

The PISA Science Assessment for 2015 and the Implications for Science Education: Uses and Abuses

Jonathan Osborne, Magnus Oskarsson, Margareta Serder,
and Svein Sjøberg

Introduction

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 17 countries, PISA was seen to be “very influential”, 11 others identified it as “moderately influential” and only 5 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’ educational system. Indeed performance on PISA has been shown to correlate with economic growth (Hanushek and 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. However, PISA is not without its critics. In a series of articles, Meyer and colleagues argue that PISA has become part of “a pervasive normalizing discourse, legitimizing historic shifts from viewing education as a social and cultural project to

J. Osborne (✉)
Stanford University, Stanford, California, USA
e-mail: osbornej@stanford.edu

M. Oskarsson
Mid Sweden University, Härnösand, Sweden

M. Serder
Malmö University, Malmö, Sweden

S. Sjøberg
University of Oslo, Oslo, Norway

an economic one engendering usable skills and ‘competences’” (Meyer et al. 2014). Labaree has argued that PISA assesses what nobody teaches (Labaree 2014) which Münch details in his exploration of how an Anglo-Saxon model of education has been forced on the German system (Münch 2014). To what extent are these criticisms justified? As science is the major focus of the tests in 2015, the findings will be particularly salient for the science education community. This chapter therefore offers a summary of a set of papers presented at a symposium at ESERA 2015 and seeks to explore the value of PISA and the legitimacy of such criticisms.

The PISA Science Assessment Framework: Advancing What It Means to Teach and Learn Science?

Jonathan Osborne

The goal of PISA is to define a set of competencies in reading, mathematics and science. Competencies are seen as “more than just knowledge and skills” requiring the ability “to meet complex demands, by drawing on and mobilising psychosocial resources (including skills and attitudes) in a particular context” (Rychen and Salganik 2003: 4). Thus the assessment framework for PISA offers an opportunity to define what constitutes a leading-edge conception of what the outcomes of formal education might achieve. 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 currently be.

The PISA framework draws on the view of many countries that an understanding of science is so important that it should be a feature of every young person’s education (Confederacion de Sociedades Cientificas de España 2011, Millar and Osborne 1998, National Research Council 2012, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, KMK 2005). Many of these documents and policy statements give pre-eminence to an education for citizenship. Likewise, the emphasis in the PISA frameworks is 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. The particular focus of the PISA science framework is on scientific literacy which is defined for 2015 as the competency to:

1. Explain phenomena scientifically.
2. Evaluate and design scientific enquiry.
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. PISA defines this very vaguely in terms of what it might be reasonable to expect a 15-year-old student to know. To do more would be impossible given that over 70 countries participate in the test. 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 what has variously been called “the nature of science” (Lederman 2007), “ideas about science” (Millar and 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? The first major contribution of the PISA assessment framework for 2015 has been to clarify these forms of knowledge required for science literacy. In addition, it is the first document to bring into being the construct of epistemic knowledge as an explicit feature of assessment and, by inference, an explicit feature of teaching and learning.

Second, the new framework has introduced a means of assessing the cognitive demand of items using a definition which defines the depth of knowledge required for any task – a feature which was absent from the previous frameworks (Webb 2007). 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 – producing a more valid assessment of student competency.

However, this view of the positive value of PISA is challenged in the following section by an argument that there are negative policy implications for educators, teachers and students.

PISA: A Global Educational Arms Race?

Svein Sjøberg

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 entire school system. An OECD report on the policy impact of PISA proudly

states that “PISA has become accepted as a reliable instrument for benchmarking student performance worldwide, and PISA results have had an influence on policy reform in the majority of participating countries/economies” (Breakspear 2012: 4).

Similarly, Andreas Schleicher (2012), director of PISA and recently also of Directorate of Education and Skills in OECD, in a TED talk starts his presentation by stating that PISA is “really a story of how international comparisons have globalized the field of education that we usually treat as an affair of domestic policy”.

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 2016). Governments are blamed for low scores, and governments are quick to take the honour when results are improving. 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 PISA undertaking is also a well-funded multinational “techno-scientific” exercise, undoubtedly the world’s largest and costliest 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 a 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: 27)

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 cannot tell us much about cause and effect, creativity blossoms and educational reforms that are not empirically founded are introduced, often overnight. National curricula, cultural values and priorities are pushed aside.

Consequently, in many countries new curricula have been introduced, caused by “PISA shocks”, (e.g. Norway, Denmark, Sweden, Germany and Japan). In many countries new national standards as well as new systems of obligatory national testing have been introduced. Some of these are directly influenced by PISA documents, as also proudly noted in a comprehensive report by the OECD itself (Breakspear 2012).

Many countries publish their own national test scores as league tables using them to rank school districts and schools. Some countries have introduced incentives such as salary systems that use test scores for teachers and (in particular) principals. Free

choice of schools further exacerbates 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 the rankings also influence the price of neighbourhood housing, thereby widening socioeconomic gaps between districts.

The strive for better test scores also serves commercial interests. Companies deliver products such as tests and teaching materials that are supposed to increase scores, and cramming schools make substantial profit from preparing students to achieve higher test scores. The largest PISA contractor is the US-based non-profit assessment and measurement institution ETS. Maybe more important is that the world’s largest commercial educational company, Pearson Inc., was involved in PISA 2015 and won the bid to develop the framework for PISA 2018. The joint press release from OECD and Pearson explains:

Pearson, the world’s leading learning company, today announces that it has won a competitive tender by the Organisation for Economic Co-operation and Development (OECD) to develop the Frameworks for PISA 2018. [...] The frameworks define what will be measured in PISA 2018, how this will be reported and which approach will be chosen for the development of tests and questionnaires. (OECD and Pearson 2014)

The partnership with PISA/OECD is also a strategic door opener for Pearson “with 40,000 employees in more than 70 countries” into the global educational market. In company with the OECD, Pearson also produces “The Learning Curve”, a ranking of nations according to a set of test-based indicators. PISA leader Andreas Schleicher sits on the Advisory Panel of The Learning Curve. These rankings get media coverage and further create anxiety among politicians and policymakers. The result is a further pressure towards doing “something” to climb in the league tables.

PISA is now used to legitimize neoliberal policies and reforms that are duly labelled New Public Management (Møller and 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). The process is also called “governing with numbers” and the “PISA effect in Europe” (Grek 2009).

The PISA testing framework (OECD 2012) is a most interesting document that could 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 is, of course, 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 use of PISA for political purposes is very selective. While the rankings of nations get a lot of attention, other results are ignored. It seems, for instance, that pupils in high-scoring countries 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, the inquiry-based instruction and the use of ICT. The fight to improve PISA rankings may conflict with the work to make science education relevant, contextualized, interesting and motivating for young learners. Whether one “believes in PISA” or not, such intriguing results need to be discussed.

As a contribution to that discussion, the next two sections draw on data from Sweden – one country whose performance on PISA has declined in recent years. These papers explore to what extent there is, or is not, a positive value to the outcomes of PISA and the comparisons that are made.

School Science in a Market-Driven School System

Magnus Oskarsson

Sweden’s results in PISA have shown the largest drop of all countries in all three subjects over the last 12 years. From results which were above the OECD average and with a high degree of equity in PISA 2000, these have fallen by 2012 to results which were below average in science as well as in math and reading and with a sharp increase of low-performing students and low-performing 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 to be fully financed by public means through the vouchers. This was followed in the mid-2000s by new control and steering mechanisms with an expanded grading system, a vast increase in the number and occasions of national tests, and a school inspectorate.

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 different 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 showed that Swedish students were above mean in all subjects (OECD 2001).

From the 1990s 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 were free to choose their school. Private schools were allowed, fully financed by public means. The effects of the reforms were minor during the 1990s and during the first years of the following decade. From around 2003, however, changes have become more noticeable (Lundahl et al. 2013).

The majority of the schools are still public, but PISA shows that Sweden has had the fastest growth in the proportion of private schools between 2003 and 2012 of all

OECD countries. One effect is that differences between schools have increased. A number of schools seem to have been abandoned by the most ambitious students, and segregation both by socioeconomic background and ethnic background has increased – a finding which supports Sjøberg's critique.

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, and several of these programmes have been directed to math and science.

When the PISA 2012 results were presented, it showed that the performance of Swedish students was 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 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 efforts by the government.

There was, however, no significant change either in the number of top-performing schools or in the number of top-performing students in science. Rather, 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 was a larger deterioration in boys' results.

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 et al. (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 et al. 2013).

While the Swedish science results have dropped not only in PISA but also in TIMSS, the same is not true, however, for national grading or the national testing. Data for national test results and for final grades in science neither show increasing numbers of failing students nor increasing differences between boys and girls, and there is growing evidence that neither the national tests nor grades are stable over time (Lundahl et al. 2013).

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% or less in the national school development programmes.

These large differences between schools and regions also have the effect that teachers' employment choices are more diverse. For instance, 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 a variety of 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 has been pointed out in a recent report from OECD (2014b) where Sweden is one of the countries where low-performing schools report both lack of resources and the largest difficulties in recruiting competent teachers.

Moreover, there are not only increasing differences in students' results but also in students' attitudes. Some students feel more motivated while others feel increasing social exclusion, and 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 others show more negative response to the greater pressure.

The value of PISA, however, is that it can reveal this kind of important information about national school systems, such as the Swedish – a point which Sjøberg does not consider. In a period with many reforms, the pros and cons of such change are often unknown when the next reforms are introduced. The Swedish example shows that when a nation's assessment system is not stable over time, international comparisons give invaluable data to educators and policymakers about the state of the educational system.

The final section looks not at the national data but at how students interpret the assessment tasks and their cultural specificity arguing that the assessments are too culturally specific to offer valid cultural comparisons.

An Interpretation of PISA Results from a Science Classroom Perspective

Margareta Serder

In Sweden, a series of educational reforms have been launched in the last 10 years, often explicitly addressing the decreasing results of the Swedish students in international assessments and PISA (Ringarp and Rothland 2010). A few recent examples of reforms intended to strengthen the outcomes of Swedish students are a new teacher education, a new grading system, a new curriculum and 1 h more of mathematics for all students. Still, the negative Swedish trend has not shifted.

In the academic conversation about PISA, several concerns with the assessment have been raised. Those address, for instance, the increasing impact that OECD/PISA has on educational policy (Sellar and Lingard 2013), methodological weaknesses (Allerup 2007) and translation issues that affect international comparability (Arffman 2010). Meanwhile, in science education, researchers have argued that the concept of scientific literacy as 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 argue that the role that literacy in its more fundamental, linguistic sense (Norris and Phillips 2003) plays in the PISA results needs to be more clearly emphasized in the scientific literacy framework (Olsen 2012).

While the reforms above are uses of PISA on a policy level, this paper leaves behind the statistical assumptions that inform policy to offer, instead, an empirical investigation of the effects of PISA from a pedagogical (classroom) perspective. More specifically, it describes a study that seeks to understand the interaction between the students and the items used for testing. It assumes that observation of problem-solving in action can give us information such as: What impediments or difficulties develop in students' encounters with specific test questions? What meanings are offered by and produced from the science problems as presented in the test? The focus of these questions is more on the validity of the data.

A Design to Explore Test Items in Scientific Literacy

To explore these questions, a study was designed in which 15-year-old students collaboratively answered 11 PISA test questions in scientific literacy (OECD 2007) during a science lesson (this design is thoroughly described in, e.g. Serder and Jakobsson 2015). Three PISA units were included: Greenhouse, Acid Rain and Sunscreens (PISA units S114, S447, S485). In total, 21 groups of 3–4 students were formed. This collaborative design was chosen based on a sociocultural understanding of knowledge construction (Wertsch 1998) which emphasizes that knowledge/knowing is shaped in action and obtains meaning from real-life situations. The interactions were video recorded producing 16 h of video data for further observation and semantic analysis (Mäkitalo et al. 2009). The focus for the analysis presented here was what specific problems the students experienced with the test questions.

A Stereotyped Portrayal of Science to Resist and Terms with Hybrid Meanings

In the analysis of the data, two main themes were discerned: (1) how the students discussed science as it was portrayed in the test items and (2) how words and formulations of the test items were used in the student conversations. PISA scientific literacy test items are required to address “real-life issues” (OECD 2012: 102) which infer that the problems are contextualized in everyday life situations. According to the analysis, this condition seems to affect how the students in this study approached the problems. A common difficulty for the groups was to identify the intended meaning of various words. Words with hybrid meaning, that have different meanings in different contexts, were found to be frequently negotiated by the students (Serder and Jakobsson 2016). The Swedish words, e.g. *pattern*, *factor*, *reference*, *constant* and *better* are examples from the study, all with differing

meanings depending upon whether they are used in an everyday, science or mathematical context. In order for the groups to respond successfully in the test, it appeared to be crucial to ignore all alternative, possible contexts – including the inferred everyday context. Two examples from the observed student conversations are the word “pattern” that could be discussed as a mathematical term (in the sense *regularity*) instead of as (intended) a result of a scientific experiment and the word “factor” used to denote a “sun cream”, instead of a scientific variable. The results also indicate that some meaning is likely to be added, or lost, in translation (Serder and Jakobsson 2016). The second analytic theme concerned meaning in a different sense, namely, the meaning of science itself. The students tended to discuss the manner in which the fictive characters of the test items were speaking and approaching the scientific problems presented to them in their imaginary everyday life. While so doing, the students often expressed resistance towards the image of science that they were presented and interacted with. However, in order to approach the test questions productively, the students need to accept the artificial aspects of the “real-life” problems, as well as the authoritative, highly stereotyped way in which science was (implicitly) portrayed in those items (Serder and Jakobsson 2015).

Use or Misuse?

The group situations used for the purpose of this study were different from individual testing situations. However, the design permits an insight in the interactions that possibly take place in the test situation, of which very little is known. The reasons for the decreasing Swedish results are likely to be various. This research poses questions about comparability between different national versions of the PISA test because the hybrid meanings of the words that were negotiated by the student groups are unlikely to be overlapping in different languages (see Arffman 2010). This is a finding that merits further exploration. As for the testing of “real-life skills”, the work shows that the everyday aspects of the test questions invite a number of various discourses and meanings. The conclusions support previous research that suggests that scientific competence is intimately linked to students’ discursive knowledge (Norris and Phillips 2003; Olsen 2012). Moreover, the inferred everyday contexts in PISA seem to enhance the way science gets portrayed as a very particular – and peculiar – human culture with certain norms and values (cf. Aikenhead 1996). For instance, the students questioned “everyday” situations inferred in the test items such as conducting experiments outside school, putting marble chips in vinegar or discussing graphs in a library (Serder and Jakobsson 2015).

Sweden may be an exceptional case, both in the decline in its PISA scores and the number of rapid education reforms in recent years. Unfortunately, there might be little correspondence between national school reforms launched to address the political anxieties of *PISA shock* and the actual problems lived by students and teachers in the school system that need addressing. In order to reduce misuse of the PISA scores, these findings raise the question of whether there is a need for the scholarly community to stress less the need to attend to the general outcomes and

superficial interpretations of the results and attend more to a fine-tuned analysis of the results and their implications. This study, in which the interactions of students and test items were studied, was an attempt to show what such analyses reveal and why they might be valuable.

Postscript

What conclusions then can be drawn from these somewhat disparate perspectives? The first is to recognize that there are both positive and negative aspects to PISA. The positive is that this international test attempts to measure competency – not just knowledge and understanding but the ability to use scientific knowledge to undertake science-specific tasks which might reasonably be expected of a science literate 15-year-old. These competencies are defined in a manner which represents the best thinking about what should be the outcomes of a contemporary science education. Moreover, the testing for PISA is undertaken in 72 countries in a manner which is as rigorous and systematic as can be given the constraints of producing one test in multiple languages in a short time scale. As a consequence, it produces an enormous quantity of data which raises issues for individual countries rather than definitive answers. Embedded within the data are some clear trends and patterns. For instance, Canada, Estonia, Germany and Hong Kong (China) have all attained high levels of performance with high or improving levels of equity in the 2015 results. What these countries would appear to share to a greater or lesser degree is (OECD 2016: 7):

1. A clear education strategy to improve performance and equity.
2. Rigorous and consistent standards are applied across all classrooms.
3. Teacher and school leader capacity has been improved.
4. Resources are distributed equitably across schools – preferentially to those schools and students that need them most.
5. At-risk students and schools are proactively targeted.

If such findings succeed in raising questions or providing pointers to where we might concentrate our efforts to improve the quality of the education that each country might offer its young people, then PISA has value. If, in contrast, PISA is simply seen as part of an international educational arms race from which little can be learnt, then all will have been in vain.

References

- Aikenhead, G. S. (1996). Science education: Border crossing into the sub-culture of science. *Studies in Science Education*, 27, 1–52.
- 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.

- Allerup, P. (2007). Identification of group differences using PISA scales – Considering effects of inhomogenous items. In S. T. Hopmann & G. Brinek (Eds.), *PISA zufolge PISA – PISA according to PISA* (pp. 175–201). Vienna: Universität Wien.
- Arffman, I. (2010). Equivalence of translations in international reading literacy Studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59.
- Ball, S. J. (2012). *Global education Inc: New policy networks and the neo-liberal imaginary*. London: Routledge.
- Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance, OECD Education working papers, No. 71*. Paris, France: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/5k9fdqffr28-en>
- Confederacion de Sociedades Cientificas de España. (COSCE). (2011). Informe ENCIENDE: Enseñanza de las Ciencias en la Didáctica Escolar para edades tempranas en España. [Report ENCIENDE: Teaching science in school education for early ages in Spain]. Madrid, Spain: COSCE. Retrieved from http://www.cosce.org/pdf/Informe_ENCIENDE.pdf
- Davidsson, E., Karlsson, K.-G., & Oskarsson, M. (2013). Trender och likvärdighet: Svenska elevers resultat på PISA naturvetenskap i en internationell jämförelse [trends and equivalence: Swedish students' performance in the PISA science in an international comparison]. *Utbildning & Demokrati* 2013, 22(3), 37–52.
- Fensham, P. J. (2009). Real world contexts in PISA science: Implications for context-based science education. *Journal of Research in Science Teaching*, 46(8), 884–896.
- Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267–321.
- Labaree, D. F. (2014). Let's measure what no one teaches: PISA, NCLB, and the shrinking aims of education. *Teachers College Record*, 116(9), 1–14.
- Lederman, N. G. (2007). Nature of science: Past, present and future. In S. Abell, & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–879). Mahwah: Lawrence Erlbaum.
- Lundahl, L., Erixon Areman, I., Holm, A.-S., & Lundström, U. (2013). Educational marketization the Swedish way. *Education Inquiry*, 4(3).
- Lundgren, U. P. (2011). PISA as a political instrument. In M. A. Pereyra, H. G. Kotthoff, & R. Cowen (Eds.), *PISA under examination. Changing knowledge, changing tests, and changing schools*. Rotterdam: Sense publishers.
- Mäkitalo, Å., Jakobsson, A., & Säljö, R. (2009). Learning to reason in the context of socioscientific problems. Exploring the demands on students in 'new' classroom activities. In K. Kumpulainen, C. Hmelo-Silver, & M. Cesar (Eds.), *Investigating classroom interaction. Methodologies in action* (pp. 7–26). Rotterdam: Sense Publishers.
- Meyer, H.-D., Tröhler, D., Labaree, D. F., & Hutt, E. L. (2014). Accountability: Antecedents, power, and processes. *Teachers College Record*, 116(9), 1–12.
- Millar, R., & Osborne, J. F. (Eds.). (1998). *Beyond 2000: Science education for the future*. London: King's College London.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public Management reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–353.
- Münch, R. (2014). Education under the regime of PISA & co.: Global standards and local traditions in conflict—The case of Germany. *Teachers College Record*, 116, 1–16.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Norris, S. P., & Phillips, L. (2003a). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87, 224–240.
- OECD. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD Publishing.

- OECD. (2007). *PISA 2006: Science competencies for tomorrow's world. Volume 1: Analysis*. Paris: OECD Publishing.
- OECD. (2012). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.
- OECD. (2014a). *PISA 2012 results: What students know and can do*. (Vol. I, Rev. edn, February 2014) [Electronic resource]: OECD publication.
- OECD. (2014b). *Resources, policies and practices in Sweden's schooling system: an in-depth analysis of PISA 2012 results*. Retrieved from http://www.lcis.com.tw/paper_store/paper_store/86952f1a-2015226205027484.pdf
- OECD. (2016). *Country note: Key findings from PISA 2015 for the United States*. Paris, OECD Publishing.
- OECD & Pearson. (2014). (Joint press release) *Pearson to develop PISA 2018 Student Assessment 21st Century Frameworks for OECD*. <https://www.pearson.com/news/announcements/2014/december/pearson-to-develop-pisa-2018-student-assessment-21st-century-fra.html>
- Olsen, R. V. (2012). An exploration of cluster structure in scientific literacy in PISA: Evidence for a Nordic dimension? *Nordic Studies in Science Education*, 1(1), 81–94.
- Oskarsson, M., Eliasson, N., & Karlsson, K. (in progress). *School science in a market-driven school system*.
- Östh, J., Andersson, E., & Malmberg, B. (2013). School choice and increasing performance difference: A counterfactual approach. *Urban Studies*, 50(2), 407–425.
- Ringarp, J., & Rothland, M. (2010). Is the grass always greener? The effect of the PISA results on education debates in Sweden and Germany. *European Educational Research Journal*, 9(3), 422–430.
- Rychen, D. S., & Salganik, L. H. (Eds.). (2003). *Definition and selection of key competencies: Executive summary*. Göttingen: Hogrefe.
- Schleicher, A. (2012). *Use data to build better schools*. TEDGlobal. Available at http://www.ted.com/talks/andreas_schleicher_use_data_to_build_better_schools?language=en
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss* (Jahrgangsstufe 10). Berlin, Germany.
- Sellar, S., & Lingard, B. (2013). PISA and the expanding role of the OECD in global educational governance. In H. Meyer & A. Benavot (Eds.), *PISA, power and policy* (pp. 185–206). Oxford: Symposium Books Ltd..
- Serder, M., & Jakobsson, A. (2015). “Why bother so incredibly much?”: Student perspectives on PISA science assignments. *Cultural Studies of Science Education*, 10(3), 833–853.
- Serder, M., & Jakobsson, A. (2016). Language games and meaning as used in students encounters with scientific literacy test items. *Science Education*, 100(2), 321–343.
- Sjøberg, S. (2016). OECD, PISA, and globalization. The influence of the international assessment regime. In C. H. Tienken & C. A. Mullen (Eds.), *Education policy perils. Tackling the tough issues*. New York: Routledge.
- Skolverket. (2012). *Educational equity in the Swedish school system? A quantitative analysis of equity over time*. Retrieved from <http://www.skolverket.se/>
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7–25. doi:10.1207/s15324818ame2001_2.
- Wertsch, J. V. (1998). *Mind as action*. New York: Oxford University Press.