In the grey zone: large-scale assessment-based activities betwixt and between policy, research and practice

Daniel Pettersson\textsuperscript{a,b}, Thomas S. Popkewitz\textsuperscript{c} and Sverker Lindblad\textsuperscript{b,d}

\textsuperscript{a}Faculty of Education and Business Studies, University of Gävle, Gävle, Sweden; \textsuperscript{b}Department of Education, Uppsala University, Uppsala, Sweden; \textsuperscript{c}Department of Curriculum and Instruction, University of Wisconsin-Madison, Madison, WI, USA; \textsuperscript{d}Department of Education and Special Education, Gothenburg University, Gothenburg, Sweden

\textbf{ABSTRACT}

This paper elaborates on a systematic research review containing 11,000 articles on international large-scale assessment (ILSA) research. Several activities operating under the ‘formal radar’ of science and governmental policy are observed, which we analytically name ‘grey zone’ activities. These activities are historicised, presented and discussed. An analytical division into three different reasons for performing the activities is made: an entrepreneurial policy reason, an entrepreneurial profitable reason and an appurtenance reason. This division highlights some of the actors in the educational grey zone. The paper is theoretical and elaborate, and contains examples of the activities that can be found in a grey zone.

\textbf{ARTICLE HISTORY}

Accepted 25 March 2017

\textbf{KEYWORDS}

Policy; large-scale assessments

\section*{Introduction}

This paper historicises what we analytically claim to be ‘grey zone’ activities involved in the interpretation of international assessments of student performance and how International Association for the Evaluation of Educational Achievement (IEA) and Organisation for Economic Co-Operation and Development (OECD) comparative assessments (such as e.g. Trends in Mathematics and Science Study (TIMSS) and Programme for International Student Assessment [PISA]) use a specific reasoning (Hacking, 1992) to describe these activities. In many respects, these activities are dependent on socio-cognitive networks, referred to as either ‘invisible colleges’ (Crane, 1972), ‘scholarly tribes’ (Becher & Trowler, 2001) or different ‘epistemic cultures’ (Knorr Cetina & Cicourel, 1981). Study of the dissemination and production of knowledge is a well-established disciplinary field, with notable scholars such as Jasanoff (2004), Merton (1973), Latour (1988), Latour and Woolgar (1979), and Shapin and Schaffer (1985). Their studies have investigated how knowledge transforms into ‘facts’ and ‘truths’ in a process of rivalry. This study is connected to these two aspects, and in particular investigates how reasoning within the sphere of (ILSA) is developed and constituted, and the activities that emanate from this reasoning. Further, we discuss the similarities and differences in the reasoning between them.

The study is performed within a tradition of historicising (Popkewitz, 2015). Historicising is performed to understand how multiple and unrelated events and trajectories come together as a grid at different times to shed new light on the self and objects of reflection and action. The notion of the grid can be likened to a cake recipe, in the sense that ingredients are mixed together to create a new ontological existence – ‘the cake’. In this paper, we present some of the ingredients that make up ‘the cake’ of today’s educational landscape. The study locates significant events and trajectories. These can be considered as units of analysis to highlight what are regarded as important ‘new’ actors and reasoning in the international field of education. The paper can be interpreted as a theoretical contribution to the understanding of some of the educational activities that ‘go under the radar’ in terms of grey-zone activities.

The idea of (and term) the grey zone emerged in a previous study on activities using or discussing numerical data produced by ILSA. In this study, which is a systematic research review consisting of some 11,000 scientific articles identified via the database Scopus (Lindblad, Pettersson, & Popkewitz, 2015), the conclusion is that in recent decades, a number of new influential international agencies have emerged to interpret the international school system results and make recommendations and specific national proposals for educational improvements (e.g. McKinsey and Pearson). When analysing the articles, it was found that more regional based agencies were now translating and adapting ILSA reasoning to non-Western world contexts (e.g. Latin American Laboratory for the Assessment of the Quality of Education [LLECA], Southern and
Eastern African Consortium for Monitoring Educational Quality [SACMEQ], Conférence des Ministres de l’Éducation des Pays Ayant le Français en Partage [COMFEMEN], Twaweza and Pratham). In this paper, these observations are taken one step further and historicise the located agencies as producing activities to use or spread specific ILSA reasoning and to make claims about educational development. This is done by analysing documents published by the different agencies, by Web searches and by analysing the various agencies’ websites. Through this process, the idea is to identify the specific reasoning of the different agencies that describes the historical development of international large-scale assessments. We are aware that the agencies discussed here differ greatly in character and scope. It should also be noted that the agencies discussed in the paper are only a few examples of all the possible agencies involved in these kinds of activities.

The analytical concept of the grey zone derives from the status of international comparisons and the impact of transnational expertise in current education discourses. Within this development, a comparativistic paradigm in education can be noted. Here, comparisons are less about education in different contexts and more about hierarchies of performances. In this context, the comparison of school systems and their performances serves as a base and focuses on ranking positions and performance trajectories over time. This comparativistic paradigm, which makes use of new algorithms to classify educational systems at the intersection of international actors and national policy and science, is repeatedly expressed in education policy debates, in the mass media and in conversations with transnational education experts about how to improve education. Hence, we analytically claim that there is an educational space, conceptualised as a grey zone, that can be investigated in terms of who has access, what is on the agenda and which principles and preferences are at work. We devote our attention to presenting the different activities, and suggest reasons for participating in a discussion about education. This can also be understood in terms of a crossroads between and betwixt policy, research and practice.

Grey-zone activities operate under the formal radar when research output or policy arenas are examined. This can be because these kinds of activities do not come under the canons of traditional research communities, or have the same responsibility as policymakers or elected officials. However, to a large extent they conform to principles and knowledge that delineate the problems of education according to its ‘needs’, statements about educational systems and the targets of change. In our view, the location of grey-zone activities is important for understanding how theories and abstractions of the social (and educational) sciences ‘act’ when mobilising and ordering social problems and methods of change. The discursive area in which agencies operate is not simply an area between and betwixt policy, research and practice; it is also a site for constituting educational judgements, recognising types of objects and drawing conclusions about manageable fields of existence that are not just numbers. The activities in this site also signal societal beliefs about how education is conceived and, perhaps more importantly, what it should be like.

ILSA produces human kinds and differences that exclude individuals and groups from creating characteristics and hierarchies of individuals. ILSA activities, such as the studies produced by IEA and OECD, can also, to some extent, be considered grey-zone activities. However, this paper elaborates on how the data and reasoning emanating from studies such as TIMSS and PISA are used in activities other than governmental and traditional research. Instead, we locate some of the influential agencies that either use ILSA data or reasoning for policymaking or profit (in this, we specifically highlight McKinsey and Pearson), or use specific ILSA reasoning to stage comparative assessments outside the traditional Western world (e.g. LLECA, SACMEQ, COMFEMEN, Twaweza and Pratham). Drawing on Porter (2012), we state that the seemingly technical appearance of the numbers created by ILSA enters into cultural realms and creates a situation in which numbers become a codification and standardisation of what constitutes reality and planning. As such, numbers operate as road maps to a desired future. They can also be road maps to what is not desired – not only in education but also in society at large.

Before describing some grey-zone activities, an outline of the reasoning behind ILSA is provided. This is followed by a brief description of two of the most well-known producers of ILSA – IEA and OECD. This is done in order to understand the context in which grey-zone activities have evolved. Our main research interest is in how numbers have historically characterised education. Hacking (1983) argued that numbers are parts of systems of communication in which technologies create distances from phenomena by appearing to summarise complex events and transactions. The seeming rigour and uniformity appear to be transported across time and space so as to not require intimate knowledge and personal trust. This creates a specific relation between education and a specific technology that is framed by ideologies of modernity and meritocracy and is understood as a selection of different and hierarchical positions in society by means of educational performance. From this point of view, education can be given a meritocratic meaning (cf. e.g. Bourdieu, 1971) that numbers become a way of.
assessing and measuring educational performance. In describing this development historically, the current study is framed within the scientific fields of educational history and the sociology of knowledge. Empirically, we historicise activities in the contemporary educational discourse, where numbers have been important for describing education. These activities are analytically and theoretically developed from the various reasoning on involvement in education.

**Faith in numbers**

In explaining the development of educational research during the 20th century, Lagemann (2000) claimed that, in an American context, ‘One cannot understand the history of education […] unless one realises that Edward L. Thorndike won and John Dewey lost’ (p. xi). Whether we agree with this claim or not, it has to be admitted that research using quantitative (e.g. statistical) techniques multiplied during the first decades of the 20th century. Many different, yet often interrelated, factors were responsible for this development, such as the acceptance of positivism as a dominant scientific reasoning, a rapid growth of educational institutions with an interest in comparisons and data usage, the socialisation of new researchers influenced by the use of statistics in psychology (Stigler, 1992; cf. Hacking, 1992), a supremacy of meritocratic values in modern societies (Pettersson, Popkewitz, & Lindblad, 2016) and the constant need to legitimise these through ‘objective’ and ‘neutral’ research (Smeyers & Depaepe, 2010; cf. Porter, 1995).

Thorndike maintained that the method of testing and the use of statistics is central, as is the belief that everything in life can be measured. The increased use of statistics to address societal and research issues can be expressed as striving to achieve *ordo ab chao* (to create order out of chaos). At the time, this is expressed in terms such as ‘whatever exists at all, exists in some amount […] anything that exists in amount can be measured […] measurement in education is in general the same as measurement in the physical sciences’ (McCall, 1922, pp. 3–5). Statistics as such came to be both an academic discipline and part of a broader educational context. The power and the efficiency of statistics gave rise to a faith in measurement and metrics (Smeyers & Depaepe, 2010; cf. Porter, 1995). The growth of scientific statistics as a dominating reasoning creates a belief that the more data we gather and the more comparisons we make, the more we will know. The use of comparisons and data in statistics carries a number of presuppositions: that reality can be represented in numbers, that it can be controlled and that risks can be managed (Pettersson et al., 2016).

When educationalists embraced the new method of testing and statistics, at a societal level this is nothing new (see e.g. Igo, 2007). During the 19th century, a desire for internal educational reforms led to a quest for data and information about education in other countries. By turning to a more empirical approach consisting of data and comparisons, social science was seemingly able to distance itself from its moral and progressive value-laden social-activist roots and appear to be neutral. Consequently, the empirical turn created an opportunity for social science to act in a way that in the past was more connected to natural science.

The empirical turn also led to the manifestation of the new science of statistics. Statistics created facts about social life and became part of the change that traversed different social sectors (Poovey, 1998). Moreover, statistical comparisons led to recognition of differences between nomenclatures as a problem to be eliminated. In doing so, a grid was constructed that appeared to be valid for and unresponsive to national contexts or time. Hence, information about contemporaneous taxonomies was preserved rather than dissolved. This view, as discussed by Desrosières (1991), also marked a clear rupture from the classical ways of social science, where numbers were used to describe things that existed independently of the conventions that established them. The act of coding therefore constructed equivalence classes between diverse objects, which meant that the class, rather than individual objects, was judged and described. The process of constructing equivalence classes thus made objects comparable. This resulted in the individual being lost in favour of overall descriptions and numbers, which were then used for an overall emphasis, rather than an emphasis on the individual (cf. Igo, 2007; Lemann, 1999). As such, numbers were seen as a technology of distance and were used to claim objectivity instantiated by moral and political discourses (Porter, 1995).

Numbers thus became visualised as social facts, whose objectivity was important in the making of citizens. Consequently, numbers are regarded as a social technology that creates consensus and harmony in the world. The uniformity that numbers give brings order into social life by regulating relations (Rose, 1999). However, although numbers ‘act’ as real, they also embody implicit choices about ‘what to measure, how to measure it, how often to measure it and how to present and interpret the results’ (Rose, 1999, p. 199).

In sum, it can be stated that in order to understand the qualities of governing, we first need to consider numbers as defining a problematised space, where subjects and objects are stabilised. Numbers are seen as technical, objective and calculable – a calculation technology – and embody the idea of everybody
having equal opportunities and representation. Numbers also standardise the subject of measurement and assessment and the act of exchange, so that they are no longer seen as dependent on the personalities or status of those who perform the measurements or assessments. We will now present a specific educational branch of using numbers to say something about the world – large-scale assessments.

The phenomenon of large-scale assessments

Countries inspired by the importance of statistical comparisons and calculation technology have developed tools and techniques for evaluation and assessment as part of their efforts to improve student learning outcomes. This came about when education became a central requirement for national economic development and political democratisation. Since the end of the 19th century, the production of numerical data has been used to create new visions of the social and economic world. The new construction of epistemic references to define ‘reality’ is linked to the creation and management of the self-defined ‘democratic’ state. Numerical data also provides more than an ‘objective way’ of seeing reality. It ‘institutes’ reality by creating a ‘common cognitive space’ that can be both observed and described through data (Lussi Borer & Lawn, 2013). This ‘common cognitive space’ has been framed and developed by, for instance, the reasoning on international assessments.

Data has gradually been considered a more objective way of understanding this ‘reality’ (Lussi Borer & Lawn, 2013). One offspring is ILSA, which measures and presents student learning outcomes. It was created with the vision that if custom and the law defines what is educationally permissible within a nation, educational systems beyond national boundaries could suggest what is possible educationally (Foshay, Thorndike, Hotyat, Pidgeon, & Walker, 1962). This argument was used to introduce a pilot study in the late 1950s that not only described the origins of an emerging field, but also predicted exceptional growth in comparative assessments (Owens, 2013).

IEA: the founder of an assessment technology

One of the first organisations to be formally established to perform ILSA was the IEA. The organisation initially conceived the world as a natural education laboratory, where different school systems were experimented with to achieve optimal results. It assumed that science could obtain evidence from different national education systems and make education more effective (Pettersson, 2014a). The first assessment performed by IEA differed from previous comparative studies in that it sought to introduce an explicit empirical approach into the methodology of comparative education – a field initially said to rely on cultural analysis (Foshay et al., 1962, cf. Kazamias, 2009). IEA embarked on the task and ran a pilot study (beginning in June 1959 and ending in June 1961), which concluded that cross-national comparisons of educational performance could be made with satisfying results (Foshay et al., 1962). Such findings were startling at the time, but even more important was the clear sense that researchers from different cultures and educational systems could agree on a common approach to testing and evaluation (Purves, 1987).

In 1961, researchers from 12 countries met to discuss the pilot study. The study was considered a success and plans for another study in mathematics took shape. The major purpose of the inquiry was to measure mathematical achievement and relate that achievement to the relevant factors in the home, school and society. The project, called the First International Mathematics Study (FIMS), attempted to assess the efficiency or productivity of different educational systems and practices (Bloom, 1969). The study revealed that there was a difference between how a subject was actually taught in the classroom and how it was described in the curriculum, and that this was a good predictor of the differences in student performance. FIMS also showed that there was a lack of equity between the performances of different groups of students. After this study, IEA performed a variety of studies on different subjects, time spans and periodicity, in which TIMSS today is the most recognised (see Lindblad et al., 2015 for a compilation of the various assessments).

OECD gets involved

Indeed, the IEA studies led to many assessments being undertaken in various countries and subjects. The PISA study, a project of OECD, was similar to the IEA studies in many respects. Although OECD has primarily been concerned with economic policy, education has become increasingly important (Pettersson, 2008). By means of statistics, reports and studies, OECD has activated a ‘common-sense’ approach to political decision-making by stating that scientific ‘proofs’ are indisputable (Martens, 2007). Martens (2007) argued that OECD’s greatest impact can be seen in its agenda with indicators and its role in constructing a global policy field of governance by comparison (cf. Carvalho, 2012; Grek, 2009). Nóvoa and Lord (2002) stated that comparisons may not be regarded as a method, but can in fact be seen as policy. The policy is driven by an expert discourse that, by means of comparative strategies, tends to impose common-sense answers in national settings.
(cf. Pettersson, 2008). While OECD serves national policy makers well with a comparable discourse in terms of statistics, it also provides them with a global policy lexicon for what education is and ought to be (cf. Pettersson, 2014b).

PISA assessments started in the year 2000, and have been conducted several times since then. In every assessment, students’ literacy, scientific literacy and knowledge of mathematics are tested, together with consideration of their interests and backgrounds. The emphasis is on ‘real-life’ circumstances and the capacity to enter the labour market with the relevant skills. This has been said to shift PISA’s focus away from less explicit educational aims that are more complicated to measure (Grek, 2009).

Even though PISA is both constructed and operates under a clear policy framework that is designed to improve future results, it is not just a testing regime. PISA should also be seen in the light of its ability to improve and attract economic and human capital investments. For policy makers, PISA is thus a two-sided coin in that it tests outcomes and attracts economic investment. In view of this, PISA can be said to have two functions – economic and educational (Pettersson, 2008). As these two aspects are interwoven and strengthen each other, they can hardly be analysed separately. Besides the cycles of PISA, OECD has also staged and presented a number of other studies (see Lindblad et al., 2015 for a compilation of the various assessments).

The entrance of new activities

ILSA results, in the shape of numerical performances that are comparable, are sent to governmental and non-governmental organisations, where they are used as descriptors of the quality of education systems. Our analysis explores how the ‘truth’ about educational matters is told by elaborating on the discursive area betwixt and between policies, research and practice in which the ILSA discourse operates. With regard to the field of ILSA activities, we find that we can no longer only use the traditional division of governmental and non-governmental organisations to describe how ILSA data is used and reasoned about and, in addition, cannot only focus on research activities. More recent developments of agencies using ILSA data, and the reasoning for staging the various activities or being inspired by ILSA methodology and techniques to stage their own assessments, are therefore exemplified. Here, three different but interconnected and intertwined constructed categories are used to describe the activities taking place in the discursive area we analytically call the grey zone. First, international entrepreneurial organisations use ILSA data for their business ventures under the slogan of giving national policymakers advice for change. Second, companies involved in what has been called the Global Education Industry (GEI) (e.g. Verger, Lubienksi, & Steiner-Khamsi, 2016) use ILSA to make a profit. Although both these categories have a lot in common, there are also differences. The most important difference is at the rhetorical level in relation to the explicit claims about ‘advice for change’ or ‘selling advice’. In one sense, both these claims ‘sell’ solutions, but there is a rhetorical difference that is further exemplified and discussed in the empirical part of the study relating to the differences in reasoning. The third category is a related yet scattered field connected to ILSA, in which assessments are developed by more locally active organisations that are largely inspired by the ILSA methodology. This field is interesting for several reasons, one of which is that organisations import ILSA reasoning as a technology to stage their own assessments of education, mainly to become influential in policy making outside the traditional Western world.

In the following, three different intellectual reasons for activities are presented, all of which use ILSA as inspiration. We call these reasons entrepreneurial policy reasons, entrepreneurial profitable reasons and appurtenance reasons. We are aware of the ambiguity of the term ‘appurtenance’, which in its simplest lexicographic form means ‘belongingness’, although it is also used in a legal context and in Gestalt theory. The Gestalt theory meaning implies a relation between two things that influence each other, thereby creating new interpretations of what is actually seen. In a legal sense, the term implies relations between two objects. In 1919, appurtenance was described in an American legal context by the Supreme Court of Minnesota as an object that ‘[…] belongs to something else. Something annexed to another thing more worthy’ (Supreme Court of Minnesota, 1919). When we use appurtenance reasons to describe the assessments that take place outside the traditional Western world using ILSA technology as inspiration, we elaborate on all these understandings of the term.

The activities in the above three reason categories are historicised and discussed in the examples of agencies with a new or rejuvenated interest in ILSA data. The ways in which the agencies are constituted and organised in their work with ILSA data is also presented. The activities in the three categories differ in their scope and aim, which will become obvious from the given examples. It is important to discuss them as similar phenomena that are facilitated by the analytically constructed concept of grey-zone activities using or inspired by ILSA reasoning and the data produced in this context. By giving examples, we argue that ILSA reasoning does not just take place in traditional governmental and non-governmental
organisations and research, but is also apparent in an international entrepreneurial policy sector, in the
GEI, and in assessments impacting local policy in large parts of the non-Western world. ILSA is, as such, described as a conglomerate of activities, and is not only limited to surveys such as TIMSS and PISA, national governmental reports presenting national results, or research. Instead, ILSA is part of a globally spread reasoning, interpreted as ‘common sense’ or ‘evidence’, which creates educational ‘facts’ and ‘truths’ that are evident in contemporary statements on education. Below, we historicise the three different reasons and give examples of how they are discussed and presented.

**Entrepreneurial policy reasons: the McKinsey activities**

In 2007, McKinsey published the report *How the World’s Best-Performing School Systems Come Out on Top* (Barber & Mourshed, 2007). The report addressed a question that is often asked by policy makers: ‘How does a system with modest performance become great?’ In 2010, another report was published called *How the World’s Most Improved School Systems Keep Getting Better* (Mourshed, Chijioke, & Barber, 2010). In the foreword, Professor Emeritus Michael Fullan, from the University of Toronto, stated the following:

There is a recent and rapidly growing appetite for figuring out and accomplishing what I call ‘whole system reform’ – how to improve all schools in a district, a region, a state, province of country. For a long time, there has been the realization that better education is the key to societal and global productivity and personal and social well-being. Only recently are we beginning to see that interest turn into specific questions about how you actually go about whole system reform. What pathways, from what starting points, are going to get results in reasonably short time frames? How do we actually 'raise the bar and close the gap' for all students? (Fullan, in Mourshed et al., 2010, p. 6)

The report is presented as filling a particular vacuum in the policy field in order to understand the complexities through a specific application of organisational system models and management theories. Hence, in its reports McKinsey presents an objective representation of the functioning of educational systems using, for instance, ILSA data. By using different data to measure the output of national systems, McKinsey is able to compare national systems and create hierarchies between successful and less successful systems.

McKinsey and Company is a private consulting firm customised for large corporations. Its reports are self-financed and provide recommendations for the public and governmental policy arenas. McKinsey states that it is one of the oldest global management consulting firms, with a social obligation to address problems that are not only of importance for the specific client, but also highlight national and human development issues. The latter societal concerns are not independent of the company’s core function of management, but are part of its larger corporate public responsibility. For this reason, McKinsey has created a non-profit economic think tank, the McKinsey Global Institute, which provides management knowledge for foundations, non-profit organisations and multilateral institutions on issues related to disease, poverty, climate change and natural disasters.3

The educational reports that McKinsey has produced embody this societal commitment by providing knowledge that is said to make a contribution to solving complex societal problems. One thing that becomes apparent in the reports is that they all motivate their existence with the argument that as school systems do not examine themselves in terms of success, the expertise of the company can contribute to international educational improvement. The company says that it does this by identifying how specific elements of educational systems have broader universal relevance for proving significant, sustained and widespread student outcome gains, and examines why it has succeeded where so many others have failed (see Mourshed et al., 2010; Mourshed, Farrell, & Barton, 2013).

The McKinsey reports (a total of three major reports have been produced on education) provide clear and direct outlines or pathways for policy decision makers on how to maintain or improve the organisational features of educational systems. The summaries and interpretations of the vast research measurement programmes link international measures of school performance to abstractions that serve as structures and functions for educational systems and management involved in change. Key factors are described for improving educational attainment in PISA measures. The reports indicate which management practices policy makers should pay attention to in terms of the macro organisation of educational systems, such as the allocation of financial resources, the level of commitment to change among stakeholders, the need for continuous professional development and the level of teachers’ salaries. The three reports (Barber & Mourshed, 2007; Mourshed et al., 2010, 2013) also pay attention to the characteristics of children’s and family backgrounds that contribute to higher achievement.

Based on analysis of the reports (Barber & Mourshed, 2007; Mourshed et al., 2010, 2013), it can be concluded that McKinsey uses ILSA data to offer clients educational solutions that can help them to improve the educational sector, as well as other
sectors of society. Hence, education is one of the ‘bricks’ that is used in the corporate reasoning of how society functions and develops. Even though McKinsey explicitly says that its educational interest is pro bono, this is not entirely accurate. Rather, the company’s involvement in education is interpreted as a corporate key factor for managing societal change more effectively. Education is consequently used as a corporate technique to achieve national economic and social development. From analysing the reports, it can be concluded that the technique of recognising successful education systems is largely based on the numerical data of student achievements. ILSA is thereby considered as a natural resource for locating deficits and opportunities in policy and regarded as ‘evidence’.

**Entrepreneurial profitable reasons: the Pearson activities**

Pearson is now the largest company operating in edubusiness. Its corporate motto is ‘always learning’. The company currently operates in over 90 countries and has an extensive portfolio of textbooks, testing, test analyses, statistical services, online learning and various software solutions. Pearson is also the most profitable international player in the global edubusiness, with a turnover of over £4 billion in the year 2012. In short, Pearson is one of the most active companies in the very profitable global education market (e.g. Burch, 2009). Pearson as a company is obviously among the first to recognise this new context and has purposefully worked to transform the corporate enterprise from a print-publishing business to a global integrated education company (Hogan, Lingard, & Sellar, 2015).

Pearson has an explicit corporate mission, which is to help people make progress in their lives through learning. Accordingly, Pearson argues that the company has a responsibility ‘to support educational improvement’ and actively share its ‘experience on models that work and those that do not’ (Pearson, 2012, p. 38). However, the mission is twofold, in that one aim is to help people to learn and the other is to make a profit. The basic prerequisite is that Pearson is not accountable to the students, teachers, schools and systems it aims to help, but is instead economically accountable to its shareholders. This duality is a delicate balance when it comes to advertising the company, given that the corporate focus is on the one hand on a specific approach known as corporate social responsibility, and on the other hand on making a profit. In the company, these two foci are discursively intertwined.

In 2013, Pearson released its Efficacy Framework in two separate publications: *Asking More: The Path to Efficacy* (Barber & Rizvi, 2013a) and *The Incomplete Guide to Delivering Learner Outcomes* (Barber & Rizvi, 2013b). Michael Barber is one of the editors of these publications. Before being employed by Pearson, Barber was head of global education at McKinsey and wrote one of the McKinsey reports on education (Barber & Mourseshed, 2007). He is also a well-known education expert and advisor to Tony Blair, the former prime minister of Great Britain. Barber appears to be an important person in the grey-zone activities described and exemplified in this paper. His movements across invisible or visible lines between policy, entrepreneurial policy and GEI can also consequently serve as an example of how intertwined these sectors are, in terms of both personnel and reasoning.

The efficacy strategy that was introduced can be seen as a key technique for a new corporate management, in which various performative mechanisms are employed to create an organisational focus on improvement and effectiveness (Hogan et al., 2015). Efficacy, in the Pearson interpretation, is inspired by how the term is used in the pharmaceutical industry, and is thought to signal a promise that the company’s education products and services are provided with an evidence-based guarantee of improving outcomes.

Another cornerstone in Pearson is *The Learning Curve* (TLC) programme. TLC is a policy report (Pearson, Economist Intelligence Unit, 2012) with an associated website and databank. It was launched in 2012 with a pool of synthesised international comparative performance data and analyses. Michael Barber is a member of the Advisory Board of TLC, as are PISA director Andreas Schleicher and recognised scholars such as Eric Hanushek from Stanford University, with an extensive scientific production within the field of educational economics, and Pamela Sammons from the University of Oxford, who has a major focus on school effectiveness research.

TLC was presented to promote what is called more ‘evidence-informed’ policymaking. The TLC technology is made up of well-established datasets, such as PISA, in an easy-to-read format with explicit policy prescriptions. TLC offers the following five key lessons for education policymakers: there are no magic bullets, respect teachers, culture can be changed, parents are neither impediment to nor saviours of education, and educate for the future, not just the present (Pearson, Economist Intelligence Unit, 2012; cf. Hogan, Sellar, & Lingard, 2016). Despite the range of these statements, TLC is presented as effective in framing policy problems to which Pearson has saleable solutions.

The Efficacy Framework and TLC became even more evident as a corporate strategy for selling solutions in December 2014, when Pearson announced that it had won a competitive tender to develop the
Already from the beginning, the practice of judging educational systems based on numerical data and comparisons of students’ performances in the shape of ILSA has become more common in the Western world in the last two decades. The examples below show that, even without participating in the most globally spread ILSAs, this is also true for ‘outsiders’ and affects both policy and practice.

The agencies presented as performing different regional assessments for what we term appurtenance reasons are LLECA, SACMEQ, the COMFEMEN, Twaweza and Pratham. All these regional bodies spur and develop important regional comparative assessments with the support of international, regional and national experts, and are funded by national and international bodies. They also, to a large extent, cooperate with ILSA personnel and international entrepreneurial policy actors and have close and tight connections with various actors in the GEI.

SACMEQ

SACMEQ was first established at a meeting between Zimbabwe’s Minister for Education and the director of UNESCO’s International Institute of Educational Planning. At the meeting, they agreed on a major research and training project called the Indicators of the Quality of Education Study. The project was undertaken in order to (a) assess the quality of education provided by primary schools, (b) involve the staff of the Ministry’s Planning Unit in integrated research and training activities and (c) provide meaningful advice related to policy concerns expressed by senior decision makers (cf. Pettersson et al., 2016). The initial project resulted in a report written by researchers Kenneth Ross and Neville Postlethwaite (Ross & Postlethwaite, 1991). Ross, and especially Postlethwaite, have both been prominent in IEA; for example, in the late 1980s they cooperated in the Reading Literacy Study (reported in Postlethwaite & Ross, 1992).

Starting in Zimbabwe, the project eventually resulted in the establishment of a wider association, with more countries participating under the acronym of SACMEQ. The organisation took on the challenge to develop cross-national cooperative activity. After the first technical reports were published in 1995, both the organisation and the cross-national project became well established in and was appreciated by the member countries. Already from the beginning, a tight connection with ILSA can be observed. For example, one of the authors of the early technical assessments, John Fallon, Pearson’s chief executive, said: ‘We are developing global benchmarks that, by assessing a wider range of skills, will help young people to prosper in the global economy.’ In addition, Andreas Schleicher, the head of PISA, stated that: ‘PISA 2018 has the potential to be the start of a new phase of our international assessments.’ Conclusively, after winning this bid, Pearson was not only in a position to sell solutions to education, but was also able to produce educational challenges. As such, Pearson became able to take firmer control of the problems that need evidence-based solutions by selling them.

The next section explores a field that is much more scattered, but nevertheless related to a discussion about grey-zone actors – regional comparative knowledge assessments.

Appurtenance reasons: regional comparative knowledge assessments

Inspired by the technology and reasoning embedded in ILSA, assessments have been developed to evaluate students’ knowledge in countries that are not very active in ILSA. Like ILSA, the assessments are based on numerical data for comparing ‘best learning practices’, which we analyse with the aid of websites and publications managed and published via regional assessment agencies in the non-Western world. Another explicit slogan is ‘learning from elsewhere’. The regional assessments that are analysed and discussed in this paper are all performed as comparisons between nations, but are limited to a smaller region, for example by language or other cultural and/or historical factors.

In this section of the paper, the focus is on regional educational assessment activities for low-income countries or countries that contend that specific contextual factors make ILSA unsuitable. The assessments were chosen by means of a search undertaken via Google to identify regional comparative assessments. Some of the assessments resulted from the search and some were referenced in reports published by other agencies. The choice of assessments is probably incomplete, although they can be used as examples to discuss and historicise a specific reasoning. Arguments for creating regional comparative assessments that are inspired by the technology and reasoning used in ILSA are thus taken into account. When it comes to framing a characteristic for the reason for staging regional assessment activities in this specific part of the grey zone, we conclude that this has to do with the duality of the concept of appurtenance. The involved actors establish contact with ILSA technology and personnel and thereby signal a desire for ‘attachment’ to how education is discursively formulated in an educational context inspired by ILSA reasoning. Despite this, inequalities can be observed in how these regionally performed assessments are related to ILSA in terms of which part is dependent on which. It can also be noted that the practice of judging educational systems based on numerical data and comparisons of students’ performances is related to ILSA in terms of which part is dependent on which. It can also be noted that the practice of judging educational systems based on numerical data and comparisons of students’ performances in the shape of ILSA has become more common in the Western world in the last two decades. The examples below show that, even without participating in the most globally spread ILSAs, this is also true for ‘outsiders’ and affects both policy and practice.

The agencies presented as performing different regional assessments for what we term appurtenance reasons are LLECA, SACMEQ, the COMFEMEN, Twaweza and Pratham. All these regional bodies spur and develop important regional comparative assessments with the support of international, regional and national experts, and are funded by national and international bodies. They also, to a large extent, cooperate with ILSA personnel and international entrepreneurial policy actors and have close and tight connections with various actors in the GEI.
reports was Andreas Schleicher, who later became the PISA director (Schleicher & Saito, 1995). The mission was to undertake integrated research and training activities that would develop and expand opportunities for educational planners and researchers, by (a) training people in the technical skills required to monitor, evaluate and compare the general conditions of schooling and the quality of basic education, (b) generating information that could be used by decision-makers to plan the quality of education and (c) utilising innovative information dissemination approaches and a range of policy dialogue activities to ensure that the results were debated, discussed and understood by stakeholders, and then used as the basis for policy and practice. To date, SACMEQ has conducted four major assessments of students’ knowledge and has collected data related to their socioeconomic backgrounds. In recent decades, these have become dominant factors in East African educational policy (Pettersson et al., 2016).

**LLECA**

Latin American countries have only sporadically participated in ILSA, although in two regional tests their desire to participate became much more evident. In 1997, under the auspice of UNESCO’s Regional Bureau of Education for Latin America and the Caribbean, LLECE\(^6\) carried out the First International Comparative Study in Language, Mathematics and Associated Factors in the Third and Fourth Grades of Primary Education (discussed in Hanushek & Woessmann, 2012). The assessment tested the mathematical and literacy performances of representative samples of students in each participating country. In 2006, another test, said to be specially designed for Latin American countries, was launched in the region. It was called the Second Regional Comparative and Explanatory Study. The second study also tested mathematical and literacy performances in combination with the collection of information on the students’ socioeconomic backgrounds. In addition to these knowledge assessments, LLECA produces a range of publications containing comparative numerical data with a view to improving education in Latin America and the Caribbean. The reports are presented in Spanish for the Latin American setting and are based on an organisational strategy to promote dissemination in the participating countries. Due to this strategy, the reports, which are largely managed in a similar way as ILSA, are regarded as important in Latin American educational policy contexts (Pettersson et al., 2016).

**CONFEMEN**

Another regional organisation with a somewhat different history is CONFEMEN, which was founded in 1960 by francophone nations. In 1991, it was stated that there was a need to bring quantitative and qualitative aspects of educational systems together and identify the most effective educational strategies. This resulted in the Programme d’Analyse des Systèmes Educatifs de la CONFEMEN (PASEC). For two decades, PASEC’s mission to evaluate performances has resulted in 35 national assessments in more than 20 countries in Africa and Asia. Since 2012, PASEC has also implemented international comparative assessments, the main objectives being (a) to measure student performance and identify effectiveness and equity factors, (b) to provide national policy indicators for comparisons in space and time, (c) to continue the development of an internal and permanent evaluation and (d) to disseminate international assessment results and contribute to the quality of education. PASEC measures achievements in French (and/or the national language if that is the language of instruction) and mathematics, together with contextual, institutional, social, economic and cultural data.\(^7\)

**Twaweza**

Regional comparative knowledge assessments are also performed by Twaweza, which in Swahili means ‘we can make it happen’.\(^8\) This organisation is located in Uganda, Kenya and Tanzania. The aims of the organisation are to help children to learn, enable citizens to exercise agency and encourage governments to be more responsive in their policymaking. The flagship of the organisation is the comparative knowledge assessment Uwezo, which is Africa’s largest annual learning assessment. What is evident in the Uwezo reports (e.g. Uwezo, 2014) is that children are not learning the basic skills of literacy and numeracy as required by the curriculum. By performing the tests, it is hoped that a ‘best practice’ will be found as to how to increase and improve learning. Uwezo is somewhat innovative, in that the tests are administered by households rather than governmental officials (see below for another example of household surveys). Knowledge tests are also carried out via mobile phone, although these tests are relatively new. In a very short space of time, Twaweza has become an important part of policymaking in the participating countries and highlights some of the most important issues for policymakers to address in the educational sector. The issues are presented as numerical data so that different regions can be compared, thereby making it possible to trace what is considered as ‘best practice’ in the promotion of better-quality education.
Pratham

The last example of a regional comparative knowledge assessment comes from India and the work of Pratham. Although the assessments take place in India as a whole, rather than regionally, they can be discussed as regional in terms of how the structure of Indian education is organised. Since 2005, Pratham has performed the Annual Status of Education Report (ASER), which is India’s largest knowledge assessment and the largest household survey of children to be conducted by citizens’ groups. It is carried out by more than 25,000 volunteers and covers over 700,000 children in 15,000 villages each year. It is also the only annual source of information about the learning levels of children in India today. The household survey methodology has been adapted several times for use in other low-income countries, such as Pakistan (2009), Mali (2011), Senegal (2012) and Mexico (2013). It is also the inspirational methodology for Uwezo (see above). Of interest in this development is the technique of letting households and volunteers administer the tests. To our knowledge, this has not been tested in a Western country on a large scale, but in low-income countries it may very well develop into a regular practice for aggregating data on students’ knowledge. This means of performing knowledge tests can even be discussed in terms of deregulating the governmental power of control to a more society-based way of testing students’ knowledge, contrary to the outsourced control given to, for instance, McKinsey and Pearson in Western countries.

Based on these international examples of regional knowledge assessments, it can be concluded that the techniques and methodology used in ILSA are global, and that numerical data is used on a global scale to present the ‘truth’ and ‘facts’ about education. What is also evident in these reports and assessments is a willingness to use comparisons of what are considered ‘best practices’ in order to achieve better attainment and, through that, develop ‘better-quality’ education.

Different reasons for grey-zone activities

Horkheimer and Adorno (1948) argued that civil society tends to make the incommensurables comparable by reducing them to abstract quantities. This strategy is based on a belief in numbers as more objective (Porter, 1995). Porter illuminated that strict quantification through measurement, counting and calculation is one of the most credible strategies for perceiving objectivity. In education, this strategy can be discussed in relation to reasoning that links political theories of government to notions of democracy and merit, which began to appear in the 19th century on numbers providing narratives about equality and social progress. The emergence of merit tied to individual capabilities and qualities is an invention that replaces manners and gentlemanly conduct as a way of thinking about truth and competency (Sapin, 1994). What became apparent is that differences in the emerging modern society could no longer be legitimised by reference to birth, rank or economic preconditions. In addition, in the 19th century a reasoning evolved of being suspicious of different privileges, and instead meritocracy developed as a safely elitist form of democracy (Porter, 1995), which meant that relationships between the individual and society had to be reviewed. As long as governmental bodies controlled and measured the achievements of students, the legitimacy of the system could be preserved and argued for in a meritocratic system. However, when new actors – which we in this paper call grey-zone actors – started to get involved in both creating and solving policy issues in education, this legitimacy began to be questioned. This development started with ILSA, but now that actors other than governments and researchers have begun to use this data, new societal relations have been created and other reasons for performing assessments and using assessment data have appeared.

Historically, reasoning about meritocratic selection has been justified with references to equality of life chances. This is interpreted as individuals with the same talents and a desire to make use of them having the same chances in life. The only hierarchy that can be accepted is based on meritocratic ideas aggregated from evaluations of individual performance. Consequently, inequality is accepted in terms of who has access to education and social position, but only if this is based on merit. However, just as inequality is based on other prerequisites, so too is the definition of equality. Equality should be staged through merit, although merit also leads to inequality. In other words, meritocracy is both an ideology and a state-sanctioned technology that promotes the elimination of a traditional heritage based inequality, but at the same time legitimises inequalities based on individual performance. This is also a system that is increasingly challenged in terms of who has the power to aggregate and evaluate individual merit. Due to the increasing number of agencies involved in the aggregation and evaluation of merit, these are gradually being transformed to signal the values of entire nations. Nationally based merits, such as those presented in, for instance, PISA and TIMSS, are now important currencies in the global economic market. Consequently, as nations strive to succeed in these rankings, they need to consider how to increase the currency. When student performances become the currency for a greater involvement in the global economy, agencies have more opportunities to sell their solutions. In addition, low-income countries
are more able to use ILSA techniques and methodologies to signal greater involvement and participation in the global economy. As such, ILSA enables actors to use ILSA reasoning to sell recommendations for or solutions to perceived educational policy issues, create new educational challenges and advise how to measure and assess educational knowledge. The consequence of this is that agencies described as being involved in the grey zone are given an opportunity to participate in both the creation and solution of educational issues.

Overall, this paper focuses on two central knowledge problematics: the relation between education and a specific technology, framed by ideologies on meritocracy and understood as a selection of different and hierarchical positions in society by means of education performances, and the development and expansion of various assessments and their increasing use in educational practice, policy and bureaucracy. It also highlights that the comparison and use of data in education has evolved into a specific technology for numerically framing education. Educational numbers were as such transformed from representations of education into education per se. This may have been influenced by the societal and historical connections to reasoning about meritocracy, which was considered as central to the development of the state and society. Porter (1995) argued that the reason why numbers became so central in the development of society is that they are perceived as ‘objective’ and ‘neutral’. However, this is both inaccurate and contradictory. Rather, numbers should be perceived as a technology for steering and managing society and the state that is based on connotations of ‘objectivity’, distance and neutrality. Consequently, grey-zone activities are involved in creating a ‘neutral’ vision of education and a ‘neutral’ vision of what education should be like, which is in itself reason enough to further investigate the increasing number of grey-zone activities involved in forming today’s educational policy.

In the paper, attention has been drawn to three different reasons for using ILSA data or being inspired by ILSA when performing various activities. The first is an entrepreneurial policy reason, which is explained by the example of McKinsey. McKinsey is primarily involved with ILSA data because education is considered to be one of the ‘bricks’ when it comes to building better opportunities for development in nations. At McKinsey, this ‘brick’ is constructed using ILSA data. McKinsey’s focus on ILSA data leads to nations working with the recommendations and solutions provided by McKinsey, which in turn strengthens their dependency on ILSA and the explicit focus on student achievement as a good measure of the national quality of education. The second is an entrepreneurial profitable reason, exemplified by Pearson. Pearson has strengthened its connections to ILSA and the reasoning is apparent in its activities because it allows the company to both solve and create educational problems. This is done under the flag of social responsibility – a responsibility that is also connected to profit. The third and final reason is the appurtenance reason, which is exemplified by examples of organisations performing assessments and evaluations in low-income countries. What is evident in these examples is that low-income countries try to make connections to, and participate in, a globalised economic market. By connecting to ILSA techniques and methodologies, an ‘attachment’ is created to the importance of improving students’ knowledge and participation in the global market, economically as well as educationally.

As indicated in the introduction, the examples of grey-zone activities provided here are not exhaustive. There will no doubt also be other reasons for agencies to tighten their connections to ILSA than those highlighted here. This paper should thus be considered an historicised effort to show that a number of activities are connected to ILSA – outside governmental and non-governmental organisations and research – and are involved in spreading ILSA reasoning. We also acknowledge that further studies need to be conducted in this particular field of research.

Notes

1. Twaweza is a consortium staging large-scale assessments in Uganda, Kenya and Tanzania called Uwezo.
2. Pratham is India’s largest non-governmental organisation working for quality education for underprivileged children.
3. http://www.mckinsey.com/about_us.
4. All information about Pearson has been retrieved from https://www.pearson.com/.
5. Information about SACMEQ has been retrieved from www.sacmeq.org.
6. Most of the information about LLECA has been retrieved from www.unesco.org.
7. The information about CONFEM has been retrieved from http://www.confemen.org.
8. The information about Twaweza and Uwezo has in part been retrieved from http://www.uwezo.net.
9. The information about Pratham has been retrieved from http://www.pratham.org.
10. http://www.asercentre.org is the official website for the ASER assessment.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Barber, M., & Mourshed, M. (2007). How the world’s best-performing school systems come out on top. Chicago, IL: McKinsey & Company.
Barber, M., & Rizvi, S. (2013a). Asking more: The path to efficacy. New York, NY: Pearson.
Barber, M., & Rizvi, S. (2013b). The incomplete guide to delivering learner outcomes. New York, NY: Pearson.
Becher, T., & Trowler, P. R. (2001). Academic tribes and territories. Buckingham: Open University Press.
Bloom, B. (1969). Cross-national study of educational attainment: Stage I of the IEA investigation in six subject areas. Final report. Volume I. Chicago, IL: University of Chicago.
Bourdieu, P. (1971). Systems of education and systems of thought. In M. F. D. Young (Ed.), Knowledge and control: New directions in the sociology of education. London: Collier-Macmillan.
Burch, P. (2009). Hidden markets: The new education privatization. Hoboken: Routledge.
Carvalho, L. M. (2012). The fabrications and travels of a knowledge-policy instrument. European Educational Research Journal, 11(2), 172–188. doi:10.2304/eerj.2012.11.2.172
Crane, D. (1972). Invisible college: Diffusion of knowledge in scientific communities. Chicago, IL: University of Chicago Press.
Desrosières, A. (1991). How to make things which hold together: Social science, statistics and the state. Dordrecht: Discourses on society. Springer Netherlands.
Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., & Walker, D. A. (1962). Educational achievements of 13 year olds in twelve countries. Hamburg: UNESCO Institute of Education.
Grek, S. (2009). Governing by numbers: The PISA ‘effect’ in Europe. Journal of Education Policy, 24(1), 23–37. doi:10.1080/02680930902412669
Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge: Cambridge University Press.
Hacking, I. (1992). “Style” for historians and philosophers. Studies in History and Philosophy of Science Part A, 23 (1), 1–20. doi:10.1016/0039-3681(92)90024-Z
Hanusch, E. A., & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. Journal of Development Economics, 99 (2), 497–512. doi:10.1016/j.jdeveco.2012.06.004
Hogan, A., Lingard, B., & Sellar, S. (2015). Edu-business and education policy: The case of Pearson. Professional Voice, 10(2), 24–30.
Hogan, A., Sellar, S., & Lingard, B. (2016). Commercialising comparison: Pearson puts the TLC in soft capitalism. Journal of Education Policy, 31(3), 243–258. doi:10.1080/02680939.2015.1112922
Horkheimer, M., & Adorno, T. W. (1948). Dialektik der Aufklärung: Philosophische Fragmente. Frankfurt: S. Fischer Verlag.
Igo, S. E. (2007). The averaged American: Surveys, citizens, and the making of mass public. Cambridge, MA: Harvard University Press.
Jasanoff, S. (2004). States of knowledge: The co-production of science and social order. Princeton, NJ: Harvard University.
Kazamias, A. M. (2009). Forgotten men, forgotten themes: The historical-philosophical-cultural and liberal humanist motif in comparative education. In R. Cowen & A. M. Kazamias (Eds.), Second international handbook of comparative education (pp. 37–58). Dordrecht: Springer.
Knorr Cetina, K., & Cicourel, A. V. (Eds.). (1981). Advances in social theory and methodology. Toward an integration of micro- and macro-sociologies. London: Routledge & Kegan Paul.
Lagemann, E. C. (2000). An elusive science. The troubling history of education research. Chicago, IL: University of Chicago Press.
Latour, B. (1988). The pasteurization of France. Cambridge. Harvard University Press.
Latour, B., & Woolgar, S. (1979). Laboratory life: The construction of scientific facts. Princeton, IL: Princeton University Press.
Lemm, N. (1999). The big test: The secret history of the American meritocracy. New York, NY: Farrar, Straus & Giroux.
Lindblad, S., Pettersson, D., & Popkewitz, T. (2015). International comparisons of school results: A systematic review of research on large scale assessments in education. Stockholm: Vetenskapsrådets rapporter.
Lussi Borer, V., & Lawn, M. (2013). Governing education systems by shaping data: From the past to the present, from national to international perspectives. European Educational Research Journal, 12(1), 48–52. doi:10.2304/eerj.2013.12.1.48
Martens, K. (2007). How to become an influential actor – The “comparative turn” in OECD education policy. In K. Martens, A. Rusconi, & K. Lutz (Eds.), Transformations of the state and global governance. London: Routledge.
McCall, W. (1922). How to measure in education? New York, NY: The Macmillan Company.
Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. Chicago, IL: University of Chicago Press.
Mourseshed, M., Chijioke, C., & Barber, M. (2010). How the world’s most improved school systems keep getting better. Chicago, IL: McKinsey & Company.
Mourseshed, M., Farrell, D., & Barton, D. (2013). Education to employment: Designing a system that works. Chicago, IL: McKinsey & Company.
Növoa, A., & Lord, M. (Eds.). (2002). Fabricating Europe: The formation of an education space. Dordrecht: Kluwer Academic.
Owens, T. L. (2013). Thinking beyond league tables: A review of key PISA research questions. In H.-D. Meyer & A. Benavot (Eds.), PISA, power and policy: The emergence of global educational governance. Oxford Studies in Comparative Education; 23:1. Oxford: Symposium Books.
Pearson. (2012). Annual report and accounts. Retrieved from http://www.pearson.com/content/dam/pearsoncorporate/files/cosec/2013/15939_PearsonAR12.pdf
Pearson, Economist Intelligence Unit. (2012). The learning curve: Lessons in country performance in education. New York, NY: Pearson.
Pettersson, D. (2008). Internationell kunskapsbedömning som inslag i nationell styrning av skolan (Acta Universitatis Upsaliensis. Uppsala Studies in Education No 120). Uppsala: Uppsala universitet.
Pettersson, D. (2014a). The development of the IEA: The rise of large-scale testing. In A. Nordin & D. Sundberg (Eds.), Transnational policy-flows in European education: Conceptualizing and governing knowledge. Oxford Studies in Comparative Education. East Greenwich: Symposium Books.
Pettersson, D. (2014b). Three narratives: National interpretations of PISA. Knowledge Cultures, 2(4), 25–45.
Pettersson, D., Popkewitz, T. S., & Lindblad, S. (2016). On the use of educational numbers: Comparative constructions of hierarchies by means of large-scale assessments.
Espacio, Tiempo y Educación, 3(1), 177–202. doi:10.14516/ete.2016.003.001.10
Poovey, M. (1998). A history of the modern fact: Problems of knowledge in the sciences of wealth and society. Chicago, IL: University of Chicago Press.
Popkewitz, T. S. (2015). Curriculum studies, the reason of “reason”, and schooling. In T. S. Popkewitz (Ed.), The “reason” of schooling: Historicizing curriculum studies, pedagogy, and teacher education. New York, NY: Routledge.
Porter, T. M. (1995). Trust in numbers: The pursuit of objectivity in science and public life. Princeton, NJ: Princeton University Press.
Porter, T. M. (2012). Funny numbers. Culture Unbound, 4, 585–598. doi:10.3384/cu.2000.1525.124585
Postlethwaite, N., & Ross, K. N. (1992). Effective schools in reading: Implications for educational planners: An exploratory study. The Hague: IEA.
Purves, A. C. (1987). The evolution of the IEA: A memoir. Comparative Education Review, 31(1), 10–28. doi:10.1086/446653
Rose, N. (1999). Powers of freedom: Reframing political thought. Cambridge: Cambridge University Press.
Ross, K. N., & Postlethwaite, T. N. (1991). Indicators of the quality of education: A national study of primary schools in Zimbabwe. Paris: UNESCO.
Sapin, S. (1994). A social history of truth: Civility and science in seventeenth-century England. Chicago, IL: University of Chicago Press.
Schleicher, A., & Saito, M. (1995). Data building and data management. Paris: IIEP.
Shapin, S., & Schaffer, S. (1985). Leviathan and the air-pump: Hobbes, Boyle and the experimental life. Princeton, NJ: Princeton University Press.
Smeyers, P., & Depaepe, M. (2010). Representation or hard evidence? The use of statistics in education and educational research. In P. Smeyers & M. Depaepe (Eds.), Educational research: The ethics and aesthetics of statistics. Educational Research 5. New York, NY: Springer.
Stigler, S. M. (1992). A historical view of statistical concepts in psychology and educational research. American Journal of Education, 101, 60–70. doi:10.1086/444032
Supreme Court of Minnesota. (1919). Cohen vs. Whitcomb 1919:142, Minnesota 20 (the citation is retrieved August 17, 2016, from a description of appurtenance from the website: http://www.duhaime.org/LegalDictionary/A/Appurtenance.aspx)
Uwezo. (2014). Are our children learning? Literacy and numeracy across East Africa 2013. Nairobi: Author.
Verger, A., Lubienski, C., & Steiner-Khamsi, G. (2016). The emergence and structuring of the global education industry: Towards an analytical framework. In A. Verger, C. Lubienski, & G. Steiner-Khamsi (Eds.), World yearbook of education 2016: The global education industry. New York, NY: Routledge.