Invited symposium:  
The PISA Science Assessments and the Implications for Science Education: Uses and Abuses

Symposium Abstract
In 2015, science will be the major focus of PISA. Consequently, the framework for the science assessment was rewritten in 2012. This symposium will begin with a presentation of that framework by Jonathan Osborne, the chair of the science expert group who will provide an explanation of the changes and possible improvements from the 2006 framework. In addition in 2015, all of the assessments will be undertaken on a computer-based platform that has consequences for the form and type of testing. This presentation will be followed by three presentations that take a critical look at the major social and political impact that PISA is having on education systems, schools and the learning of science. Svein Sjøberg will argue that PISA should be seen in a political and cultural context, and as an instrument of power. He will claim that PISA has led to a global race, and that many countries use PISA to legitimize neoliberal school reforms that are detrimental to the values usually promoted by educators. In contrast, Magnus Oskarsson and Margareta Serder will look at the effects of PISA in one country – Sweden. Oskarsson will argue that the PISA results provide one external indicator and measure of the performance of the system. In Sweden performance has declined significantly in comparison to other countries and the divergence between high and low performers increased. As such it provides a useful contrast to internal measures which portray a different picture. Finally, Serder will finish by presenting a study that has explored how groups of 15 year-old students from an average comprehensive school interpret the PISA items and construct responses. Her findings cast doubts on the validity and comparability across countries.

Chairperson: Jonathan Osborne School of Education, Stanford University United States osbornej@stanford.edu

Organizer List: Svein Sjøberg University of Oslo, ILS, Dept of Teacher Education and School Research Norway svein.sjoberg@ils.uio.no
The PISA Science Assessment Framework for 2015

Abstract
In 2015, the main focus of the PISA assessment will be science. As a consequence, the framework for the assessment in science was rewritten by a group of experts led by Dr Jonathan Osborne of Stanford University. In this paper, the rationale for the new framework and how it has modified and changed the previous version which has been used for all assessments since 2006 will be presented. The assessment framework for science in PISA is an attempt to measure three specific competencies - the ability to: explain phenomena scientifically; the ability to evaluate and design scientific enquiry; and the ability to interpret data and evidence scientifically. In the new framework, these three competencies are seen as requiring content knowledge, procedural knowledge and epistemic knowledge – a change from the previous framework which defined the knowledge required as a combination of content knowledge and knowledge about science. The paper will explain why this is seen as an improvement and how it contributes to a deeper understanding of what must be taught in science. In addition, the new framework now categorizes the cognitive demand of any item in terms of three levels which reflect the depth of knowledge required to undertake the task. Finally, in 2015, the PISA assessment will be undertaken using computers. This permits a new range of assessments and a measure of adaptivity – examples of which will be shown.

Summary
For many national governments, the outcome of the triennial, OECD PISA tests matter. For instance, in a survey conducted of the impact of PISA for seventeen countries the influence of PISA was seen to be ‘very influential’, eleven others identified it as ‘moderately influential’ and only five countries saw PISA as ‘not very influential’ (Breakspear, 2012). The Director of PISA, Andreas Schleicher, sees PISA as a tool for identifying poor performance in any countries education system. Indeed performance on PISA has been shown to correlate with economic growth (Hanushek & Woessmann, 2012). Along with the results of the TIMSS study, these tests have become an international benchmark that enable a country to judge the performance of its education system against that of other countries. Germany, for instance, suffered a severe blow to its sense of self-esteem when the 2000 results showed that their performance was merely mediocre (Breakspear, 2012). As a result, both Germany and Switzerland initiated significant programs of reform in response to lower than expected performance.

Moreover, the OECD sees PISA as a means of defining what constitutes a leading-edge conception of the outcomes of formal education. In the case of PISA, the operationalization of what should be assessed in these programs is a product of a dialogue between the OECD directorate, the PISA governing body and small panels of experts who draft the framework for consideration. For science, these outcomes are defined by the frameworks written for assessment in 2000, 2006 and 2015. The 2015 framework (OECD, 2012) is the first revision since 2006 and, hence, can be seen as an important contribution to defining an international perspective on what the outcomes of formal science education should be.
The PISA framework draws on the focus of many educational systems that an understanding of science is so important that it should be a feature of every young person’s education (American Association for the Advancement of Science, 1989; Confederacion de Sociedades Cientificas de España, 2011; Fensham, 1985; Millar & Osborne, 1998; National Research Council, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK), 2005; Taiwan Ministry of Education, 1999). Many of these documents and policy statements give pre-eminence to an education for citizenship. Likewise, the emphasis in the PISA frameworks is definitively on science for citizenship seeking to assess the competency of 15 year old students to become informed critical consumers of scientific knowledge – a competency that all individuals are expected to need during their lifetimes. This paper, provided by the chair of the science expert group, Dr Jonathan Osborne of Stanford University, provides an explanation and rationale for the nature of the latest assessment framework and the ways in which it will be assessed.

In PISA 2015 scientific literacy is defined by three competencies. These are the ability to:

1. Explain phenomena scientifically;
2. Evaluate and design scientific enquiry; and
3. Interpret data and evidence scientifically.

These competencies are seen to lie at the heart of what it means to reason scientifically and require a knowledge of science or what is commonly called content knowledge. The second and third competencies, however, require more than a knowledge of what we know. Rather, they depend on an understanding of how scientific knowledge is established and the degree of confidence with which it is held. Historically, specific calls have been made for teaching about what has variously been called ‘the nature of science’ (Lederman, 2006), ‘ideas about science’ (Millar & Osborne, 1998) or ‘scientific practices’ (National Research Council, 2012). Within PISA, the 2006 framework operationalized this aspect of science using the term ‘knowledge about science’. The major innovative feature of the PISA framework for 2015 has been to demarcate a knowledge of the standard procedures of the diverse methods and practices used to establish scientific knowledge – what is commonly called procedural knowledge from what is called epistemic knowledge. The latter is needed to understand the rationale for the common practices of scientific enquiry, the status of the knowledge claims that are generated, and the meaning of foundational terms such as theory, hypothesis and data.

Procedural and epistemic knowledge are necessary to identify questions that are amenable to scientific inquiry, to judge whether appropriate procedures have been used, to ensure that the claims are justified, and to distinguish scientific issues from matters of values or economic considerations. What then might be the constructs of procedural knowledge and epistemic knowledge? How might they be defined and how might they be demarcated from each other? This presentation will show how the new framework has defined these constructs and provide a rationale for their importance.

In addition, the new framework has attempted to define the cognitive demand of any items using a definition which defines the depth of knowledge required for any task – a feature which was absent from the previous frameworks. Finally, 2015 will be the first year in which the assessment will be undertaken using a computer
based platform. Computer based assessment not only offers some adaptive testing but also a wider and more diverse form of assessment of student competency and examples will be shown of the differences.

References


PISA – a global educational arms race?

Abstract

The PISA project has to be understood as a social phenomenon and a political project, in essence a well-funded instrument of power. 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-stakes test. Governments are blamed for low scores, and governments are quick to take the honour when results are improving. National curricula, cultural values and priorities are pushed aside.

The PISA project has positive virtues in potentially raising debates about the purpose and contents of science education and science for literacy and citizenship. Science educators may appreciate the thinking behind the framework for testing, but we also need to consider the wider social, ideological context of PISA. In particular how PISA rankings are used to legitimize educational reforms that may contradict principles and ideals that educators value.

Many of the reforms that are legitimized by PISA can be characterized as New Public Management and neoliberal policies. Key words in these reforms are globalization, standardization, belief in competition, free choice and market thinking. The presentation will give examples of the uses and misuses of PISA, and will raise some of the problems caused by PISA as a global measure of quality. The battle to improve PISA rankings may conflict with our work to make science relevant, contextualized, interesting and motivating for young learners.

Summary

Since the first publication of PISA results in 2001, the results have become a kind of global “gold standard” for educational quality - a single measure of the quality of the education system. In many countries, educational reforms have been launched as direct responses to the PISA results. While some try to copy the PISA winners; others do just the opposite of what high achieving countries actually do.

The intentions of PISA are, not surprisingly, related to the overall political aims OECD and the underlying concern for economic development in a competitive global free market economy. PISA is constructed and intended for the 30+ industrialized and wealthy OECD countries, but has later been joined by a similar number of other countries with developing "economies". When the PISA results are presented, they are seen as an indicator for future competitive edge in a global economy (Sjøberg, 2015).

The "common sense" assumption that the country's future depends on success in PISA, may also account for the extreme importance that is now attributed PISA rankings: Bad rankings on PISA are thought to be bad signals for the future of the country. Research on the relationship between a PISA (and TIMSS)-scores and economic development and other indicators on welfare shows that this assumption does not hold true when the analysis is limited to the comparable OECD-countries (Tienken, 2008). Much of the political panic that is created by what is perceived as low national PISA-scores is undermined by such findings.
The PISA undertaking is a well-funded multinational “techno-scientific” exercise, undoubtedly the world's largest and most costly empirical study of schools and education. Given the size and importance, PISA has to be understood not just as a study of student learning, but also as a "social phenomenon" in its wider political, social and cultural context, as also acknowledged by people who played a key role in the OECD preparations of PISA. As chair of Centre for Educational Research and Innovation (CERI) in the OECD, Professor Ulf P Lundgren had until 2000 the key role in the preparation of PISA. Ten years later, he writes:

"The outcomes of PISA we hoped could stimulate a debate on learning outcomes not only from an educational perspective but also a broad cultural and social perspective. Rarely has a pious hope been so dashed. […]. (Lundgren, 2011).

PISA rankings create anxiety and discomfort in practically all countries, even 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 us much about cause and effect, the creativity blossoms, and educational reforms that are not empirically founded are introduced, often overnight.

Consequently, in many countries new curricula have been introduced, caused by "PISA-shocks", (e.g. Norway, Denmark, Sweden, Germany and Japan). In many countries national standards as well as new systems of obligatory national testing have been introduced. Some of these are directly influenced by PISA documents. Many countries publish such test scores as league tables, where school districts and schools are ranked according to these scores. Some countries have introduced incentives such as salary systems related to test scores for teachers and (in particular) principals. Free choice of schools further exacerbate the importance of the rankings, often widening the gap between schools, as well as creating ways to "improve" test rankings. Such rankings have several consequences, like the obvious "teaching to the test", but also influence the price of neighbourhood housing, widening socio-economic gaps between districts.

The strife for better test scores also serves commercial interests. Companies deliver products like tests and teaching materials that are supposed to increase scores, and cramming schools make a fortune by preparing students to achieve higher test scores. It is interesting to note that the world's largest educational company, Pearson Inc. now is directly involved in the running of PISA 2015. The partnership with PISA/OECD is also a strategic door-opener for Pearson into the global educational market. In company with OECD, Pearson also produces "The Learning Curve", a ranking of nations according to a set of test-based indicators. These rankings get media coverage and further create anxiety among policymakers. The result is a further pressure toward doing "something" to climb the league tables.

PISA is now used to legitimize neoliberal policies and reforms that are duly labelled New Public Management (Møller & Skedsmo, 2013). The PISA outcomes are also leading to an emerging global governance and standardization of education, as also noted by key educational experts (Ball, 2012; Rinne, 2008). The process is also called "governing with numbers" and the "PISA effect in Europe (Grek, 2009).

This presentation will have two critiques (Sjøberg, 2015). The first is an inherent feature of the PISA undertaking, and hence cannot be “fixed”. The PISA testing framework (OECD, 2013) is a most interesting document that should be used to inspire discussions about the purpose and contents of science curriculum and teaching. However, 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. 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”. 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 critique draws on some of the 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.

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
School science in a market driven school system

Abstract

Sweden’s results in PISA have shown the largest drop of all countries in all three subjects in the last twelve years. From results above the OECD mean with a high degree of equity in PISA 2000, to PISA 2012 with results below mean in science, math and reading and with a sharp increase of low performing students and schools. One important reason behind this seems to be the decentralization and market adaptations of the Swedish school since the mid 1990s. A free choice of school and a voucher system were introduced together with new legislation that allowed private schools fully financed by public means through the vouchers. This followed in the mid 2000s by new control and steering mechanisms with an expanded grading system, a vast increase in the number and incidence of national test and a school inspectorate.

Positive signs in results as a consequence of these reforms are hard to find. The increasing differences between students and between schools are not only in results. There are increasing differences in student attitude where some students feel more motivated while others feel increased social exclusion. Among schools it is easier for high achieving schools to recruit teachers and bigger cities are more likely to participate in school development programs and in service training.

These changes are not easy to track in the national testing or the national grading system, were there have only been minor changes. There is increasing evidence that both the grading system and the increased competition between schools have caused grade inflation.

PISA is sometimes connected with the new public management and the increasing trend to measure student results and the increased national testing. Sweden could be one example of this but what this paper shows is the strength in PISA to assess these kinds of reforms.

Summary

Sweden like the other Nordic countries has a long history of successful efforts to create a comprehensive school system with good results and a high degree of equity. The differences between high and low achievers and the difference between schools have been smaller than in many other countries and the same has been true for the impact of social background. The first PISA study 2000 also reported that the results for Swedish students were above mean in all subjects (OECD, 2001).

From the 1990’s Sweden has developed a more decentralized school system with a new curriculum and a benchmarked grading system. Influenced by New Public Management theories a School voucher system was established and the students became free to choose their school. Private schools were allowed, fully financed by public means. The effects of the reforms were minor during the 1990’s and during the first years of the following decade. From around 2003, however, changes have become more noticeable (Lundahl, Erixon Arreman, Holm, & Lundström, 2013).
The majority of the schools are still public, but PISA shows that that Sweden has had the fastest growth in the proportion of private schools between 2003 and 2012 of all OECD countries. One effect of this is that differences between schools have increased. A number of schools are abandoned by the most ambitious students, and segregation both by socioeconomic background and migration background have increased.

Since 2006 many new school reforms have been introduced. Examples of reforms are a formation of a national school inspectorate, increased quality control, more national testing and again an expanding grading system. A new mandatory national test in science for grade nine was piloted in 2009 and fully operational in 2010. A number of partly government-funded, school development and in-service programmes have been offered to schools. Several of these programmes have been directed to math and science.

In December 2013 Sweden experienced its first PISA shock. In PISA 2012, Sweden performed below the international average in all tested domains. The drop was the largest in OECD in all three domains since the start of PISA. A closer look at the science results from 2006 and onwards reveals not only a drop in the mean result but also an increasing difference between low and high achievers (OECD, 2014a). The decline has been more rapid since 2006 and especially after 2009 despite all of the efforts by the government.

There has been no significant change either in the number of top performing schools or in the number of top performing students in science. Instead, it is the number of low achievers that has increased and the same is true for the number of low achieving schools, where the proportion of schools with a mean below 450 in PISA science has increased from less than 5 % in 2006 to 20 % in 2012. Results for both boys and girls have dropped, but there has been a larger deterioration in boys’ results. While the proportion of girls not reaching PISA level 2 in science has increased from 15 to 19 percent, this proportion has increased from 17 to 25 percent for boys.

Differences between schools are also increasing. Ambitious students choose schools with high reputation while other students are left behind in less advantaged schools. And as PISA shows, there are few winners and many losers. Several reports show that the school-choice system and the voucher system are two important reasons (Skolverket, 2012). A study by Östh, Andersson, & Malmberg, (2013) shows that the cause of increasing differences between schools is school choice rather than increasing residential segregation. Another recent study shows covariance between increased between-school variance and decreasing PISA science results in a number of countries (Davidsson, Karlsson, & Oskarsson, 2013)

While the Swedish science results have dropped not only in PISA but also in TIMSS the same is not true for national grading or the national testing. Registry data for national test results and final grades in science show neither increasing numbers of failing students nor increasing differences between boys and girls. Preliminary results indicate that neither national tests nor grades are stable over time. The grading system and the increased competition between schools may have caused grade inflation (Oskarsson, Eliasson, & Karlsson, in progress).

A recent report has pointed out that participation in school development programmes varies greatly between different regions. Larger cities and towns with universities have participated in a majority of these programs while smaller communities in more remote areas only have a participation rate at around 15 percent or less in the national school development programmes.
These large differences between schools and regions also have the effect that what is offered to teachers’
varies. It becomes more attractive to work at schools with not only high-achieving students but also good in-
service training and strong support for professional development. This means that schools, that for different
reasons do not or cannot take part in development programmes, are likely to show decreasing results and
increasing difficulties in recruiting teachers. This is exactly what is pointed out in a recent report from OECD
(2014b) where Sweden is one of the countries where low performing schools have the largest difficulties in
recruiting competent teachers.

There are not only increasing differences in students’ results but also in students’ attitudes. Some students feel
more motivated while others feel increased social exclusion. Several of these background factors have strong
correlations with students’ results. More control and testing seem to have increased the extrinsic motivation
among some students while other show a more negative response to the greater pressure.

PISA has revealed important information about the development of the Swedish school system. In a period with
many reforms the pros and cons of these changes are often unknown when the next reforms are introduced.
When a nation’s assessment system are not stable over time as in the case of Sweden, international
comparisons give invaluable data to educators and policy makers about the state of the educational system.

References

naturvetenskap i en internationell jämförelse (in Swedish).
*Education Inquiry*, 4(3).
OECD. (2001). *Knowledge and skills for life : first results from the OECD Programme for International Student
[Elektronisk resurs]: OECD Publications Centre.
OECD. (2014b). Resources, policies and practices in Sweden’s schooling system: an in-depth analysis of Pisa
[http://www.skolverket.se/](http://www.skolverket.se/)
Östh, J., Andersson, E., & Malmberg, B. (2013). School choice and increasing performance difference: A
What do low PISA scores mean in the science classroom?

Abstract
The three scientific literacy competencies assessed according to the new framework for PISA 2015, place an emphasis on the most important skills to be taught in the science classroom. This paper takes the position that the assessment of competencies, which eventually render PISA scores that indicate high or low quality, need to be interpreted also on a classroom level. What do, for instance, low PISA scores mean when it comes to teaching, or understanding what students really are capable of doing, or not? The PISA debate, so far, mainly discusses the quality of school systems based on the scores achieved by the students, but critically misses the opportunity to use these scores to understand why students in a particular country obtained these scores. The purpose of this paper is to present the rational and the results of an empirical classroom study, designed to approach what low PISA scores might mean more specifically. The data were collected in the fall of 2010, which means they are based on the previous, but still similar, PISA framework from 2006. In the study, 15-year-old students from an average Swedish compulsory school collaboratively worked with test items from PISA Science, to discuss and respond to a selection of PISA questions. The analyses of 16 hours of student group work show the importance of high linguistic skills for engaging in the test questions, but also pose questions about what knowledge about science might mean and how international comparisons can overcome problems that arise when it comes to such issues as translation. The paper asks how policy discussions can be made to engage less in the general results and more the in local complexities of the science classroom and what is taught.

Summary
As argued by Osborne (2015, this symposium), PISA results matter. In Sweden, where this study is geographically situated, the former right wing government (2006-2014) implemented a series of educational reforms, for which several of the needs were based on PISA and TIMSS results: for example a new teacher education, replacing the one from 2001, was implemented with reference to decreasing PISA results. In addition, a new grading system and a new curriculum were reforms intended to strengthen the outcomes of Swedish students. One consequence of these reforms is that new primary science teachers in the new system must teach more school subjects than before and hence can be less educated in science. Another consequence is that the new national curricula contain far more detailed information about what to teach in, for instance, science. It seems as more control of the school system, and perhaps more learning from “PISA winners” (such as Finland), is the response to the low results. Meanwhile, the PISA results indicate an increasing gap between high scoring and low scoring schools, more low performing students and an increasing disillusionment about going to school (Skolverket, 2013).

In the academic PISA literature, there are several themes for the discussion, concerning the influence that PISA has on educational policy (Sellar & Lingard, 2013), methodological weaknesses in the statistical procedures (Allerup, 2007), international comparability (Arffman, 2010), and the increasing societal “sociology of measurement” (Gorur, 2014). Meanwhile, in science education, researchers have argued that the concept of scientific literacy articulated in PISA might have a positive effect in offering a good example of the goals and emphases of science education (Fensham, 2009). Other scholars in this field, such as Olsen (2012) argue that the
role that literacy in its more fundamental, linguistic sense (Norris & Phillips, 2003) plays in the PISA results needs to be more clearly emphasized in the scientific literacy framework (Olsen, 2012).

This paper contributes to this discussion by asking how a school system in the lower parts of the PISA results table, with Sweden as the example, might understand its PISA results on a classroom level. More specifically, when students do not know how to answer a PISA test question, or when they do know but fail to answer, what could be the possible reasons?

**Method**

In order to explore this issue, a study in which Swedish students interacted with and collaboratively answered test questions from PISA science (OECD, 2007) during a science lesson, was designed (Serder & Jakobsson, 2014). Three PISA units were included: Greenhouse, Acid Rain and Sunscreens (PISA unit number S114, S447, S485). In total, 21 groups of 3-4 students each were formed. This collaborative design was chosen based on sociocultural theories (Wertsch, 1998) which emphasize that knowledge is shaped in action, and gets meaning in real-life situations. The interactions were video recorded for further observation and semantic analysis (Mäkitalo, Jakobsson, & Säljö, 2009) of the 16 hours of video data that were collected. The focus for the analysis presented here was what specific problems the students experienced with the test questions.

**Results**

Two main themes in the data were how the students’ discussed science as it was portrayed in the test items, and how they interpreted different words/formulations of the test items. PISA items are contextualized in everyday life, a condition that seems to affect how the students approach the problems. A common difficulty for the groups was to identify the meaning of various words in their intended sense. Instead, words with different meanings in different contexts - everyday, science and mathematical - were frequently negotiated by the groups (Serder & Jakobsson, in review). The Swedish words for e.g. pattern, factor, reference, constant and better are examples from the study. For success, it appeared crucial to ignore all other interfering, possible contexts (including the everyday context). The results indicate that some meaning is also likely to be added, or lost, in translation procedures. Further, the students discussed, and questioned, the manner in which the fictive characters of the test items were speaking and approaching the scientific problems presented to them in their everyday life. In order to approach the test questions productively, the students needed to accept the artificial aspects of the problems, as well as the authoritative, highly academic way in which science was portrayed implicitly in those items (Serder & Jakobsson, 2014).

**Discussion**

This study was undertaken in student groups, which means that the situations were different from individual testing situations. This point is important to consider while interpreting the results. However, the study permits an insight in in the actual difficulties that might arise in a situation, of which we know very little. The reasons for the decreasing results for Swedish students are likely to be manifold. My research poses questions about comparability between different national versions of the PISA test, due to the problem of translation (c.f. Arffman, 2010), as well as about testing of real-life skills in school-like situations. The research suggests that scientific competence is highly linked to students’ discursive knowledge (c.f. Norris & Phillips, 2003; Olsen, 2012), both of what constitutes science as a particular human culture with certain norms and values, and as a particular way to talk about the world.
Unfortunately, there might be little correspondence between what school reforms are implemented in a country and the actual problems signaled by its school system. How can a research society, like ESERA’s, contribute to policy discussions in order for them to engage less in the general outcomes (such as fear for decreasing results) of PISA, and more in the local complexities of schools and science classrooms?

References


