Recherches
en Education

Page d'information pour le suivi de votre article
Reserches en Education Numéro 14 - Septembre 2012

PISA : politique, problèmes fondamentaux et résultats surprenants

PISA: Politics, fundamental problems and intriguing results

English manuscript, ***La Revue, Recherches en Education*** **Numéro 14 - Septembre 2012**

Svein Sjøberg, University of Oslo

svein.sjoberg@ils.uio.no, Home page http://folk.uio.no/sveinsj/

Abstract

The PISA project has in the last decade steadily increased its influence on educational policies in participating countries. Tables of PISA scores and country rankings are often taken at face value, not only in media, but also by policymakers and politicians. The PISA undertaking is a well-funded multinational "techno-scientific" machinery, there are few critical voices and very limited informed debate and critical research. This article will raise some of the concerns that would require further, more systematic scrutiny.

In this article, I will start by raising some basic questions about why OECD has initiated the PISA project. The intentions of PISA are, not surprisingly, strongly related to the overall political aims OECD and the underlying commitment to a competitive global free market economy. Examples are given on how PISA exerts influence on educational policy.

Then I will provide 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". I will argue 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 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 all local, current and topical issues must be excluded. This runs against most current thinking in e.g. science education, where "science in context" and "a localized curriculum" are ideals promoted by e.g. UNESCO, educators as well as in many 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, 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.

Introduction

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.

Since the first publication of PISA results in 2001, based on the testing in 2000, the results have become a kind of "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.

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 of education. Let us consider some details.

The PISA home page introduces PISA like this:

> "*Are students well prepared for future challenges? Can they analyse, reason and communicate effectively? Do they have the capacity to continue learning throughout life? The OECD Programme for International Student Assessment (PISA) answers these questions and more, through its surveys of 15-year-olds in the principal industrialised countries. Every three years, it assesses to what extent students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society."* (Complete full quote from http://www.pisa.oecd.org , accessed Jan 31, 2012)

One can hardly object to an ambition like this. It would be great if PISA really "answers these questions and more". But PISA should be judged not by the ambitions, but what they actually do measure.

In the above statement and elsewhere, OECD does *not* claim that PISA scores measures the quality of *school 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 reports also stress that they do *not* measure according to national school curricula, but based the definitions and the framework made by the OECD PISA experts (OECD 2006). The PISA Technical Report (OECD 2009) clearly states 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." 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 OECDs 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 economical 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 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, 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.

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*." It is also interesting to note that the importance of PISA is defined in terms of the fraction (90 percent) 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 read 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, economical 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 report to Norway in 2008, OECD expert 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 2008 OECD report to Norway had the education system as the focus. With PISA data as input for calculations, the essence of OECD educational advice was the following: Norwegian schools can become better by Closing smaller schools, Increasing class size, Introduce more testing, Publish results at school (and teacher) level, Base 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. This is hardly an objective, neutral, "scientific" advice. One should also not that "better schools" are defined as "more cost-effective", and further operationalised to mean more PISA points per dollar.

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. 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 PISA2000 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!" (which was the case in the recent cross country championship) (Aftenposten, January 2001).

The headlines in all the newspapers told us, with war-like headlines, that "Norway is a 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[1]".

[1] For some reason, Norwegians had expected that we should be on top – as we often are on other indicators, like the UNDP Human Development Index (UNDP 1990-2012) and in winter sports. When we are not the winners, we regard ourselves as being losers.

The results from PISA (and TIMSS as well) have shaped the public image of the quality of our school system, not only for the aspects that have in fact been studied, but for more or less all other aspects of school. It has now become “common wisdom” that Norwegian schools in general have low quality, and that Norwegian classrooms are among the noisiest in the world (although this has not been studied). The media has presented tabloid-like and oversimplified rankings. It seems that the public as well as politicians have accepted these versions as objective scientific truths about our education system. There has been little critical public debate, and the researchers behind the PISA study have done little to modify this false impression and remind the public about the limitations of the study. In sum; PISA has 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 serous 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, published a book shortly afterwards in which he presented the "inside story" on the reforms made while they were in power. The key features of the book are 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).

Later, in the parliamentary election campaign in 2009, the prime minister candidate for the same party had the following main message, even with a personal signature in the newspapers:

> *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 Parliament unanimously had passed a new law as the value foundation for Norwegian schools, and when new national curricula were in the process of implementation. 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. A new White paper on lower secondary schools (KD 2011) was presented to the Parliament in 2011. It refers to the PISA project and “OECD experts” on practically every page, while the key words in the newly passed law with the value statements for Norwegian schools are barely mentioned.

Many examples indicate how PISA in effect redefines national curricula and overrules nationally decided purposes of schools and education. Similar examples could be given from e.g. Denmark and Sweden, but limited place makes this impossible here.

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. Traditional public services are increasingly subjected to competitive bids where they compete with private actors. This trend seems to characterize the development in several countries.

The other and related political/economical perspective is that of globalization. The world market 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, like the Bologna process and its introduction of a common degree system in higher education. A key role also for OECD is to promote common standards, indicators and measures.

The PISA project does not have these political/economical aspects as explicit goals, but PISA-based calculations are heavily used by OECD experts to provide "independent, expert advice" to member countries. These advices are given considerable political weight in member countries (and also in the more than 30 PISA countries that are not OECD members.) The PISA "philosophy" and testing provide ideals and models for the countries that take part, and these often overrule national curricula. They also influence new educational reforms, as showed above with Norway as an example.

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 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 quantifiable and measurable standards that can be used for calculations. PISA test scores are ideal for this purpose, whether the researchers behind the projects like it or not.

Basic, inherent problems with PISA: Mission impossible?

Universally valid "real life" indicators?

The basic problem in PISA lies 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 more or less identical for young people across countries and cultures[2]. Although life in many countries do have some similar traits, one can hardly assume that the 15-year olds in e.g. Japan, Greece, Mexico and Norway are preparing for the same challenges and need identical life skills and competencies. One should remember that PISA is testing the *whole age cohort* 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.

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 decision from intention to test is of course the selection of the knowledge domains (or school subjects) that should be included. OECD chose three domains 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.

In fact, does PISA live up to its main declaration? Can it provide answers to its basic questions: "*Are students well prepared for future challenges? Can they analyse, reason and communicate effectively? Do they have the capacity to continue learning throughout life?*"

[2] In fact, OECD and PISA does restrict this statement to be valid for "industrialized countries", i.e. the OECD member countries, but the influence of PISA is equally strong in the ca 30 PISA countries that are not OECD members.

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.

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 know internationally 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 2006). 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". Trough a complicated process with initial screening and selection, national and international piloting, pre-field trials, main field trial round and psychometric analysis that involves many actors and subcommittees and many meeting for negotiations and debate, a final series of test items are decided. The complication of (just one of the many stages i) 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 "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 exploitation 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 Norwegian university) 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 translations. 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 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 is a support to such a claim, as also noted in many research papers. 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 components 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, sometimes 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 Inga 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 TIMMS, 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 "amateurs". 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, Cloning, Dolly the sheep

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

The stem text of the original, in English, is reproduced in Fig 1. (the other original, in French, is reproduced at the end of this paper, for use in the French translation)

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 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 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 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 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) 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 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 Feb 20 2012)

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 laid down 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 Danish and Swedish texts changed the layout and formatting of the text, making the original short text lines into long lines. Line numbers were removed in the text as well as in the reference to the text in question 2.
- 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 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 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 cursory 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 2 ½ 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.

Competencies for the future – as seen by 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. Here are but two examples:

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 (66 %)**
- **Sense for Strategic thinking (61%)**
- **Ability to motivate (60%)**
- **Concern for staff and co-workers (58%)**

- **Self-confidence and self-efficacy, (56%)**
- **Solid educational background (48%)**
- **Visionary (33 %)**
- **Understanding numbers and quantities (24%)**

(Abelia, 2008, 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 many countries[3]. 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 2 ½ 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 many more issues that are problematic. The final test consists of items that are selected, 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. As well known, 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 actual work on PISA data. Svend Kreiner, professor of biomedical statistics at Copenhagen University, argue that he can get Denmark up from a very low position to nearly top PISA rank 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. (The Journal *Information,* April 14, 2011, http://www.information.dk/265524)

Problematic and intriguing results

The main concern in this article, as apparent from the above, is about the political/economical aspects of PISA, and the overall, basic weaknesses with the project. But there are also serious concerns that should be

[3] PISA now involves some "70 countries and economies, representing 90% of the World economy"

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 such 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 in particular 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 rank, 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, the Norwegian minister of education argued that it was 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 negatively[4] 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 with cause

[4] It 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".

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 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 official 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 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 rather meager 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 and 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 who are counterintuitive and at odds with other policies.

Conclusions

This article has focused on the problematic sides of PISA, mainly because these are so neglected in research as well as in the public debates. 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. This author is worried that the latter perspective is dominant in PISA, already from the design phase, and also in the way it is used by the OECD, and how it functions in public debates as well as in educational policy in many countries.

The official intentions of PISA 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 economical 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 the overall political/economical 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). 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 seem 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 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.

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

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

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.

Grisay A. de Jong J.H., Gebhardt E., Berezner A., Halleux-Monseur B. (2007) Translation equivalence across PISA countries. *Journal of Applied Measurement*. 2007;8(3):249-66.

Jenkins, E. W. (2009). Reforming school science: a commentary on selected reports and policy documents *Studies in Science Education* vol 45, no 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. [Norwegian PISA report] Oslo: Universitetsforlaget.

Loveless, T. (2009). How well are American students learning? *The Brown Center Report on American Education*: Volume II, Number 3

OECD (2001). *Knowedge 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.

Olsen, R. V., & Lie, S. (2011). Profiles of students' interest in science issues around the world: Analysis of data from PISA 2006. *International Journal of Science Education*, 33(1), 97–120.

Osborne, J, & Dillon, J (ed) (2008). *Science Education in Europe: Critical Reflections* A Report to the Nuffield Foundation, King's College London

Sjøberg, S. (2007). PISA and "Real Life Challenges": Mission Impossible? In Hopmann, S.T., Brinek,G. & Retzl, M. (Eds.) (2007) *PISA zufolge PISA – PISA According to PISA,* LIT Verlag, Berlin

KD (2011). [The Ministry of Education and Research]. *Motivasjon – Mestring – Muligheter* White paper 22 to the Parliament

UNDP (1990-2012). *Human Development Report*, New York and London, Oxford University Press

CV/resume

Svein Sjøberg is professor in science education at Oslo University, Norway. He has worked extensively with international and comparative aspects of science education through e.g. UNESCO, OECD, ICSU and the EU, and has won several prizes and awards for his research and promotion of science literacy. Home page http://folk.uio.no/sveinsj/

French version of PISA item "Clonage". Dolly (retrieved Feb 2010 from http://www.oecd.org/document/38/0,3746,en_32252351_32236173_34993126_1_1_1_1,00.html

CLONAGE

Lisez l'extrait de presse suivant et répondez aux questions qui l'accompagnent.

Une machine à copier les êtres vivants ?

Aucun doute : s'il y avait eu des élections pour désigner l'animal de l'année 1997, Dolly les aurait remportées haut la main ! Dolly est la brebis écossaise que vous voyez sur la photo. Cependant, Dolly n'est pas une brebis quelconque : elle est le clone d'une autre brebis. Un clone signifie une copie conforme. Cloner signifie « copier à partir d'un original unique ». Les chercheurs ont réussi à créer une brebis (Dolly) identique à une autre brebis qui a servi d'« original ».

Le chercheur écossais Ian Wilmut a été le concepteur de ce mécanisme à copier les moutons. Il a prélevé un minuscule fragment de la mamelle d'une brebis adulte (brebis 1). De ce fragment, il a extrait le noyau, ensuite il a transféré ce noyau à l'intérieur de l'ovule d'une autre brebis (brebis 2). Il avait préalablement retiré de cet ovule tous les éléments qui auraient contribué à donner les caractéristiques de la brebis 2 à l'agneau qui en serait né. Ensuite, Wilmut a implanté cet ovule manipulé de la brebis 2 dans une troisième brebis (brebis 3). La brebis 3 est devenue pleine et a donné le jour à un agneau : Dolly.

Certains savants pensent que, dans quelques années, il sera également possible de cloner des êtres humains. Cependant, de nombreux gouvernements ont déjà établi des lois qui interdisent le clonage des humains.

Question 1 : CLONAGE

À quel mouton Dolly est-elle identique ?

A À la brebis 1.
B À la brebis 2.
C À la brebis 3.
D Au père de Dolly.

Question 2 : CLONAGE *S128Q02*

Les lignes 15-16 décrivent la partie de mamelle utilisée par le chercheur comme « *un minuscule fragment* ». Le contenu de l'article permet de comprendre ce que veut dire ce « *minuscule fragment* ».

Le « *minuscule fragment* » est :

A une cellule.
B un gène.
C le noyau d'une cellule.
D un chromosome.