who embrace PISA and who 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 has no influence?

Already from the first PISA round, OECD produced graphs and indicators that showed 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 the quality.

It is interesting to note that 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, for instance, is highest in PISA score, but lowest in spending. This is used by political debate: Finnish teachers have difficulties in asking for higher salaries and more funding, since they already are on top of the ranking, and no changes need to be done. Norway, on the other hand, is much lower on the PISA ranking, but with higher in public spending on schools. Based on PISA, Norwegian politicians have argued that it has been proved that more spending would not increase the quality of schools. As noted earlier, the OECD (2008) Economic Report to Norway actually warns Norway to increase spending on schools, stressing that "this will not improve the quality".

PISA findings on cost and funding, like the above, are frequently used in influential OECD publications,

like the annual Education at a Glance. They conclude that "averaged across OECD countries, there is potential for reducing inputs by 30.7 % while maintaining outputs constant (OECD, 2007b).

High PISA science scores → 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 also has many questions about attitudinal aspects of how young people relate to science. This was an important element of the PISA2006 study, when science was the core subject. The definition of science literacy in PISA2006 actually included "willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen" (OECD 2006). The indices and constructs that were developed for this broad category were, however, not included in the PISA scores that were used for rankings etc. A special issue of International Journal of Science Education (2011, vol, 33, No1) presents several interesting results from analysis based on these data.

The simplest and possibly most surprising finding is that many countries with the highest mean PISA science score were at the bottom of the list of students' interest in science (Bybee & McRae, 2011). Finland and Japan are prime examples: at the top on PISA science score, and at the very bottom on constructs like "interest in science", "future-oriented motivation to learn science" as well as on "future science job", i.e. inclination to see themselves as scientists in future studies and careers. In fact, the PISA science score correlates negatively¹ with Future science orientation ($r = -0.83$) and with Future science job ($r = -0.53$) (Kjærnsli & Lie, 2011).

Such findings are most disturbing. If the students in PISA top ranking countries leave compulsory school with strong negative orientations towards science, one need 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 a cause and effect, but one should at least think twice before using these countries as educational models and ideals to be copied.

It is also interesting to note that many of the winners in the PISA science score also have the largest gender differences in PISA score. Finland is again a prime example, where girls outperform boys on all three

¹It should be noted that the above negative relationships are when countries are the units of analysis. When individual students within each country are the units, some of the correlations are positive. The unjust statistical inference from differences between groups to individual differences is actually labeled "ecological fallacy".

subjects in 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”.

Class size does not matter, large schools better than small?

Another set of PISA results related to the use of resources concern aspects like class size and size of schools. PISA scores are slightly negatively related to such variables, and this leads to advice that schools may become more cost effective if class sizes as well as the size of schools are increased. This was, as noted, also the advice given to Norway by OECD (2008).

Of course, such calculations are purely statistical, and do not take into account the geographical spread of the population or cultural and national concerns about social and regional policies.

Traditional teaching → better results?

The PISA students’ questionnaire has a series of questions about family background, home environment, cultural artifacts etc. It also contains a series of questions to students about the teaching methods and classroom practices that characterizes 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 with the results is that they run contrary to current advice from science educators as well as “accepted wisdom” among policy-makers, curriculum specialists etc. The following is a brief indication of some problematic results.

A trend in current science education is the stress on active learning and inquiry methods. Such views are expressed by panels of science education specialists (e.g. Osborne & Dillon, 2008) as well as from expert reports from the OECD (2006b) and from the influential EU report Science education NOW! (EU, 2007). These and other policy statements have channeled much of the research and development work supported by the European Union’s current Frame Programme FP7 into Inquiry Based Science Education (IBSE).

Reference is often made to the French programme *La main à la pâte* organized by l’Académie des Sciences. The key person in the project is Pierre Lena, well-known astrophysicist as well as former “Directeur général des enseignements supérieurs et de la recherche au ministère de l’Éducation nationale” This program was inspired by a US programme 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 et al., 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, inquiry-based science education (IBSE), 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 such advice, PISA rather provides counter-evidence. This possible contradiction should at least be seen as problematic.

The use of ICT leads to lower scores?

PISA has several questions regarding the use of the use of Information technology (ICT) in schools, and has made two constructs based on this. One construct or index is related to the use of internet at schools, the other is related to the use of software and educational programs. In a detailed study of the 5 Nordic countries, Kjærnsli et al (2007) document clear negative relationship between the use of ICT and PISA score. It is also interesting to note that the PISA winner, Finland, is by far the Nordic country with the least use of ICT, actually below the OECD average. In contrast, Norway is on top in the OECD in all indicators on the use of ICT in schools, but with only average PISA scores. Nevertheless, the policy advice in Norway is to increase the use of ICT in schools in order to achieve higher.

Intriguing PISA results: Concluding remark

Some of the above problematic results are not difficult to understand. A written test like PISA can hardly measure the skills and knowledge acquired 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 the 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 PISA, one has to take all the results seriously, also those which are counterintuitive and at odds with other research findings and policies that are recommended by educational experts.

CRITIQUE FROM ACADEMICS

In parallel with the increasing global influence of PISA on educational debate and policy, there has been a

growing critique in the academic world. Several readable anthologies have been published (Hopman et al., 2007; Pereyra et al., 2011; Meyer & Benavot 2013). The authors come from many countries and many academic fields and include well-known philosophers, sociologists, economist 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 in OECD. In the letter (Meyer et al., 2014) they voice a series of concerns about the growing influence of PISA. They argue that PISA is killing the joy of learning and lead to the detriment of basic values that schools should strive for. This initiative received public attention, also through coverage in *The Guardian* and other news media worldwide. The open letter has later been signed by more than 2000 academics from about 40 countries. Behind the initiative we find leading educators like 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, appointed to public office by Presidents George H. W. Bush and Bill Clinton. She is now, as distinguished professor of history and philosophy of education, the most influential critic of market-driven education policies she earlier had a strong belief in. She is the author of several publications, among these the influential book *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 the criticism of the uses and misuses of PISA now is rather common among most academics concerned about schooling and education. This critique has been fuelled in the past few years, also because PISA is extending its scope and influence in several ways. A further analysis of this development lies outside the scope of this article, but here are some aspects of the development.

PISA has recently established a close cooperation with Pearson Inc., the owner of *Financial Times*, *The Economist*, Penguin Group and Dorling Kindersley. 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. 80 % of Pearson's revenues now come from education, maybe the world's fastest growing market sector. Pearson won the bid for important parts of the PISA2015 testing and has developed strong links with OECD. Pearson has, of course, a vested interest in creating a market for its services and products. Through its close partnership with OECD it has come in a good position to expand its market as well as its influence. Diane Ravitch, mentioned above, is concerned about this influence, and

expresses it this way: "Are we prepared to hand over our children, our teachers, and our definition of knowledge to Pearson?" (Ravitch, *The Washington Post*, 7. May 2012)

PISA itself is also widening its repertoire. With a new product, "PISA for schools", individual schools or school districts may get "information and analyses [...] comparable to existing PISA scales. (<http://www.oecd.org/pisa/aboutpisa/pisa-basedtestforschools.htm> accessed 15.Oct. 2014). This, of course, may create an enormous market, and also bring competition, testing and rankings even closer to the daily activities in schools.

As mentioned, PISA is constructed to be used by the highly industrialized OECD countries, but has already some 30+ non-OECD participants. But OECD wants an even further development of PISA, and has launched a "PISA for Development" version. This is presented as "PISA survey instruments that are more relevant for the contexts found in developing countries but which produce scores that are on the same scales as the main PISA assessment." (<http://www.oecd.org/pisa/aboutpisa/pisaforddevelopment.htm> accessed 15. Oct. 2014).

With its PISA for Development, 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. Such scores will be seen as "benchmarks" and objective measures of quality by both donors and national authorities. OECD may thereby push aside the influence of UN-organizations like UNESCO and UNICEF, which may have rather different priorities than those of OECD.

While PISA has 15-year olds as target population, an emerging OECD project has been labeled "PISA for adults": Programme for the International Assessment of Adult Competencies (PIAAC). Results from the first full-scale data collection in 24 countries were published in 2013 (OECD, 2013), and the programme is likely to increase its importance, also for the sector of higher education and training. (<http://www.oecd.org/site/piaac/> accessed 15. Oct. 2014).

CONCLUSIONS

This article has focused on the problematic sides of PISA. 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 life as well as for the wellbeing of their societies. The data bank generated by successive rounds

of PISA is tremendous, and is most likely the largest and most professional data source in the history of social science and educational research. These data are also well documented and are open for most interesting research.

International comparisons in education are important; they can open for new perspectives, and they can provide inspirations 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 be eye-openers to acknowledge and celebrate the great variety between youth, nations and cultures on aspects of education, and as such serve as a source of inspiration. But such studies can also be used normatively, providing a pressure to oblige and fit to allegedly universal and common standards set from the authority of external specialists. We experience what is seen as a prime example of New Public Management (Møller & Skedsmo, 2013) as well as a kind of global governance and standardization of education, as also noted by key educational experts (Ball, 2012; Rinne, 2008). As indicated above, academics from several disciplines have raised concerns about various sides of the PISA undertaking, and OECD acting like a global ministry of education. The open letter, now signed by a large number of leading academics and school leaders is also a sign of a growing concern about how PISA is used to overrule national priorities in education.

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

The inherent difficulties in measuring what PISA asserts that it measures are seldom realized. The road from the brave intentions to an actual test instrument and valid data is long and murky. This article has pointed to some of the problematic issues in this process. This relates to the selection of subjects, (and of ignoring other subject/s). Fundamental problems are also inherent in the development of an international, fair test, which by necessity leads to context-free items. Further complications arise when items are to be translated to other languages. In this article and elsewhere (Sjøberg, 2007) I argue that it is not just problematic to live up to the intentions laid down in the overall statements of PISA. I argue that it is in fact “mission impossible”.

The public, media and policy makers, however, often take the PISA numbers and rankings as given facts. They trust that the experts know what they do, and that the numbers are objective and neutral measures. They trust that PISA scores measure what the intention has been.

No test is better than the items it consists of. The secrecy over most PISA items makes critique and scrutiny from the academic community and even the public difficult. Many of the published PISA items have met serious critique, both for its contents and for its language and relevance. Translations into the many different languages have only to a limited degree been examined, but it is easy to find flaws and even substantive changes and mistranslations. More research is needed here. The problematic use of statistics should 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 as well with what politicians prescribe as medicine to improve the quality of schools. Many politicians want to copy the PISA winners, but to do so, they often prescribe measures that are the opposite of what these winners actually do. There is a need to address seriously these paradoxical results. If one really believes in PISA, one has to accept and to address also some of these intriguing findings.

PISA has a profound influence on educational policy in many countries, and this 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 a 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 PISA/OECD and the global, commercial actor Pearson.

Science education and education in general, is always part of a larger picture. Our landscape is constrained, influenced and often determined by factors and forces that we do not always understand. As educators and researchers we are often too embedded in our own subject. The need to see our situation and activities in a wider political and cultural context is more important now than ever before. PISA and OECD are important elements in this larger picture.

Large resources are used to run the PISA project and to produce their 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.

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.
- 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: University of Jyväskylä, Institute for Educational Research.
- Arffman, I. (2010). Equivalence of Translations in International Reading Literacy Studies, *Scandinavian Journal of Educational Research*, Vol. 54, No. 1, February 2010, 37–59.
- Ball, S.J. (2012). *Global education Inc: New policy networks and the neo-liberal imaginary*. London: Routledge.
- Bergesen, O. H. (2006). *Kampen om kunnskapsskolen* (Eng: The fight for a knowledge-based school) Oslo: Universitetsforlaget.
- Bybee, R., & McCrae, B. J. (2011) Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science, *International Journal of Science Education*, 33(1), 7 — 26.
- EU (2007). *Science Education Now: A renewed pedagogy for the future of Europe*, (The Rocard report), European Commission, Brussels: EC.
- Grek, S. (2009). Governing by numbers: the PISA ‘effect’ in Europe. *Journal of Education Policy*, 24(1), 23-37
- Hopmann, S.T., Brinek, G. & Retzl, M. (Eds.) (2007) *PISA zufolge PISA – PISA According to PISA*, LIT Verlag, Berlin
- Jenkins, E. W. (2009). Reforming school science: 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 elevs kompetanse i naturfag, lesing og matematikk i PISA 2006. [Norwegian PISA report] Oslo: Universitetsforlaget.
- Kreiner, S & Christensen, K. B. Bang (2013). Analyses of Model Fit and Robustness. A New Look at the PISA Scaling Model Underlying Ranking of Countries According to Reading Literacy. *Psychometrika*, June 2013
- Lawn, M & Grek, S (2012) *Europeanizing Education: Governing a new policy space*. Symposium Books
- Loveless, T. (2009). How well are American students learning? *The Brown Center Report on American Education*, 2(3).
- Lundgren, U. P. (2011) PISA as a Political Instrument in Pereyra, M.A., Kotthoff, H.G and Cowen, R (Eds.) *PISA Under Examination. Changing Knowledge, Changing Tests, and Changing Schools*. Rotterdam: Sense Publishers.
- Meyer, H. D et al. (2014). Open letter to Andreas Schleicher on PISA. (<http://oecdpsaletter.org/>, accessed 15. Oct. 2014).
- Meyer, H.D & Benavot, A. (Eds) (2013). *PISA, Power and Policy: the emergence of global educational governance*, Oxford: Symposium Books.
- Møller, J.& G. Skedsmo (2013). Modernising education: New Public Management reform in the Norwegian education system, *Journal of Educational Administration and History*, 45(4), 336-353.
- OECD (1999). *Measuring Student Knowledge and skills. A New Framework for Assessment*. Paris: OECD Publications.
- OECD (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD Publications.
- OECD (2006a). *Assessing Scientific, Reading and Mathematical Literacy, A framework for PISA 2006*. Paris: OECD Publications.
- OECD (2006b). *Evolution of Student Interest in Science and Technology Studies*, Policy Report from OECDs Global science forum. <http://www.oecd.org/dataoecd/16/30/36645825.pdf>
- OECD (2007a). *PISA 2006: Science competencies for tomorrow's world Vol 1 Analysis*. Paris: OECD Publications.
- OECD (2007b). *Education at a Glance* Paris: OECD Publications.
- OECD (2008). *OECD Economic Surveys, Norway* Volume 2008/13. Paris: OECD Publications.
- OECD (2009). *PISA 2006 Technical Report* Organisation for Economic Co-operation and Development. <http://www.oecd.org/dataoecd/0/47/42025182.pdf>
- OECD (2010a). *PISA 2009 Results: What Makes a School Successful? Volume IV* Paris: OECD Publications.
- OECD (2010b). *PISA computer-based assessment of student skills in science*. Paris. OECD Publications.
- OECD (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, Paris: OECD Publishing.
- OECD (2013b). *PISA 2012 Results in Focus. What 15-year-olds know and what they can do with what they know*. Paris: OECD Publishing
- OECD (2013c). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.
- Osborne, J, & Dillon, J (ed) (2008). *Science Education in Europe: Critical Reflections* A Report to the Nuffield Foundation, King’s College London
- Popkewitz, T. (2011). PISA: Numbers, Standardizing Conduct, and the Alchemy of School Subjects. In Pereyra, M.A., Kotthoff, H.G and Cowen, R (eds.) *PISA Under Examination. Changing Knowledge, Changing Tests, and Changing Schools*. Rotterdam: Sense.
- Pereyra, M. A., Kotthoff, H.-G., & Cowan, R. (eds). (2011). *PISA under examination: Changing knowledge, changing tests, and changing schools*. Rotterdam: Sense.
- Ravitch, Diane (2011). *The Death and Life of the Great American School System. How Testing and Choice Are Undermining Education*. New York: Basic Books
- Rinne, R (2008) The growing supranational impacts of the OECD and the EU on national education policies and the case of Finland, *Policy Futures in Education*, 6, 665–680

- Sjøberg, S. (2007). PISA and "Real Life Challenges": Mission Impossible? In Hopmann, S.T., Brinek, G. & Retzl, M. (Eds.) (2007) *PISA zfolge PISA – PISA According to PISA*, LIT Verlag, Berlin
- Steffen B. & C Höble (2014). Decision-making Competence in Biology Education: Implementation into German Curricula in Relation to International Approaches *Eurasia Journal of Mathematics, Science & Technology Education*, 2014, 10(4), 343-355.
- Tienken, C. (2008). Rankings of International Achievement Test Performance and Economic Strength: Correlation or Conjecture? *International Journal of Education Policy and Leadership*, 3(4).

