

Svein Sjøberg is professor emeritus in science education at Oslo University. His current research interests are the social, ethical and cultural aspects of science education, in particular the impacts and influence of large scale assessment studies like PISA and TIMSS. He has worked extensively with international and comparative aspects of science education through e.g. UNESCO, OECD, ICSU, the EU and the European Commission and has won several prizes and awards for his research, teaching, and promotion of science literacy and public understanding of science.

SVEIN SJØBERG

Department of Teacher Education and School Research (ILS), University of Oslo, Norway
svein.sjoberg@ils.uio.no

The power and paradoxes of PISA: Should Inquiry-Based Science Education be sacrificed to climb on the rankings?

Abstract

Since publication of the first PISA results in 2001, the PISA scores have become a kind of global “gold standard” for educational quality. Climbing on the international PISA rankings have become a high priority for national educational policies world-wide, also in the Nordic countries. This article first explores why and how the OECD, with PISA as the main instrument, has emerged as the key defining organization for educational quality and policy. Some of the underlying assumptions, ideologies and values are critiqued. Secondly, the article draws attention to PISA findings that are surprising, unexpected and problematic. The most problematic finding for science education is that PISA-scores correlate negatively with nearly all aspects of inquiry-based science teaching (IBSE), the kind of teaching that is recommended by scientists as well as science educators.

INTRODUCTION: PISA AS A GLOBAL STANDARD

International large-scale comparative studies of students' achievement, often labelled ILSA, have existed since the 1960's, mainly organized by the IEA (International Association for Educational Achievement). Acronyms like TIMSS (Trends in Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) have become widely known, also outside the education sector. Results are reported in the media, and they have over the decades been used by policymakers to inform national reforms and initiatives.

The influence of ILSAs increased dramatically when the OECD (Organization for Economic Cooperation and Development) launched its PISA (Programme for International Student Assessment) project in 2000.

On the basis of broad analysis and reviews of educational reforms and development, an OECD- report on the policy impact of PISA states that

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

(Breakspear, 2012)

Similarly, Andreas Schleicher, director of PISA and recently also of the Directorate of Education and Skills in OECD, in a well visited 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.”* (Schleicher, 2013).

PISA-leader Andreas Schleicher in TED-talk 615,228 views

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”*



http://www.ted.com/talks/andreas_schleicher_use_data_to_build_better_schools?language=en

Figure 1. PISA leader, Andreas Schleicher (2013), presenting what PISA is really about in a TED-talk.

PISA has increasingly influenced the education debate and policy in nearly all participating countries. The three Scandinavian countries experienced “PISA-shocks” that have led to major reforms as well as conflicts. These are treated in articles, for instance by Serder and Jakobsson (2015, 2016) and Sjøberg (2014) as well as books, like the Norwegian anthology edited by Elstad and Sivesind (2010), the Swedish by Landahl and Lundahl (2017) and the Danish/Nordic anthology by Klitmøller and Sommer (2018).

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 and “economies”. The intentions of PISA are, of course, related to the overall aims of the OECD and its commitment to a competitive global free market economy. When PISA is presented, its importance is stated by claiming that participation *“make up ninetyths of the world economy.”* (OECD, 2010a, p 3). This is a surprising way of counting pupils, but it indicates the focus of the PISA-project: the economy.

This article first explores why and how the OECD, with PISA as the main instrument, has emerged as the key defining organization for educational quality and policy. It will look critically into some basic features of PISA as a political and normative project and examine how the OECD exerts its global influence through its statistics and a well-planned media policy.

Secondly, the article draws attention to PISA findings that are surprising, unexpected and problematic. The most problematic finding for science education is that PISA-scores correlate negatively with nearly all aspects of inquiry-based science teaching (IBSE), the kind of teaching that is recommended by scientists as well as science educators. Whether one "believes in PISA" or not, these results need to be discussed critically by science educators.

PISA AS A POLITICAL AND NORMATIVE PROJECT

The great influence of PISA mainly stems from the status and authority of the OECD, an organization that is "owned", governed and financed by the member states. The political nature of PISA is evident in the organization of the project: The PISA Governing Board is composed of representatives of OECD members, appointed by their education ministries. The Board determines the policy priorities for PISA and makes sure that these are respected during the implementation of each PISA survey.

The OECD expressed the aims and ambitions of PISA already before the first PISA testing:

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

(OECD, 1999, p 7)

These words have been repeated in later PISA reports. If the above aims of PISA had been stated as "research questions" in a research proposal, it is unlikely to have passed a peer review process. But this OECD initiative is politically determined, and not meant for peer review by researchers.

In the first report from PISA/OECD, the joint commitment of the OECD owners was clearly stated: "*PISA represents a new commitment by the governments of OECD countries to monitor the outcomes of education systems in terms of student achievement, within a common framework that is internationally agreed.*" (OECD, 1999, p11). In later reports, the normative nature of PISA is even more explicit. The PISA 2009 report states that PISA "*provides a basis for international collaboration in defining and implementing educational policies.*" (OECD, 2010a, p 3).

The Swedish professor Ulf P. Lundgren played a major role in the preparations of PISA. Some 10 years later, he reflects on what he calls "*PISA as a political instrument*" and concludes that

PISA is an example of what in a global world nationally is perceived as the answer to what is going to be taught, who it is going to be taught to and how the outcomes of teaching will be judged and used for control and political governing.

(Lundgren, 2011)

The above cited ambitions clearly distinguish PISA from other ILSAs, in particular TIMSS, which is also testing achievement in mathematics and science. It is important to remember that PISA is explicitly *not* testing school knowledge or testing according to the country's curriculum, but "*within a common framework that is internationally agreed.*" TIMSS, on the other hand, tries to test achievement according to the (common elements of) national curricula. Moreover, TIMSS items are often

similar to those used in more traditional exams and tests. TIMSS is descriptive and analytical, has no political or other “agenda”, and researchers have always played a major role in its development and reporting. PISA, on the other hand, is political and normative, and has the ambition to influence education policy according to contents, norms and values as defined by the OECD, mainly as preparation for the global labour market. It should be noted, however, that also the IEA have become more politicized in recent years. Governments and their education ministries are IEA member organizations, provide the funding and sit on the boards. In 2018 the IEA board is chaired by a representative from the Norwegian Directorate of Education.

The PISA project can be seen as part of a wider international policy trend, where common standards is a prerequisite, and where concepts and ideas from the market economy are used in the education sector. The most visible aspect of PISA is its focus on league tables and numerical scores. The published rankings create competition, where there are winners and losers. The countries at the top are celebrated at “successful”, and PISA reports hold them up as winners and models. Everything seems to centre on having success: PISA reports celebrate successful systems, successful schools, successful reformers, successful learners (OECD, 2010a, 2015a and 2016c). The underlying belief is that competition in a market always generates quality and success. In the PISA perspective, the ultimate meaning of schools and education is to prepare the individual for active participation in this competition. It is this highly political, normative and narrowing nature of PISA that makes it different from most other ILSA studies.

PISA as “Big Science”

PISA is a huge multinational undertaking. In a way, it may be compared to what is known as Big Science in modern science research. Like NASA and CERN, PISA involves thousands of people from a multitude of nations, with several external contractors and national teams. The PISA 2015 Technical Report (OECD, 2017, p 456-463) provides names and affiliations of some 600 people involved in just the central, international running of PISA: politically appointed boards and expert committees of various kinds. In addition comes the many hundreds of people involved in the 70+ national PISA teams and the people they employ for coding and other work. Tens of thousands of principals and teachers are also involved, and of course the more than half a million students that take the test.

The reports from PISA are voluminous, each of them with dozens of authors. The six key international PISA 2015 reports (<http://www.oecd.org/pisa>) are each about 500 A4-size pages, densely packed with information, and supported with links to the PISA-databases for download of data for details. The PISA 2015 Technical report (OECD, 2017a) was published long time after the release of results, and is also 500 pages, compact with details about the technicalities of the project development, data collection and statistical analysis. Even for people who are fulltime employed in PISA, it must be difficult to keep abreast.

In addition to the voluminous international reports, each country produces national reports, and PISA produces a multitude of more specific newsletters and policy briefs. In the light of this richness and complexity of data gathering and analysis, it is noteworthy that it is the resulting league table of PISA winners and losers that attracts most attention.

Selling the PISA message

The official release of PISA results is a well-planned and synchronized global media event. When PISA results are presented in public, the audience will receive light versions of the voluminous PISA reports, often in the form of ready-made media packages. This reporting is targeting the media, politicians and policymakers. These PISA products are glossy and colourful, well written, with simple messages, conclusions and recommendations. Presentation videos and interactive data animations are also made available, often in close cooperation with commercial providers, the largest being Pearson Inc.

the world's largest provider of educational tests, textbooks, services and private schools. Pearson is a key partner also in the development of PISA 2018. The invitations to the PISA press briefings and the release of reports clearly state that the PISA results should be seen as indicators for the future economic competitiveness. The well-arranged PISA release is described as follows in an in-depth study:

A few weeks before the release of PISA results in December of the year following that in which the test was administered, notification is sent to a list of approximately 4500 journalists globally, alerting them to the imminent publication of PISA results. The media team then works with the Education Directorate on a choreographed release of PISA results. (Addey et al., 2017, p14)

The resulting country rankings of mean PISA-scores receive most of the attention, and create what Sellar, Thomson and Rutkowski (2017) describe as a “a global education race”. They strongly warn against trying to copy the celebrated PISA-winners. The final conclusion in their most readable book is that “PISA envy is toxic” (ibid, p 99).

PISA and Human Capital production

The focus on the economy is a key for understanding the extreme importance that is now attributed to the PISA rankings: it seems “common sense” that high scores on reading, mathematics and science are predictors for the country's future economic competitiveness. Hence, low rankings on PISA are assumed to be bad signals for the future of the nations' competitive economy. This is explicitly stated in the international PISA reports as well in the press briefings for the release of results.

Important underpinnings regarding the importance of international test results for economic prosperity are the works of the US economist Eric Hanushek. Over decades he has published extensively on the relationship between economic investment and educational quality, and is widely used by the World Bank and the OECD. With his companion, the German professor Ludger Woessman, he authored the OECD report on “*The long run Economic Impact of Improving PISA Outcomes*” (OECD, 2010b). In this report, they provide numbers on how much each country will earn on improving the national PISA-score. They provide different scenarios for the implications of different magnitudes of PISA improvements.

Concretely, they assert that an increase in 25 PISA points (a quarter of a standard deviation) over time will increase the Danish GDP with 586 billion USD (OECD, 2010b, p23). Norway would earn 841 billion USD and Sweden 1019 billion. They also claim that if Germany improves its PISA-score to the level of Finland, “*Germany would see a USD 16 trillion improvement, or more than five times current GDP. All of these calculations are in real, or inflation-adjusted, terms.*” (ibid, p25).

These and other findings based on Hanushek's economic modelling have for decades been strongly rejected by scholars from many academic fields, simply by discarding the validity of such projections. They claim that no social science can make predictions like these. Recently, also the calculations are challenged in an article that asserts that they are based on invalid statistics. For an academic article, the title is sharper than one often sees, even naming the target for the critique: “*A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth.*” (Komatsu & Rapple, 2017). The authors have used exactly the same data, and come to completely different results. They end the article by stating that “*Our intent is to contribute to a more rigorous global discussion on education policy, as well as call attention to the fact that the new global policy regime is founded on flawed statistics.*” (Komatsu & Rapple, 2017).

Expanding and extending PISA

Seen from the OECD, PISA has been a remarkable success. By providing rankings and indicators based on its data, PISA sets the scene for discussions about quality of schooling and entire school systems. And in most countries, politicians and policy-makers follow suit. Given this success; it is easy to understand that the OECD is also broadening its scope and influence in the education sector with other “PISA-like” studies, ranging from kindergarten to retirement age, from the national level to school level, and from highly developed OECD countries to developing countries. Examples follow.

“**Starting Strong**” is one of several OECD-programs to address preschool/kindergarten level (ECEC: Early Childhood Education and Care), also by comparing attainments and competencies and the return of investments in early child care (OECD, 2017b).

“**PISA-based test for Schools**” is a PISA-like test that may be used to test how well a school or school district compares with each other or with the PISA-winners. This may thereby bring the power of influence on policies closer to school districts, local authorities and even particular schools and their teachers. The product is commercially available in the USA, UK and Spain (OECD, 2018c).

“**PIAAC, Survey of Adult Skills**” (often called “PISA for adults”) is measuring skills and competencies of the adult work-force (16-65 years), on a scale similar to the PISA scale. The survey measures adults’ proficiency in key information-processing skills - literacy, numeracy and problem solving in technology-rich environments - and gathers information and data on how adults use their skills at home, at work and in the wider community. In each country, a representative sample of about 5 000 are interviewed in face-to-face settings. Some 40 countries took part in the first testing round, and data are published and available in many formats, see for instance (OECD, 2016d).

“**PISA for Development**” is a version of PISA that is meant to be used by low- and middle income countries. It will do this using *“enhanced PISA survey instruments that are more relevant for the contexts found in middle- and low-income countries but which produce scores that are on the same scales as the main PISA assessment”*. In this project, the OECD also defines supposedly globally valid competencies that are needed for young people in all developing countries. Results are likely to be used as benchmarks for development assistance from the World Bank and other donors. PISA for Development publishes regular policy briefs with progress reports and findings. (OECD, 2018a).

Education at a glance: OECD Indicators is an annual publication that brings indicators and statistics from the OECD/PISA and other sources, and is widely used by policymakers and researchers world-wide. Data are also available in different formats (like Excel) to be downloaded for analysis. The reports provide *“key information on the output of educational institutions; the impact of learning across countries; the financial and human resources invested in education; access, participation and progression in education; and the learning environment and organisation of schools.”* (OECD, 2017c).

Over the years, the OECD has become a key global provider of statistics, not only for the economy, but also in the education sector. The OECD statistics is increasingly being used by other global actors, including the European Union, the World Bank and gradually also UN-organizations like UNESCO.

The main concern from the many critics of the OECD and PISA is that school policies, which used to be the core aspect of all nations’ cultural heritage, values and identity has become the subject of external and global influence, where the prime perspective is the economic competitiveness in a harsh global race. This implies a de facto redefinition and narrowing of the purpose and meaning of schools.

Governance by numbers, statistics and indicators

The OECD has no formal legal power, but exerts what scholars describe as “soft power”, they “govern by numbers and indicators” (Grek, 2009).

Good and reliable statistics is, of course, important. But statistics and indicators do not simply describe reality, they construct and shape reality. What you choose to measure also defines what is seen as important. How you construct an indicator builds on underlying assumptions and value-based priorities that are soon forgotten when league tables are constructed and presented.

Simon Breakspear is making the same point in a report with the telling title *“How does PISA shape education policy making? Why how we measure learning determines what counts in education.”* (Breakspear, 2014).

The influence of PISA on educational policy takes many forms, but they all rest on the use of comparisons, rankings, statistics and indicators. Xavier Pons (2017) has provided a critical review of research on PISA effects on education governance. The review exploits a dataset of 87 references which show that PISA introduced major changes in the governance of education worldwide. He provides several case studies of how national reforms have been triggered and legitimized by reference to perceived “PISA-shocks”.

One example is Australia; where the prime minister, Julia Gillard, in 2012 stated that *“The government will use PISA [...] to track Australia’s progress compared with the rest of the world. By 2025, Australia should be top five in the world.”* (The Australian, 3. Sept 2012)

Another example is Norway. The day of the release of PISA 2015 results, the Norwegian minister of education, Torbjørn Røe Isaksen wrote the following in a leading Norwegian newspaper:

*Waking up this morning, I knew that this was one of the most important days in my life as minister of Education: **The PISA-day.** [...] Right or wrong, PISA results define a government’s period as success or failure, and this was the day!*

(Røe Isaksen, Morgenbladet, 6 Dec 2016, my translation)

Governments in other countries have made similar statements, stating that climbing on PISA-rankings is the main goal for their schools. (Sjøberg, 2015; Pons, 2017).

The PISA reporting of data, in particular the normative conclusions and recommendations, has been met with strong criticism, also from researchers involved in PISA and other ILSAs. Rutkowski and Rutkowski (2016) urge PISA to have a *“more measured approach to reporting and interpreting PISA results”*. A main point in the critique is that PISA reporting makes causal claims, while the research design does not allow causal inferences about neither “success” nor “failure”.

The PISA scores and league tables get most of the attention in the media as well as from policy-makers (Steiner-Khamsi, 2003). Less attention is given to the details and to the many surprising and paradoxical results. In the following, we go in some detail on results of importance to science education.

PISA, SCIENCE LITERACY AND INQUIRY

“Science literacy” was one of the three core components of the PISA test already from the first PISA-round in 2000. In the two first rounds of PISA, the definition of science literacy was rather short, but when science in 2006 was for the first time the core subject of PISA, the definition was developed and expanded considerably.

In 2015, when science for the second time was the core PISA domain, the definition was changed, further elaborated and extended. But when the definition of science literacy changes over time, this becomes problematic, since measuring trends is a key purpose of PISA. When you want to measure change, you cannot change the measure. Nevertheless, trends for PISA science literacy are often provided on the same graphs already from the first PISA round until the present. The PISA reports make readers aware of the changing definitions in footnotes, but most readers do probably not read these.

The new definition of science literacy for the PISA 2015 is given in short and more detailed versions. In the report that provides the first PISA 2015 results, the following definition is given:

***Science literacy is defined** as the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific inquiry, and interpret data and evidence scientifically.*

(OECD 2016b, p. 28)

This short version is expanded in the Assessment and Analytical Framework (OECD, 2016a). Well known science education researchers were involved in this work. The science literacy expert group was chaired by Professor Jonathan Osborne, and included more than a dozen other science educators, among them Robin Millar, Andrée Tiberghien, Russell Tytler, Rodger Bybee, Harrie Eijkelhof and Jens Dolin from Denmark. It may, however, seem a little strange to read that *“The framework for assessing science was developed by the scientific literacy expert group with the guidance of [4 names are given] from Pearson”* (OECD, 2016a, p 3). Pearson is, as noted, is the world’s largest commercial provider of educational services and products. It is no doubt that the close relationship with PISA gives Pearson a strong competitive advantage on the ever-growing global edu-business market. (Hogan, Sellar and Lingard, 2016).

In refining and extending the PISA definition of science literacy the expert group base their arguments and conclusions on important contributions to the “canon” of philosophy of science and the field of science education. The PISA definition also builds on previous reviews of the concept science literacy (Roberts, 2007), and on policy-oriented recommendations from working groups from the science education community, as provided in for instance Millar and Osborne (1998) and Osborne and Dillon (2008).

Most science educators are likely to find the selection of background material, the discussions and the resulting clarification to be of great interest. It may serve as a valuable source for discussions regarding the purpose and contents of school science.

The PISA 2015 definition of science literacy includes affective dimensions, and they are explicit on this: *“Attitudes form part of the **construct** of scientific literacy. That is, a person’s scientific literacy includes certain attitudes, beliefs, motivational orientations, self-efficacy and values.”* (OECD, 2016a, p36).

The framework clearly distinguishes between attitudes *towards* science (like being interested in science) and *scientific attitudes* (like valuing empirical evidence as the basis of belief).

The PISA 2015 assessment evaluates students’ attitudes *towards* science in three areas: interest in science and technology, environmental awareness, and valuing scientific approaches to inquiry.

Some affective elements were already included in the PISA 2006 science test as “embedded” in the science test units. The students were asked whether they were interested in the topic of the unit and

whether they wanted to learn more about the topic. On the basis of such responses, a construct of “Interest in science” was calculated. The responses to these questions were, however, not included when calculating the PISA score. In fact, the results for the interest score were somewhat surprising, and “revealed a negative correlation between student science topic interest and science performance at the country level.” (Bybee and McRae, 2011). Many countries with high PISA score came out low on the interest score. As Bybee and McRae (2011) also notes: “Finland, the country with the highest mean science achievement score in PISA 2006, was the lowest scoring country on the interest scale.” Such results should cause some concern when discussing educational implications of PISA, especially when it comes to learning from the winners.

For different reasons, in PISA 2015, attitudinal aspects like interest in science were no longer “embedded” in the test units, but placed in the student questionnaire. Given the comprehensive PISA 2015 definition of scientific literacy, a reader might assume that the PISA science literacy score actually includes all aspects mentioned. But, although they are part of the definition, they are *not* addressed by the test items; the published PISA science literacy score is based solely on the results from test items.

Attitudinal aspects and questions related to teaching and learning experiences, attitudes to science, epistemic beliefs etc. are, however, addressed in the PISA student background questionnaire. On the basis of the questionnaire, certain constructs that are part of the above science literacy definition are calculated. PISA reports provide the details, and also how these constructs relate to the PISA test score (OECD, 2016c). Some of these results are surprising and problematic and are explored in the following.

Inquiry in science education

The concept of inquiry is central when philosophers describe what science is all about. In fact, science and inquiry are often listed as synonyms in dictionaries. Inquiry can be a noun, the result of an investigation, and also a description of the process that leads to knowledge and insights. The term inquiry was central in John Dewey’s (1938) philosophy of education, and my own first encounter with physics education literature was through Eric Roger’s fascinating book *Physics for the Inquiring Mind* (1960) – still worthy a revisit!

Understanding the *nature of science*, often with the acronym of NOS, has long time been an important issue in science education (Lederman, 1992, 2006). Sometimes NOS is described as knowledge *about* science, as a contrast to knowledge *in* science. When stressing knowledge *about* science, it is often to draw attention to science as a human product and a dynamic process, often contrasted to considering science as a mere pile of established facts, laws and theories.

In science education, the concept of inquiry can be manifested when learning about the nature of science, and most often when we focus on the process dimension of doing science. Science curricula in many countries stress that students should learn about how scientific knowledge is developed, constructed, validated and tested through systematic inquiry. Students are also supposed to develop their epistemological understanding of science as a discipline whose knowledge claims are based on inquiry and in principle are tentative, fallible and open for scrutiny.

In school science, inquiry may also be manifested in the methods of teaching and instruction. A basic idea that the learners should “act like scientists”: they are supposed to formulate ideas to be tested, design and carry out experiments, discuss the findings and draw conclusions. By doing this, students will hopefully develop an understanding of the processes as well as the contents of science. So – by themselves working more or less like scientists, the students are supposed to learn the science contents as well as improving their understanding of the nature of science as a process and activity.

Inquiry-based teaching has always been a valued aspect of science education. In recent years it has again been brought into use again, as if it were a newcomer. *Inquiry-Based Science Education*, IBSE, is now a well-known acronym for science educators, and appear in policy-documents from a variety of institutions and organizations. IBSE is the key term in the influential EU-document *Science Education NOW: A renewed pedagogy for the future of Europe* (EC, 2007), chaired by Michel Rocard, former Prime Minister of France. The report claims that “*IBSE is effective with all kinds of students from the weakest to the most able and is fully compatible with the ambition of excellence. Moreover IBSE is beneficial to promoting girls’ interest and participation in science activities.*” (EC, 2007, p2)

Based on this report, the term IBSE became a key concept in calls for EU-funding for science education research and initiatives in Frame Programme 7 as well as for the current Horizon 2020 -programme. IBSE also plays a major role in the recommendations in the International Council for Science report to the Science Unions world-wide (ICSU, 2011). IBSE is also the key idea in the science educational initiatives of ALLEA, the European Federation of Academies of Sciences and Humanities, which brings together Academies in more than 40 European countries. The joint programme statement for science education claims that “*IBSE is a form of science education that – unlike the traditional model where the teacher provides facts and the students learn them – gives children the opportunity to explore “hands on”, to experiment, to ask questions and to develop responses based on reasoning*” (<http://www.allea.org>, visited March 05, 2018)

As we can see from the above, IBSE is promoted for many different reasons and from many different stakeholders. It is emphasized that active involvement in investigations and discussions is in line with basic ideas about the nature of science as a discipline, and also with the tenets of a constructivist view on learning. Moreover, it is expected that IBSE leads to stronger personal involvement and higher interest in science, in particular for girls. IBSE is also assumed to enhance the respect for science and the motivation to choose science-related studies and occupations.

In summary, we may identify three different reasons for recommending IBSE:

1. IBSE is an efficient way of teaching and *learning science contents*
2. IBSE is instrumental for learning about *the nature of science*
3. IBSE will improve students’ *joy, interest and motivation towards science*

To which degree IBSE lives up to these three expectations is an empirical question. PISA data might shed light on this issue.

IBSE and PISA-score

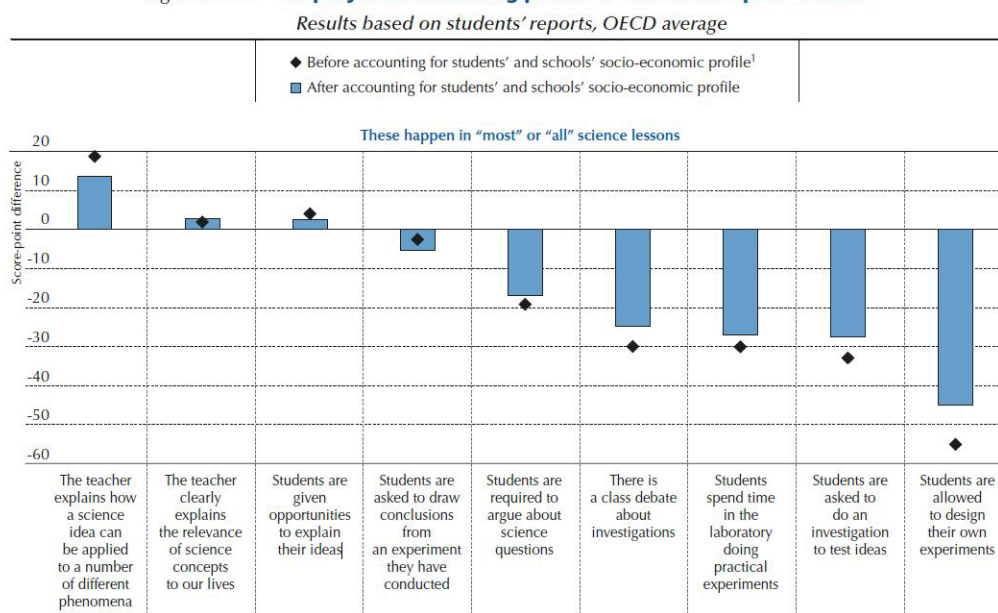
In PISA 2015, nine statements in the student questionnaire are meant to measure to which degree the students have taken part in inquiry-based teaching. The questions are the following:

Students are given opportunities to explain their ideas
Students spend time in the laboratory doing practical experiments
Students are required to argue about science questions
Students are asked to draw conclusions from an experiment they have conducted
Students are allowed to design their own experiments
Students are asked to do an investigation to test ideas
There is a class debate about investigations
The teacher clearly explains the relevance of science concepts to our lives
The teacher explains how a science idea can be applied to a number of different phenomena
 (OECD 2016c, p 242)

These questions are answered on a 4-point Likert scale and are combined to an *index of inquiry-based instruction*. This index is meant to “measure the extent to which science teachers encourage students to be deep learners and to inquire about a science problem using scientific methods, including experiments.” (ibid, p242).

The relationship between IBSE-index and the PISA score is provided in various ways in OECD, 2016c. Figure 2 show how each of the nine components on this index relate to the PISA score.

Figure II.2.20 ■ **Enquiry-based teaching practices and science performance**



1. The socio-economic profile is measured by the PISA index of economic, social and cultural status.

Note: All differences are statistically significant (see Annex A3).

Source: OECD, PISA 2015 Database, Table II.2.28.

Figure 2. The different components of the IBSE-index and the PISA-score. (OECD, 2016c, p 262)

Here we see that the overall pattern is a negative relationship between the components of the IBSE-index and PISA-score. The main exception is the component “The teacher explains how a science idea can be applied to a number of different phenomena.” One might argue that this component should rather belong to another index, the “Index of teacher-directed instruction”. The same may be the case for the component “The teacher clearly explains the relevance of science concepts to our lives.”

This Index of teacher-directed instruction varies strongly between countries, and high scoring and low-scoring countries are interestingly at both ends of this spectrum: In Singapore and Finland students report a much higher teacher-directed value than the OECD average, while Korea and Japan have low teacher-directed index. The most interesting and indeed challenging result is that within *all* countries, teacher-directed teaching correlates *positively* with the students' PISA test scores. (OECD, 2016c, p 64).

IBSE and the Nordic countries

Figure 3 shows the countries sorted by the index of inquiry-based instruction with country means based on students' reports. At the very bottom we find the high-scoring PISA countries like Japan, Korea, Taiwan and Finland. In these countries students are barely exposed to IBSE. Norway is around the OECD average, while Sweden and in particular Denmark are among the highest OECD-countries on the list. (OECD, 2016c, p 72).

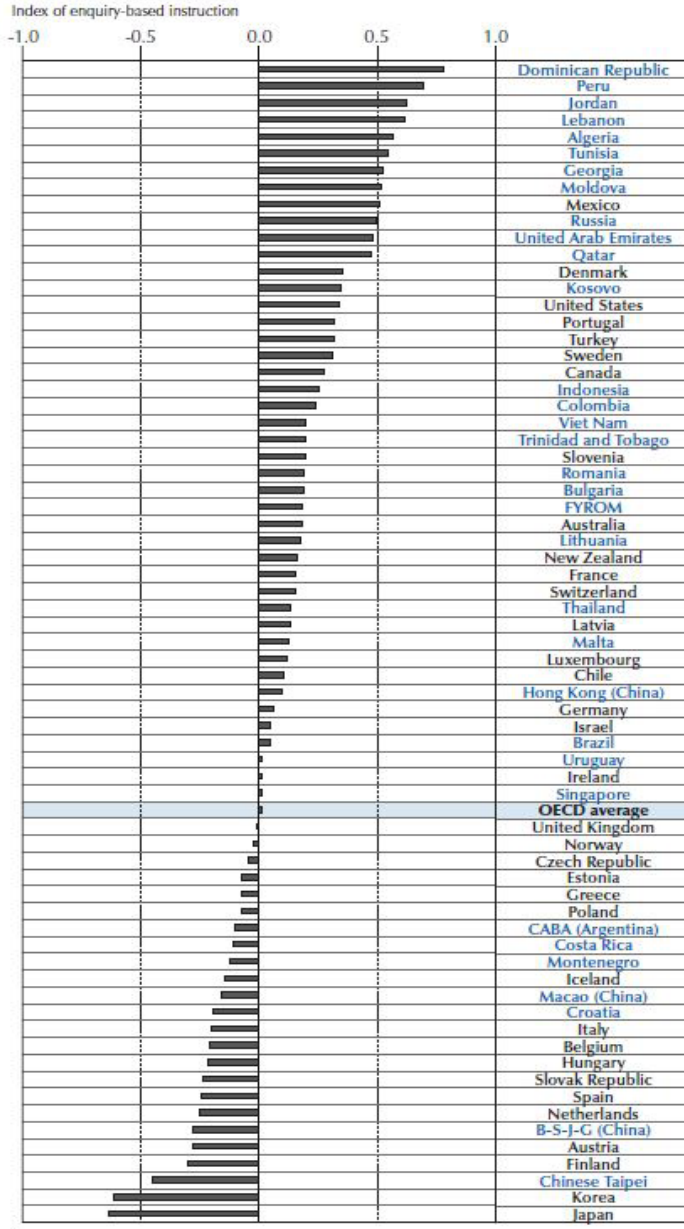


Figure 3. Countries sorted by PISA Index of Inquiry-based instruction. Country means based on students' reports. The index normalized for OECD mean value zero and standard deviation 1.0. Countries in black are the OECD countries. Source: Part of Figure II.2.19 in OECD, 2016c, p72.

The PISA-reports note that the mentioned high-scoring countries seldom use IBSE and that the general pattern is that *“in 56 countries and economies, greater exposure to inquiry-based instruction is associated with lower scores in science.”* (OECD, 2016c, p36).

Also, for the variation among students within the same country, the PISA finding is that *“in no education system do students who reported that they are frequently exposed to inquiry based instruction [...] score higher in science.”* (ibid, p71).

One of the questions in the inquiry-index may be of particular interest for science educators. *Experiments* play a crucial role in science, and have always played an important role in science teaching at all levels. In many countries, doing experiments are part and parcel of science teaching, often accompanied with concrete statements about numbers of obligatory experiments to be performed. Well-equipped school science laboratories are often seen as a prerequisite for quality science teaching. But when it comes to PISA scores, the report states that *activities related to experiments and laboratory work show the strongest negative relationship with science performance.* (ibid, p71).

But, although the relationship between IBSE and PISA test score is negative, IBSE relates positively to *interest* in science, *epistemic beliefs* and *motivation* for science-oriented future careers: *“Across OECD countries, more frequent inquiry-based teaching is positively related to students holding stronger epistemic beliefs and being more likely to expect to work in a science-related occupation when they are 30.”* (ibid, p36).

From a science education and an even broader educational perspective, these results are most interesting and challenging. It indicates that although IBSE may not increase the immediate knowledge score on tests like PISA, inquiry-based teaching may lead to a better understanding of the nature of science and also increase students' motivation to choose science-related education and careers.

SUMMARY AND DISCUSSION

This article has two main concerns: Firstly to argue that PISA should be understood in a broader political and cultural context, and secondly to draw attention to issues of direct relevance for science education.

PISA is basically a political project: formulated, governed and financed by the OECD “owners” in line with the priorities of the OECD. The project is explicitly normative, but presented and understood as a neutral and objective measure of the quality of a nations' school system and a proxy for the nations' future global competitiveness. From this perspective, PISA has been a global success, and the consequences of PISA are many. The most serious is a kind epistemological dominance: Governments' obsession with PISA-scores redefines and restricts the very meaning and purpose of schools. In many countries PISA-shocks have opened up for ill-informed school reforms. The Finnish educator Pasi Sahlberg (2011) describes these PISA-driven educational reforms by the acronym GERM: Global Educational Reform Movement, characterized by privatization, market driven reforms, free school choice, competition and test-driven accountability. He notes that *“Finland has remained immune, but other Nordic countries have moved to adopt policies that are close to GERM.”* (Sahlberg, 2011, p125).

While the above points should be a concern for everybody with an interest in the role of schools and education, the second part of this article addresses issues that are more directly of concern for science educators. Most science educators would appreciate the deliberations around the concept of science literacy as provided in the PISA Framework (OECD, 2016a). The resulting PISA definition of science literacy is in harmony with widely shared ideas and ideals among stakeholders in science as well

as in science education. (Although the PISA documents barely mention the current UN Sustainable Development Goals and the related initiatives regarding Education for Sustainable Development.)

A reader of PISA results is likely to believe that the science literacy score is a valid measure of the published definition. But this is not the case. The PISA science literacy score is calculated solely on the scores on the test units. Important elements of the definition of science literacy are not included in the score, but are measured by responses in the students' questionnaire.

More concretely, we have drawn attention to the fact that the published PISA test score correlates negatively with being exposed to inquiry-based teaching methods, in particular with doing science experiments. The celebrated PISA-winners do to a limited degree use IBSE in their teaching.

On the other hand, being exposed to IBSE is positively related to students' epistemological beliefs about science, their attitudes towards science and their interest in pursuing science later in life. From an educational perspective, such results may in the long run be more important than answering correctly on a test at age fifteen.

In an analysis of the PISA 2015-data, US-Chinese Yong Zhao (2017) points out that students in the so-called PISA-winners in East-Asia (e.g. Japan, Korea, Hong Kong, Singapore) seem to suffer from what he calls "side-effects" of the struggle to get good marks and high test-scores. He presents PISA-data that show that students in these countries have very low self-confidence and self-efficacy related to science and mathematics. He points out that there *"is a significant negative correlation between students' self-efficacy in science and their scores in the subject across education systems in the 2015 PISA results. Additionally, PISA scores have been found to have a significant negative correlation with entrepreneurial confidence and intentions."* (Zhao, 2017).

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 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 from PISA should call for some caution against trying to copy the "PISA winners".

Other problematical PISA-results should also be of interest for educators. Money and resources spent on education do not seem to matter for the PISA scores (OECD, 2016a, p184). Class size does not matter (OECD, 2016b, p207). PISA-scores correlate negatively with investment in and the use of ICT in teaching (OECD, 2015b). PISA science scores seem unrelated to the teaching-time given to science in schools (OECD, 2016c). The availability of a well-equipped school science laboratory with "science-specific" resources has no positive association to PISA-score in the majority of PISA countries (OECD, 2016c, p55). Somewhat surprising is also the finding that the index of "feedback from teachers" correlates negatively with the PISA-score for practically all countries, among them all the Nordic countries (OECD, 2016c, p67).

Such findings of positive and negative relationships and associations are often presented and understood as causality in the international PISA reports. We should, however, remember that PISA cannot by its design make any causal claims. As noted before, the OECD has been urged to have a "more measured approach to reporting and interpreting PISA results" (Rutkowski & Rutkowski, 2016).

Concluding remarks

Some of the above problematic results are not difficult to understand. If the ultimate test of quality in science teaching is the score on a test (written or digital), it is no surprise that teaching will be more

“cost-effective” than spending time on excursions, experimental work or discussion of socio-scientific issues or the nature of science.

A written (or digital) test like PISA can hardly measure the skills and competencies acquired in experimental work in a lab or on an excursion; neither can it capture the kind of interest, curiosity and enthusiasm that may be the result of argumentation, inquiry, and the search for solutions to questions that the students have formulated themselves. But these aspects are part of the definition of science literacy in PISA as well as in other sources.

The use of PISA data for policy recommendations is, at best, very selective. If one “believes in PISA”, one has to take all the results seriously, also those which are counterintuitive and at odds with other research findings and policies that are recommended by scientists as well as science educators.

The danger is that schools, in politicians’ priority to climb on the PISA rankings, sacrifice the educational strife for a better, more interesting, authentic, context-based and relevant science education for the learners.

Large resources are used by the governments to run the PISA project. Many academic institutions are dependent on contracts to run PISA and other ILSAs. In a situation where external finance is vital, these contracts are of importance for jobs as well as budget balance. Institutions, who win the bids to run ILSA projects as well as national and other testing regimes, develop a close relationship to the governmental agencies. Some sort of loyalty may be expected from the staff and the institution. Academic ideals may become subordinated and suffer when such dependencies develop. Critical research is scarce and not well funded. Given the great political and educational importance of PISA, there is a strong need for critical and independent research. Above all, science educators should address the contradictions in the messages from the different international actors, and face the cultural, political and ideological tensions between different views on the role and purposes of school science.

REFERENCES

- Addey, C., Sellar S., Steiner-Khamsi, G., Lingard, B. & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare: A Journal of Comparative and International Education*. 47(3), 434-452
<http://dx.doi.org/10.1080/03057925.2017.1301399>
- 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: OECD Publishing.
- Breakspear, S. (2014). *How does PISA shape education policy making? Why how we measure learning determines what counts in education*. Centre for Strategic Education. <http://simon-breakspear.com/wp-content/uploads/2015/09/Breakspear-PISA-Paper.pdf>
- 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
- Dewey, J. (1938). *Logic. The Theory of Inquiry*. New York: Henry Holt and Company.
- Elstad, E. & Sivesind, K. (eds.) (2010). *PISA: Sannheten om skolen?* [PISA: the truth about the school?] Oslo: Universitetsforlaget.
- EC (2007). *Science Education Now: A renewed pedagogy for the future of Europ*, (The Rocard report), Brussels: European Commission, EC.
https://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf
- Grek, S. (2009). Governing by numbers: the PISA ‘effect’ in Europe. *Journal of Education Policy*, 24:1, 23-37.

- Hogan, A. Sellar, S. & Lingard, B. (2016). Commercialising comparison: Pearson puts the TLC in soft capitalism, *Journal of Education Policy*, 31(3), 243-258.
- ICSU (2011). *Report of the ICSU Ad-hoc Review Panel on Science Education*. Paris: International Council for Science.
- Klitmøller, J. & Sommer, D. (eds.) (2018, in print). *FREMTIDSPARAT? Hinsides PISA – nordiske perspektiver på uddannelse*. [Beyond PISA. Nordic perspectives on education.] København: Hans Reitzels Forlag.
- Komatsu, H. & Rappleye, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, 53:2, 166-191, DOI: 10.1080/03050068.2017.1300008
- Landahl, J. & Lundahl, C (2017). *Bortom PISA. Internationall och jämförande pedagogik*. [Beyond PISA: international and comparative pedagogy.] Stockholm: Natur & Kultur.
- Lederman, N.G. (1992). Students' and teachers' conceptions of the nature of science: a review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.
- Lederman, N.G. (2006). Nature of science: Past, present and future. in S. Abell and N.G. Lederman (eds.). *Handbook of Research on Science Education*. Lawrence Erlbaum, Mahwah, NJ, p. 831-879.
- Lundgren, U. P. (2011). PISA as a Political Instrument in Pereyra, M.A., Kothhoff, H.G and Cowen, R (eds.) *PISA Under Examination. Changing Knowledge, Changing Tests, and Changing Schools*. Rotterdam: Sense Publishers.
- Millar, R. and J.F. Osborne (eds.) (1998). *Beyond 2000: Science Education for the Future*, School of Education, King's College, London, www.nuffieldfoundation.org/sites/default/files/Beyond%202000.pdf.
- OECD (1999). *Measuring Student Knowledge and Skills. A New Framework for Assessment*. Paris: OECD Publications.
- OECD (2010a). *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices*, Vol. 4. Paris: OECD Publications.
- OECD (2010b). (Hanushek and Woessman) *The High Cost of Low Educational Performance: The long run Economic Impact of Improving PISA Outcomes*. Paris: OECD Publishing. <https://www.oecd.org/pisa/44417824.pdf>
- OECD (2015a). *OECD's video PISA – Measuring student success around the world*. Available at: <http://www.oecd.org/pisa/aboutpisa>
- OECD (2015b). *Students, Computers and Learning: Making the Connection*. Paris, OECD Publishing. <http://dx.doi.org/10.1787/9789264239555-en>
- OECD (2016a). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264255425-en>
- OECD (2016b). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- OECD (2016c). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD Publishing.
- OECD (2016d). *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264258051-en>
- OECD (2017a). *PISA 2015 Technical Report*. Paris: OECD Publishing. <http://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>
- OECD (2017b). *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264276116-en>
- OECD (2017c). *Education at a Glance 2016: OECD Indicators*. Paris: OECD Publishing <http://dx.doi.org/10.1787/eag-2017-en>
- OECD (2018a). *PISA for Development. Policy Briefs*. Paris: OECD. <http://www.oecd.org/pisa/aboutpisa/pisafordevelopment-documentation-briefs.htm>

- OECD (2018b). *PISA in Focus. Policy Briefs*. Paris: OECD <http://dx.doi.org/10.1787/22260919>
- OECD (2018c). *PISA-based test for schools* <http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools.htm> visited Feb 18. 2018
- Osborne, J, & Dillon, J (red) (2008). *Science Education in Europe: Critical Reflections* A Report to the Nuffield Foundation, King's College London.
- Pons X. (2017). Fifteen years of research on PISA effects on education governance: A critical review. *Eur J Educ.* 2017;52:131–144. <https://doi.org/10.1111/ejed.12213>
- Rogers, E. M (1960). *Physics for the Inquiring Mind*, Princeton: Princeton University Press
- Rutkowski, L. & Rutkowski, D (2016). A Call for a More Measured Approach to Reporting and Interpreting PISA Results. *Educational Researcher*, Vol. 45 No. 4, pp. 252–257
- Sahlberg, P. (2011). *Finnish Lessons: what can the world learn from educational change in Finland?* New York: Teachers College.
- Schleicher, A. (2013). *Use data to build better schools*. TEDGlobal Presentation. http://www.ted.com/talks/andreas_schleicher_use_data_to_build_better_schools?language=en
- Sellar, S; Thompson, G, & Rutkowski, D. (2017). *The global education race: taking the measure of PISA and international testing*. Toronto: Brush Education Inc.
- 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. (2014). Pisafisering av norsk skole. En suksesshistorie fra OECD. [Pisa-fication of Norwegian schools: A success story from the OECD.] in Røvik, K A, Eilertsen T.V. & Furu, M.E. (eds.) *Reformideer i norsk skole. Spredning, oversettelse og implementering*. Oslo: Cappelen Damm Akademisk.
- Sjøberg, S. (2015). PISA and Global Educational Governance- A Critique of the Project, its Uses and Implications. *Eurasia Journal of Mathematics, Science & Technology Education*, vol.11, no.4, pp.11-127. <https://eric.ed.gov/?id=EJ1059205>
- Steiner-Khamsi, G. (2003). The politics of league tables. *Journal of Social Science Education*. Vol 1, 1-6 Online <https://doi.org/10.4119/UNIBI/jsse-v2-i1-470>
- Zhao, Y. (2017). What works may hurt: Side effects in education. *Journal of Educational Change*. DOI 10.1007/s10833-016-9294-4