

PISA truth effects: the construction of low performance

Margareta Serder* and Malin Ideland

Both of Faculty for Education and Society, Malmoe University, Sweden

Countries with large numbers of students who struggle to master basic reading skills at age 15 are likely to be held back in the future, when those students become adults who *lack the skills needed* to function effectively in the workplace and in society. Among students *who fail to reach the baseline level of performance* (Level 2) in mathematics, reading or science, most can be expected not to continue with education beyond compulsory schooling, and therefore risk facing difficulties using mathematics, reading and using science concepts throughout their lives ... Even in the average OECD country, where more than one in five students does not reach Level 2, tackling such *low performance* is a major challenge. (OECD, 2013, p. 254, our emphases in italics)

Keywords: performance; competency; PISA; students; STS; scientific literacy

This quote is from an OECD publication that reports the results from the 2012 PISA survey, on which the above analysis is based. The Swedish PISA results from that year were the lowest ever in all three knowledge areas at stake, which shocked the nation. Since low performance on the PISA is put forward as serious personal and societal risks, this concern is expected. In December 2013, the Swedish results were broadcast directly from a press conference held at the National Agency of Education. The message was that the results were very serious, overwhelming, and signaled a school system in crisis. During the press conference, Director-General Anna Ekström also emphasized her confidence in PISA: ‘Here at the National Agency for Education, we trust PISA’ (Ekström, 3 December 2013). Through this broadcast, and massive media reporting, the crisis of the Swedish school system became publicly indisputable. Although we share the view that many things probably would be in place to strengthen Swedish schools, we find this type of unreserved faith in PISA as evidence of educational failure or success, to be problematic. In this article, we make an attempt to unpack what we would term the taken-for-granted notion of *low performance*. Using the theoretical lens of Michel Foucault (1971), and particularly Bruno Latour’s (1999) STS studies, our main thesis is that low performance is not a given category, but a reality performed into being by students, test instruments, and measurement rationality.

In the global structure of education, understanding what important knowledge ‘is’, and what could be its missing parts, has become a grand narrative of our time, with help from the idea of a globalized educational system (Novóa, 2013). Actors in the global education market (or rather, instigators at the Organisation for Economic Co-

* Email: margareta.serder@mah.se

operation and Development [OECD], the European Commission, the World Bank, and UNESCO (Grek, 2009; Lawn & Grek, 2012; Morgan, 2009) for international assessments such as PISA are constructing the standards for what is a successful student or a successful nation.

However, the role of PISA deviates from other assessments in that it constitutes the very instrument that can guarantee control over the edu-political development that other actors hope to see (Lawn & Grek, 2012). Therefore, the PISA has a unique position in defining what should be the important knowledge of our society and, in Foucault's terms, producing a global truth on this matter. Clara Morgan (2009, p. 196) stresses that PISA serves as a 'mechanism for legitimating, disseminating, and regulating a certain form for official educational knowledge'. Furthermore, we argue that PISA performance differs by being a global measure. It is one scale against which all countries measure independent factors such as curricula, and school organization. It also differs in its aim to assess a specific kind of literacy—competency-oriented skills assumed to be applicable in so-called everyday situations or general life (OECD, 2006). In Morgan's (2009) words, the role of education, according to PISA frameworks, is acquiring competencies and skills needed for a knowledge-based economy. These competencies and skills are defined as categories of literacy: *reading*, *mathematical*, and *scientific*. But what is a 'literacy'? How does it become measurable?

Translations through a chain of events

Radhika Gorur (2011) and Tom Popkewitz (2011) use the notion of *translation* to understand the representational changes that PISA literacy undergoes in becoming a global, taken-for-granted, object for thought. For Latour (1999), the concept of translation means to stress the whole chain of events that makes up the knowledge about the world—our facts—and acknowledge that some things are added and some are lost in each step of translation. In asking about the nature of the scientific knowledge produced by our modern-day 'Delphic oracle', Gorur (2011, p. 77) uses translation to describe the production of PISA facts:

At every point there is a translation—from 'knowledge and skills for life' to 'literacy, numeracy and scientific literacy', to a few items that stand for these literacies. The few test items, validated by the framework and the field testing, refer to the three literacies being tested, which in turn represent the knowledge and skills for life, which in turn represent the work of millions of teachers in tens of thousands of schools in a sizeable portion of the world, and stand for the performance of education systems. (Gorur, 2011, p. 84)

Popkewitz (2011) describes the multiple translations that take place when academic disciplines are transformed into school subjects to fit into the purposes of schools, and how they travel further in the realizations and calculations of PISA. He calls this chain of translations *alchemy*. Furthermore, he historicizes the complex transition in which students' knowledge translates into numbers and then into PISA facts as: 'telling the truth about society, schooling and children' (p. 33). He asserts that numbers are systems of communication that create distance from the phenomena of concern, 'by appearing to summarize complex events and transactions' (Popkewitz, 2011, p. 33; c.f. Porter, 1995). Gorur argues that through the consecutive events of the PISA process, the facts produced are disconnected from their situated origin in people and tasks. Popkewitz writes that the apparent objectivity and impartiality of numbers have

become a standard of reason for our modern democracies (2011, p. 43). Between Gorur (2011) and Popkewitz (2011), we have identified a gap in the chain of translations, which this article seeks to fill. This gap is about the encounters in which PISA items and real students meet, which are the very events in which marks on test sheets are produced. This is intended to attach the facts produced in assessment, as evidence for performance and skill (or lack of such), to the socio-material practices from which they stem. In this article, we focus on the competencies in the scientific literacy (OECD, 2006).

It is important for our unpacking of the notion of low performance to align with Gorur's proposition to view the practices of PISA in terms of a scientific laboratory (Latour, 1999); a place where knowledge is constructed by different kinds of processes enacted for different purposes. To push our theoretical argument further, we propose to compare the laboratory activities of PISA with those in a completely different way, namely the nineteenth-century laboratory of Louis Pasteur. This might seem to be a detour at first, but we believe it is not. This analogy will help us make an argument about what it takes for any category (in our case, a PISA competency or low performance) to exist.

Latour (1999) describes the scientific work Pasteur enacted when he so to speak 'discovered' microbes. Latour asks whether or not one can assert the location or existence of microbes before Pasteur? Latour writes that many would find it commonsensical to answer that the microbes were in the same place before and after Pasteur; that Pasteur did not change the course of events—that he only discovered a pre-existing phenomenon, and named them *microbes*. However, Latour argues that things and phenomena come into reality by being given attributes in the scientific process in which defining the object takes place. The microbes in Pasteur's laboratory needed to be given circumstances, conditions, specific nutrition demands, and ways of being detected. Thereafter, they became microbes as we know them. Only then could they could be discerned as something.

We propose thinking of performance and PISA competencies through this analogy; recognizing that the existence or not of a specific PISA competence is linked to the historical event of starting to try to detect it. This leads us to ask: What are the conditions and attributes of the PISA test that make us see competence as something? Within a Latourian system of truth, we could propose that measured competencies are not 'caused' by knowledge and learned skills, but come into existence through the means constructed to detect them. Furthermore, the idea of PISA as a scientific laboratory invites us to regard the test instrument, or the PISA test, by which performances and competencies are produced, as a kind of detector (Knorr Cetina, 1999) or a device constructed to 'see' certain things but not others. For instance, the notion of a detector can be exemplified by the instruments constructed for reading the level of hemoglobin in a blood test. Following some procedures, those instruments can detect hemoglobin, but not, for instance, the content of virus or thyroid hormone. Meanwhile, the ensemble of the instrument and the blood also constructs a particular way to perceive and know about blood. Later on in this paper, we describe how PISA can be viewed as a detector.

Assumptions and critics

According to Gorur (2011), viewing PISA's facts as realities performed into being rather than pre-existing truths is 'to convert PISA knowledge from a "matter of fact",

something that is beyond dispute, into a “matter of concern”,’ (p. 78), which again becomes available for discussion (Latour, 2004). We want to interfere with a number of implicit epistemological and ontological assumptions in knowledge measurement, by recognizing the performativity of performance and competencies. Among these is a view of knowledge as individual, measurable, and culturally neutral. Moreover, test items can be produced in the sense that the acts and enactments they enable are globally the same. Perhaps we could speak of an assumption of knowledge as being the same before, during, and after assessment?

The scholarly critique of PISA addresses these assumptions. Gorur and Wu (2014) are analyzing ‘average performance scores’ in Australia and stress these to actually be uninformative and misleading. In addressing the nature of knowledge in general, scholars using sociocultural learning theories claim that the frontiers between the individual and the social, and the mind and the material are vague and fluid. Therefore, they argue that knowledge cannot be assessed on an individual basis or just in written isolation (e.g., Roth, 2007; Schoultz, Säljö, & Wyndhamn, 2001). Furthermore, the term *competence* implies a rational, individualized view of knowledge, in addition to its standards of ‘utility’ in the knowledge-based economy (Carvalho, 2012). According to the PISA framework, the knowledge measurement is ‘informed, but not constrained by the denominator of national curricula’ (OECD, 2006, p. 7). Instead, it assesses ‘things that 15-year-old students will need in the future [with a focus on the application of their] knowledge and experiences to real life issues’ (p. 7). In this sense, the competencies of real life appear as universal (Tröhler, 2013). In addition to the critique from scholars within the psychometric research community about assessment validity (Goldstein, 2004; Kreiner, 2011; see also Morgan, 2009, pp. 194–195), it is argued that the kind of international comparisons to which PISA adheres rely on assumptions about similarities in the educational systems that hardly exist (Fertig, 2004).

In spite of the critics, the PISA machinery endures and the edu-political communities of Europe and elsewhere continuously repeat the importance of its evidenced results and the necessity of taking it into account (Carvalho, 2012; Grek, 2009; 2012) as well as following the ‘good examples’ of performance (Simons, 2014). This is particularly the case in areas where ‘low performance’ is a ‘fact’, as in the current Swedish PISA debate. In the well-established global structure of education (Lawn & Grek, 2012; Meyer & Benavot, 2013), it appears that what important knowledge ‘is’, and thus what could be its ‘missing parts’ have become universal discourses of truth also in a Foucauldian sense (Foucault, 1980). Expanding on this body of literature, the present article deconstructs the notion of performance, by shifting from a representational focus to a performative one (Gorur, 2011) and by combining analyses of policy documents and studies of students’ encounters with test items.

Using the idea of PISA as a scientific laboratory (Gorur, 2011), the empirical part of this article aims to explore attributes and conditions for (low) performance and (lack of) competency in PISA scientific literacy. We approach this in two ways. First, we analyze how the categories of performance and competency are discursively constructed in written PISA materials such as assessment frameworks, test items, and results reports. This analysis of how standards for scientific literacy are fabricated inside a measurement rationality (Biesta, 2010) is regarded as a background against which the main data are read.

Our second approach explores encounters in which 15-year-old students collaboratively work with PISA scientific literacy items. These situations have been

constructed to study in moment-by-moment analysis the socio-material conditions for the construction of responses. We find it useful to understand the notion of the *detector* (the test instrument itself, and the acts it enables or not) through Andrew Pickering's (1995) theory of *mangling*. Pickering allows us to see acts and events as a contingent adjustment between material and human actors, evolving in the constant resistance and accommodation of the mangle of practice. We use his metaphor as an analytic tool for grasping how students' oral and written answers come into being in the situated encounters with the PISA materials. It invites us to see performance not as representations of held knowledge, but as a result of resistances and accommodations that arise there and then.

Methodology

Our main empirical focus in this article is to explore how performance and PISA competency come into being in item-student encounters to unpack the taken-for-granted notion of performance. In addition, we analyze the discursive frames of these categories as they are performed in various PISA materials. The PISA assessment framework (OECD, 2006), in which scientific literacy is defined, is the core material we use for analysis. The publicly released PISA test units in scientific literacy (eight units/24 items) are included in the text analysis.¹ Since we are interested in the construction of performance, we also use the reported results and analyses from PISA 2006 (OECD, 2007) and an OECD report specifically on PISA strong performers (OECD, 2010) to map the discourse. For reference, we use the Danish and Swedish PISA reports from the same period, as well as the four other published PISA frameworks focused on reading and mathematics literacy (OECD, 1999; 2003; 2009; 2013).

The analysis of the PISA documents was inspired by Foucauldian discourse theory (1971). Accordingly, we considered normality as being culturally organized; and explored how taken-for-granted ideas on performance and competency are constructed through implicit rules about what can or cannot be said and done, what discursive elements are included and excluded, and how this becomes a system of truth. As a system of truth, discourses not only reflect reality, but also produce it through the constitution of what is valid knowledge and desired subjects. During analysis of how competence and performance are 're-produced', these notions were pinned down into four related topics in order to detect the constitution of *low performance*. These topics were *science* (the object for knowledge); the *science knower* (the subject for knowledge); *scientific literacy competency* (the mode for knowledge); and *performance* (the quality of knowledge). All passages in the texts estimated to speak about these topics were underlined and themed. These topics were eventually discussed by the two authors in terms of how they connect to the topics and what constructions they reflect, according to our readings. This analysis aims to understand the attributes that allow scientific literacy competency and (low) PISA performance in science to be discursively produced in frameworks, reports, analyses, and testing materials.

The main data consists of 16 hours of video-recorded encounters between 11 test items—occurring in 'Acid Rain' [S485], 'Greenhouse' [S114] and 'Sunscreens' [S447]) from the published PISA scientific-literacy test material, and 21 small groups of 15-year-old Swedish students working collaboratively with these items during a science lesson. The instruction for the teacher-constructed groups was to discuss and

agree on one common written answer. In total, 71 Year 9 students in a small-city compulsory school in an urbanized area in Southern Sweden participated. The participants amounted to half of their cohort in this particular school, of which publicly available statistics² report average results compared to the general Swedish context. The test items were selected together with the Swedish and Danish national PISA teams to use items that were considered well-functioning in the test, used in all three PISA assessments (2000, 2003, and 2006), and fairly well reflected the remaining non-released items of the PISA assessment. At the end of the lesson, the students' written material was collected. Central for the analysis of the video-recorded, transcribed, and data material is what the students were doing with the test-items (viewed as a particular kind of detector) and vice versa: what the test-items (detectors) were doing with the students in these socio-material encounters.

In the analysis, the specific focus was situations which appeared to be experienced lack of competence. Examples include situations in which (1) no answers were found; (2) incorrect answers were suggested (that would result in no score, according to the protocol); or (3), any kind of experience of low performance, such as when students expressed feelings of stupidity, insufficiency, or a lack of understanding. We also included situations in which correct responses were produced and/or in which relevant discussions took place, although incorrect answers resulted.

Performance was grasped as an attribute of the ensemble of test items and students, in which resistance and accommodation to the situation evolved in each moment, in a mangle of practice (Pickering, 1995). For instance, when students adjusted their response to the amount of space given in the test question, this was analyzed as resistance on the part of the item and accommodation on the part of the students. The analysis of the PISA text materials and the item-student encounters were presented in parallel, starting with scientific literacy competency as performed into being in documents, then items and encounters, followed by performance presented in a similar manner.

The encounters were originally constructed for research purposes to approach the PISA test from a student perspective (see Serder & Jakobsson, 2014; Serder & Jakobsson, in review), trying to find alternative ways to understand the test situation from which PISA results are reported. When students respond to test questions in a standardized way, they work in silence. We argue that in a collaborative approach, aspects of students' co-actions with the materials become observable in talk, propositions, and doing-with the material. This study enables us to see possible, but not any of the actual, enactments that take place in the process of mangling with the test in real testing situations. Recalling Pasteur's laboratory, we are trying to grasp on a micro-level the situated conditions and attributes forming PISA answers that are then translated into scientific literacy competencies.

Scientific literacy competencies performed into being

In the PISA materials, *science* appears as a body of knowledge which has the ability to solve many of the problems of future society. It also has its evident place in people's everyday life 'problem-solving'. For example, test questions are placed in narrative contexts of a presumed real-life situation, in which scientific methods preferably are to be used when examining the effects of acid rain or the effectiveness of sunscreens (see also Serder & Jakobsson, 2014). The *science knower* is a future optimal citizen who performs as an individual, ongoingly facing challenges that

demand scientific competence and willing to engage in them (OECD, 2006, p. 20; c.f. Popkewitz, 2008; Ideland & Malmberg, 2014), and acts (as) within a scientific discourse/culture. A science-knower is:

A person who has an interest in scientific topics, thinks about science-related issues, has a concern for issues of technology, resources and the environment, and reflects on the importance of science in personal and social perspectives. (OECD, 2006, s. 24)

However, according to the PISA framework (OECD, 2006), the scientific literacy competencies assessed by the test are narrower. They are named: *Using scientific evidence*, *Identifying scientific issues*, and *Explaining phenomena scientifically*. All three are put forward as neatly defined knowledge competencies with fixed boundaries (e.g., OECD, 2006, pp. 29–33), giving the impression that they are ready to achieve and apply. For instance:

Identifying scientific issues: It is important to be able to distinguish scientific issues and content from other forms of issues. Importantly, scientific issues must lend themselves to answers based on scientific evidence. (OECD, 2006, p. 30)

This short text illustrates how different categories of knowledge may be performed into being. Here, scientific issues become isolated from other issues (such as ethical, political, and social). In the same vein, content, as a thing-like object (Latour, 1999), is constructed as either scientific or not. As such, the boundaries of these categories are asking for coders around the world to carefully study each word in written responses to uphold the division between scientific and non-scientific. The construction of *science* (and technology) also performs scientific literacy into a necessary competency for the future.

An understanding of science and technology is central to a young person's preparedness for life in modern society. It enables an individual to participate fully in a society in which science and technology play a significant role. (OECD, 2006, p. 20)

Scientific literacy is constructed through the PISA documents as a mode for knowledge that makes people make rational decisions, based on scientific arguments. We do not suggest that nuanced decisions cannot be made through informed scientific knowledge. Nevertheless, other arguments simultaneously become constructed as wrong, or as a different kind of knowledge that is not legitimate in this situation (which does not seem so lifelike). We want to point at the discursively produced, opposed, positions in relation to science: What becomes of the individual who is not interested in science in this particular manner? What if they do not privilege or reflect on how important science is in a personal and social perspective? We argue that within this discourse, such students are constructed as lacking competency. In this way, lack of scientific competence is produced, and intimately tied to personal and cultural values about the natural, social, and material world.

The next step in the analysis shows how these constructions of (lack of) competencies come into being in student-item encounters. The PISA test questions—that is, the items—are assembled in thematic units, beginning with an introductory text that often appears to connect the tasks to stories about real life. The introductory text of the PISA unit 'Sunscreens' (S447) serves as a first example. This text describes an experiment conducted by two fictional students, Dean and Mimi. The experiment is designed to compare the effectiveness of four different sunscreens

(called S1, S2, etc.), using different materials, creams of the sunscreens, and mineral oil (M) and zinc oxide (ZnO). One of the introductory passages reads as follows:

Dean placed a drop of each substance inside a circle marked on one sheet of plastic, then put the second plastic sheet over the top. He placed a large book on top of both sheets and pressed down.



Mimi then put the plastic sheets on top of the sheet of light-sensitive paper. Light-sensitive paper changes from dark grey to white (or very light grey), depending on how long it is exposed to sunlight. Finally, Dean placed the sheets in a sunny place.

Figure 1: Sunscreens (S447) introduction text.

Later on, the introductory text states that Dean and Mimi ‘included mineral oil because it lets most of the sunlight through, and zinc oxide because it almost completely blocks sunlight’ (S447; OECD, 2006). Thus, the roles of the zinc oxide and the mineral oil are supposed to be understood as extreme measuring points (references) in this experiment. Once the ‘real-life’ scene for the scientific experiment is set, it is followed by four consecutive test questions on scientific literacy, including a multiple-choice question to measure students’ knowledge about scientific methods (the competency *Identifying scientific issues*):

Question 2: SUNSCREENS S447Q02

Which one of these statements is a scientific description of the role of the mineral oil and the zinc oxide in comparing the effectiveness of the sunscreens?

- A Mineral oil and zinc oxide are both factors being tested.
- B Mineral oil is a factor being tested and zinc oxide is a reference substance.
- C Mineral oil is a reference substance and zinc oxide is a factor being tested.
- D Mineral oil and zinc oxide are both reference substances.

Figure 2: Sunscreens test item. Question 2.

According to protocol, the correct answer is D. Let us now turn to the students in the study, who are trying to find out the right answers on this item. Caroline, Lucy and two other girls select Option D, producing a correct response, which should perform scientific literacy, according to PISA standards. However, Lucy and Caroline are comparing the text introducing the test question with the test item they are answering.

Lucy: Mm ... Yes but THAT *Dean placed a drop of each substance inside a circle ...*
[reading aloud the text just above the picture in Figure 1] Then both are ...?

Caroline: Then both are *reference substances* [Option D]. (Group J; Nov 2010, 13:25–14:00)

In this encounter, Caroline and Lucy enacted a mere comparison of words. Semiotic resemblances in different parts of the item are more important than knowledge about scientific methods. The strategy appears to be about finding the word *substance* in both the description of zinc oxide and mineral oil and the combination of words in Option D. The answer intended to represent individual, measurable knowledge in science instead was produced from accidental resemblances of different parts of the text.

Further examples can be found in our material. Moreover, we observe how answers are constrained by the space afforded on the test sheet, problems in briefly formulating the discussed ideas, or it being too bothersome to construct an answer. One type of encounter that we often see in the video material concerns the meaning of specific words. In the following example, Richard, Jana, and Cornelia are operating with the same item as Lucy and Caroline. This time, the group ends up choosing C, which is an incorrect answer to ‘Sunscreens’ Question 2.

Richard: [reading] The words are so strange...

Jana: What?

Richard: *reference substance* ... [reading aloud]

Jana: [shaking her head] Yeah, like ... what is that?

Cornelia: [looking out through the window]

[All students read in silence]

Richard: If only one knew what reference means? [sighs]

[The students are reading again, searching in the text]

Jana: This [Option C] is the most reasonable because ... zinc oxide does not let sunlight through. One could say perhaps it is a factor, as in sun-protection factor ... which is supposed to stop sunlight. Just as sun protection. And they [zinc oxide and mineral oil] are like ... not the same ... so let's go for C. (Group L, Nov. 2010, 20:29–22:40).

What are the conditions by which scientific literacy competency, or rather its lack, comes into being in this encounter? First, the enactments of this encounter with the test item have very little to do with knowledge about scientific methods, which is the competence being measured. We see situated resistances and accommodations (Pickering, 1995) in relation to the meanings of the words *reference* and *factor*, and their relationship to real-life experiences of sunscreens. Since the *reference* in this encounter is unknown, the word *factor* instead becomes the focus for action. In several other encounters, we observed a situated use of the word *factor* in the sense of *sun protection*, for which it often stands in everyday Swedish vocabulary (see further Serder & Jakobsson, in review; 2014). In the quote above, Jana refers to Option C, in which zinc oxide is a factor as ‘the most reasonable’, and then suggests the (incorrect) Option C. According to the text, mineral oil and zinc oxide have contrasting properties—they are either blocking or letting through sunlight. In this student-item encounter, an answer in which zinc oxide is a factor and mineral oil is its opposite (a reference) is produced. As possible evidence about student scientific literacy, this

answer is incorrect and a token for low performance in the competency *Identifying scientific issues*. Other competencies, such as everyday language and reading skills seem to distract the scientific reasoning. The boundaries for the object and the mode for knowledge become blurred.

At other times, informed discussions take place in which the produced answer does not match the discussion. Edward and Cathrine appear to know a lot about the zinc oxide substance that occurs in the Sunscreens unit. However, as they work with Question 3, knowing much is not an advantage, according to how the encounter unfolds.

Which one of these questions were Mimi and Dean trying to answer?

A How does the protection for each sunscreen compare with the others?

B How do sunscreens protect your skin from ultraviolet radiation?

C Is there any sunscreen that gives less protection than mineral oil?

D Is there any sunscreen that gives more protection than zinc oxide?

Figure 3: Sunscreens test item. Question 3.

Also here, an incorrect option (D) is selected (Option A is correct). For this item, language translation between national versions of the PISA test actually has a role. This is because Options C and D (above) are translated into:

- C Finns det något solskyddsmedel som skyddar **sämre** än mineral olja?
D Finns det något solskyddsmedel som skyddar **bättre** än zink oxid?
(Skolverket, 2007)

Language translation procedures in PISA are highly standardized to assure assessment quality (Bybee, McCrae, & Laurie, 2009; Grisay, 2003). However, several studies have pointed at translation-related problems in PISA (Arffman, 2010; Puchhammer, 2007). The two words in bold font, *sämre* and *bättre*, stands for *worse* and *better*. This translation was used instead of the English words *less* and *more* (c.f. Serder & Sørensen, 2014). These words are actually approached in the encounter between Edward, Cathrine, and this item.

[After reading in silence, Edward turns to Cathrine]

Edward: Yes, but how is ... which one protects the best? I mean how it protects?

Cathrine: Yes, yes.

Edward: But it must be zinc oxide shuts out the most ... out almost all sunlight, and that is not good for your body [health] to have ... or zinc is no good at all ... because some are allergic to zinc and then that is not good. I mean metals and such things, but I think it's which one gives a better protection than zinc oxide [Option D], so it's closer to D, I'd say.

Cathrine: Is it this one (points)? *Is there* (marking D on the test sheet) (Group N: 19:01–20:39)

Edward's concern seems to be the properties of zinc oxide causing allergies. He also argues that 'shutting out sunlight ... is not good for your body'. Perhaps he refers to scientific data stating sunlight to be crucial for the bodily production of vitamin D.

Finding a better (bättre) sun protection (as stated in Option D of the Swedish version of the test) could be what the experiment is about. In the situated adjustments to the test instrument, another incorrect answer is selected, and another sign of low performance is produced. Once again, we can see how the boundaries dissolve for what is a scientific question, this time in relation to other biological knowledge (health and allergy). What Edward performs is an alternative scientific repertoire, which is not valid in this specific test item. Furthermore, language translation entrains the kind of translation that we know from Latour (1999): in each step of translation, meanings are added and lost.

Good that we are so damn stupid

We will now return to the PISA documents in which (low) performance is discursively constructed. In this section, we briefly describe how students in these statistically construed categories come into being as at risk, not so much for themselves, but for a successful future society. In one document, written in cooperation between OECD, the National Center on Education, and the Economy in Washington (although ‘on the responsibility of the Secretary-General of the OECD’, as stated in the foreword), the lowest category of performance acknowledged as literacy (in reading) is described as follows:

Level 2 on the PISA reading scale can be considered a baseline level of proficiency, at which students begin to demonstrate the reading competencies that will enable them to participate effectively and productively in life. Students proficient at Level 2 are capable of very basic tasks, such as locating information that meets several conditions, making comparisons or contrasts around a single feature, working out what a well-defined part of a text means even when the information is not prominent, and making connections between the text and personal experience. (OECD, 2010, p. 29)

This passage indicates that students at Level 2 barely have competencies that will enable them to participate in society effectively and productively. Our concern here is not to belittle reading skills, but to show how only certain interests and skills are packed into categories of *performance*, leaving other knowledge and preferences behind and unnoticed (c.f. Biesta, 2010). In another text (OECD, 2013), this category of students is seen as plausibly not able to ‘function effectively in the workplace or in society’ (p. 254), or as potential threats (Popkewitz, 2008) in this economic-oriented, individualistic discourse of performance (Biesta, 2010; c.f. Tröhler, 2013). Low performers become performed as non-contributors in an imagined future society or, to push it ever further, as insufficiently producing knowledge machines. The main point is how attributes are linked to specific categories of students, performed into being as a particular type of people.

The instrument actor not only produces low performance as discursive or statistical categories (marks on the test that do not give scores) but also in the quite opposite way. This occurs in events in which students experience themselves as insufficient or incapable, such as in the following encounter of students and the ‘Acid Rain’ unit (S485).

ACID RAIN

Below is a photo of statues called Caryatids that were built on the Acropolis in Athens more than 2500 years ago. The statues are made of a type of rock called marble. Marble is composed of calcium carbonate.

In 1980, the original statues were transferred inside the museum of the Acropolis and were replaced by replicas. The original statues were being eaten away by acid rain.



Figure 4: The introductory text in the PISA unit on Acid Rain (S485).

Science Literacy

Question 1: ACID RAIN

S485Q02 - 0 1 2 9

Question intent: Explaining phenomena scientifically

Normal rain is slightly acidic because it has absorbed some carbon dioxide from the air. Acid rain is more acidic than normal rain because it has absorbed gases like sulphur oxides and nitrogen oxides as well.

Where do these sulphur oxides and nitrogen oxides in the air come from?

.....

Figure 5: The first test question in Acid Rain.

Théa, Amanda, and Mickey hesitated to participate in this study, but finally decided to do so. Théa is reading the introducing text to the unit 'Acid Rain' (S485) aloud. In spite of the resistance that this encounter entails, she persistently sounds out the words that are hard for her to pronounce, while commenting on them verbally, and grimacing.

Théa: [Reading aloud] Question Two acid rain: *Normal rain is slightly acidic because it has ... it has absorba -sorbed [stumbles] some carbon dioxide from the air. Acid rain is more acidic than normal*

rain because it has abos- abos – absorbed – oh my God I read so bad – gases like sulfur oxides and nitrogen oxides... oxides I mean Ox-i-d-es [trying the word] as well. Where do these sulfur oxides [giggling] this sounds so brainless ... and nitrogen oxides in the air come from?

[The girls are looking at each other, puzzled]

Théa: Good that we are so damn stupid.

Mickey: The earth?

Théa: But can't ... can't one Google it? [pointing at a computer]

Mickey: [giggles nervously]

Théa: But just answer the question 118 118

Mickey: The air's ...?

Amanda [very quietly]: Does one have to answer?

Théa: Well, I think all questions are this hard. (Group E, Nov 2010, 12:00–16:00)

Since the lines on the testing sheet remain empty, PISA and this group of students would agree in one aspect. These girls 'are' poor performers (e.g., OECD, 2010, p. 29). But in what sense were these girls poor performers before being persuaded to answer these test questions? If we see their hesitation toward participating as a sign of remembered performances in the past, yes perhaps they were poor performers. However, if performance is produced in relation to the conditions and circumstances there and then, isn't it in encounters such as this that low performance is brought into being? Let's study this aspect closely: When Théa reads aloud, the words are hard to pronounce, so she stumbles and accuses herself of being a bad reader, and the text [it] of being brainless. Théa also ironically targets the group's standards of performance: 'Good that we are so damn stupid'. There initially is resistance, then accusation, then more accusation. However, there is also a solution: While Mickey twice suggests responses to the test question, Théa directs attention toward the nature of the knowledge asked for here. It is about facts (the competency *Explaining phenomena scientifically*), Théa suggests dialing the number 118 118 (to the company Eniro, which advertises that it can answer any question) or Googling the web. These acts are connected to how facts are found in the real life with which PISA is concerned. In the last statement, Amanda asks if they really have to answer this question. Théa seems to think yes, and that all questions in the test are as hard as this one.

In real life, the skills and abilities for which this task asks would probably have performance among the conditions to emerge, with Eniro or the web as co-actors. However, according to PISA conditions, only poor performance results. We can assume that the lack of the students' experience at these moments is also one of the realities coming into existence. In parallel to their hesitating resistance toward this task, we discern a failing endeavour, an inward accommodation (Pickering, 1995) that ends in what appears as disappointment about their self-value. Therefore, the text and its items constitute a double-edged sword. It produces a non-response due to its way of detecting that depends on the meanings of small details, which equals lack of competency. Meanwhile, it also produces the situated experience of low performance and lack of something assumed to be known. This latter experience, of course, risks

being multiplied and reproduced in the assessment over and over again: ‘Well I think all questions are this hard’.

We don’t know if, and to what extent, these students are aware that PISA is constructing them as threatening the future. However, we can see that they are inside a measurement rationality, assessing themselves in relation to cultural standards for scientifically competent citizens. To summarize, this data material of students and PISA scientific literacy test item encounters suggests innumerable different, yet similar, mangling practices (Pickering, 1995) in which performance is produced. How can the situations we present here guide our understanding of what PISA performance can be? What do they tell us about the categories of literacy and competency?

Unpacking the notion of low performance: Conclusion and discussion

For this article, we gave ourselves the task of unpacking the taken-for-granted notion of *low performance* by looking into how scientific literacy competencies are constructed in PISA documents and test items. We also studied how lack of competence comes into existence in item-student encounters, in which students adjust their operations and are adjusted in their operations with the PISA science test instrument (OECD, 2006) as a detector (Knorr Cetina, 1999). All these events are participations in the contemporary production of truth (Foucault, 1971; 1980) about what performance is and should be. The performing citizen-to-be is a discursive reality of PISA frameworks, reports, and analyses. They emerge as willing to engage, with science as an always-present endeavour, and do not fail to make the right, scientifically informed decisions. The *poor performer* (e.g., OECD, 2010) is the idea of the opposite, but performed into being as a category in themselves as students at risk. Their greatest failure probably concerns a lack of effectiveness, non-contribution, the danger of someone who makes the wrong decisions, and who does not express interest in the mandatory scientific project. As we argue in this article, all evidence of performance in the kind of standardized testing of which PISA is part results from multiple translations (Gorur, 2011; Popkewitz, 2011). We tried to construct and approach events in which marks on test sheets are produced as part of imagined translations (because the students in these encounters were not part of an actual PISA test).

Some of the most striking truths about current education are the political concern for league tables (Steiner-Khamsi, 2003), that the PISA result of a nation is of specific interest in comparison with other nations, and that a student’s achievement is valued in comparison with others on equal scales. As a consequence of this logic of competition, most countries in the world strive for achieving a top position in the PISA. Any other result is considered failure (Kamens, 2013). As argued before, the more trust put into the PISA project, the truer (in a Foucauldian sense) it becomes. In Sweden (the country in which we began formulating our questions about performance), marking, grading, and comparing are the suggested tools for achieving more and more.

However, this article questions these realities, and argues that, similar to Pasteur’s microbes (Latour, 1999), the performances that PISA speaks of are created in the very moments of scientific measurement. They are the products of resistances and accommodations (Pickering, 1995) enacted by students in their situated adjustment to the items presented to them.

Expanding the analysis, the results are also products of the test constructors in their adjustment toward the requirements of the framework and the results of the field trials of the test items. Test constructors are asked to refine formulations, place the tasks in fictional real-life contexts in which science is ever-present, or exclude items with wrong statistical attributes. Moreover, we argue that the methods for measuring students' knowledge, performed with the student by the instrument actor itself, actually teach performance (and low performance) as well. Théa ironically voiced this: 'Good that we are so damn stupid'.

Popkewitz (2011) argues that the numbers used in assessments like PISA create distance from the phenomena of concern, 'by appearing to summarize complex events and transactions' (p. 33). One of those things that become distant is the invention/construction of the categories within which students perform; the *a priori* defined PISA competencies that we claim are needed to see the concept of *low performance* as something (Latour, 1999). We argue that the boundaries of these categories shape the knowledge to be assessed and the attributes by which it comes into existence. In this way, the obviously thing-like object for thought (Latour, 1999)—universal scientific competencies—gains reality. In this article, we examine these discursive and material boundaries and the test situation as a scientific experiment in which these boundaries travel across the mangle (Pickering, 1995) to perform performance (and low performance). As such, they become the 'tool that creates and allows the creation of new problems and imagined new tomorrows' (Carvalho & Costa, 2014, p. 7).

Morgan (2009, p. 195) asserts that PISA answers our questions as long as we let it do so. It is our unremitting use of PISA that gives it a powerful position. We fear not only how PISA 'speaks matter-of-factly' (Gorur, 2011) about statistically evidenced realities of good and bad performance, but also how this is encapsulated in producing truth (Foucault, 1971) about what constitutes important knowledge. Most of all, we fear how the production of experiences of failure are performed into being during assessment. The policy debate on school tends to focus on what appears as measurable and universal, both of which are attributes that are questioned in this article. We hope this debate one day will talk more about the complexity of educational matters and the purpose of education in a deeper sense (Biesta, 2010). To do that, trust must be placed in students, teachers, and headmasters, rather than PISA. Most likely, people in those positions carry other stories about what high and low performance could be and how performance is entangled in a particular—rather than a global and comparable—context.

Acknowledgment

This work was supported by The Swedish Research Council under Grant number 721-2008-4717

Note

¹ https://nces.ed.gov/surveys/pisa/pdf/items2_science.pdf

² www.siris.skolverket.se

References

- Arffman, I. (2010). Equivalence of translations in international reading literacy studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59.
- Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Boulder, CO: Paradigm Publisher.
- Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching*, 46(8), 865–883.
- Carvalho, L.M. (2012). The fabrications and travels of a knowledge-policy instrument. *European Educational Research Journal*, 11(2), 172–188.
- Carvalho, L.M., & Costa, E. (2014). Seeing education with one's own eyes and through PISA lenses: Considerations of the reception of PISA in European countries. *Discourse: Studies in the Cultural Politics of Education*, DOI: 10.1080/01596306.2013.871449
- Eckström, A. (2013). Broadcast from press conference of the Swedish PISA results. 3 December 2013.
- Fertig, M. (2004). What can we learn from international student performance studies? Some methodological remarks. RWI: Discussion Paper, No. 23. Essen: RWI.
- Foucault, M. (1971). *L'ordre du discours*. Paris: Editions Gallimard.
- Foucault, M. (1980). *Power/Knowledge. Selected interviews and other writings 1972-1977 by Michel Foucault*, (Ed. C. Gordon). Harlow: Harvester Press Limited.
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*, 11(3), 319–330.
- Gorur, R. (2011). ANT on the PISA trail: Following the statistical pursuit of certainty. *Educational Philosophy and Theory*, 43(1), 76–93.
- Gorur, R., & Wu, M. (2014). Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse: Studies in the Cultural Politics of Education*, DOI: 10.1080/01596306.2014.930020
- Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Grek, S. (2012). What PISA knows and can do: Studying the role of national actors in the making of PISA. *European Educational Research Journal*, 11(2), 243–254.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20(2), 225–240.
- Ideland, M. & Malmberg, C. (2014). 'Our common world' belongs to 'us': Constructions of otherness in education for sustainable development, *Critical Studies in Education*, DOI: 10.1080/17508487.2014.936890
- Kamens, D.H. (2013). Globalization and the emergence of an audit culture: PISA and the search for 'best practices' and magic bullets. In H-D. Meyer, & A. Benavot (Eds.), *PISA, power and policy: The emergence of global educational governance* (pp. 117–140). Oxford: Symposium Books.
- Knorr-Cetina, K. (1999). *Epistemic cultures. How the sciences make knowledge*. Cambridge, Mass.: Harvard University Press.
- Kreiner, S. (2011). *Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment*. Copenhagen: University of Copenhagen.
- Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225-248.
- Lawn, M., & Grek, S. (Eds.) (2012). *Europeanizing education: Governing a new policy space*. Oxford: Symposium Books.

- Meyer, H.-D., & Benavot, A. (Eds.). (2013). *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books.
- Morgan, C. (2009). *The OECD programme for international student assessment: Unravelling a knowledge network*. Saarbrücken: VDM Verlag Dr Müller.
- Nóvoa, A. (2013). The blindness of Europe: New fabrications in the European educational space. *Sisyphus Journal of Education*, 1(1), 105–124.
- Organisation for Economic Co-operation and Development (OECD). (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. OECD Publications.
- OECD. (1999). *Measuring student knowledge and skills: a new framework for assessment*. Paris: Organisation for Economic Co-operation and Development (OECD).
- OECD. (2003). *The PISA 2003 assessment framework: mathematics, reading, science and problem solving knowledge and skills*. Paris: Organisation for Economic Co-operation and Development (OECD).
- OECD. (2007). *PISA 2006. Science competencies for tomorrow's world (Vol. 1, Analysis)*. OECD Publications.
- OECD. (2009). *PISA 2009 Assessment Framework. Key competencies in reading, mathematics and science*. OECD Publications.
- OECD. (2010). *Strong performers and successful reformers in education. Lessons from PISA for the United States*. OECD Publications.
- OECD. (2013). *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science (Vol. 1)*. OECD Publications.
- Pickering, A. (1995). *The mangle of practice: Time, agency and science*. Chicago: University of Chicago Press.
- Popkewitz, T. (2008). *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*. New York: Routledge.
- Popkewitz, T. (2011). PISA. In M.A. Pereyra, H.G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 31–46). Sense Publishers.
- Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Puchhammer, M. (2007). Language-based item analysis: Problems in intercultural comparisons. In S. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA according to PISA: Does PISA keep what it promises?* (pp. 127–138). Wien: LIT.
- Roth, W.-M. (2007). Toward a dialectical notion and praxis of scientific literacy. *Journal of Curriculum Studies*, 39(4), 377–398.
- Serder, M. & Jakobsson, A. (2014). Why bother so incredibly much? PISA assignments from a student perspective. *Cultural Studies of Science Education*. In press (available electronically).
- Serder, M. & Jakobsson, A. (in review). Language games: Meaning as use in student encounters with scientific literacy test items.
- Serder, M. & Sørensen, H. (2014). Comparing Danish and Swedish versions of PISA scientific literacy. Conference paper presented at NFSUN in Helsinki, June 4, 2014.
- Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Conceptual knowledge in talk and text: What does it take to understand a science question? *Instructional Science*, 29(3), 213–236.
- Simons, M. (2014). Governing education without reform: The power of the example. *Discourse: Studies in the Cultural Politics of Education*, DOI: 10.1080/01596306.2014.892660

- Skolverket. (2007). *PISA 2006: 15-åringars förmåga att förstå, tolka och reflektera – naturvetenskap, matematik och läsförståelse*, No. 306. Stockholm: Skolverket.
- Steiner-Khamisi, G. (2003). The politics of league tables. *Journal of Social Science Education*, 2(1), 1-6.
- Tröhler, D. (2013). The OECD and cold war culture: Thinking historically about PISA. In H.-D. Meyer, & A. Benavot (Eds.): *PISA, power and policy: The emergence of global educational governance* (pp. 141–161). Oxford: Symposium Books.