Cómo referenciar este artículo / How to reference this article

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: http://dx.doi.org/10.14516/ete.2016.003.001.10

On the Use of Educational Numbers: Comparative Constructions of Hierarchies by Means of Large-Scale Assessments

Sobre el uso de los números de la Educación: construcciones comparativas de jerarquías a través de evaluaciones a gran escala

Daniel Pettersson
e-mail: Daniel.Pettersson@hig.se
University of Gävle. Sweden
Thomas S. Popkewitz
e-mail: popkewitz@education.wisc.edu
University of Wisconsin-Madison. United States
Sverker Lindblad
e-mail: sverker.lindblad@ped.gu.se
University of Gothenburg. Sweden

Abstract: Our text is focusing on two central knowledge problematics. First, the relation between education and a specific technology, framed by ideologies on modernity and meritocracy, understood as a selection to different and hierarchical positions in society by means of education performances. Second, the development and expansion of national, regional and international assessments and the increasing use of them within educational practice, policy and bureaucracy is acknowledged. In doing so we note that an historical tradition within education to compare and use data evolved into a specific technology for framing education with a centrality of numbers. Educational numbers came as such to be transformed from representations of education into education per se. This could happen due to societal historical connections to reasoning about modernity and meritocracy, which were considered as central in the development of the state and society. Porter (1995) is making an argument about that the reason that numbers came to be central in the development of society had to do with that numbers are perceived as «objective» and as such «neutral», but in reality this is in many respect false, and even contradictory. Instead, numbers should be perceived as a technology of steering and managing society and the state, a technology based on connotations of «objectivity», but also as a technology of distance and neutrality. What we are making an argument about is connected to Porters statements. We state that comparisons and the use of numbered data for describing education dependent on parallel societal processes, in science, society and state, came to be transformed into that numbered data on education came to be perceived as education per se. This development can be described in several aspects, but we are primarily describing it through emphasizing some historical comparative and data aggregative collaborations within science and governmental organizations and later the growing importance of transnational agencies and international, regional and national assessments.

Keywords: Education by the numbers; meritocracy; assessments; educational comparisons; modernity.

Resumen: Nuestro texto se centra en dos problemáticas centrales de conocimiento. En primer lugar, la relación entre la educación y una tecnología específica, enmarcado por las ideologías de la modernidad y la meritocracia, entendida como una selección de diferentes y jerárquicas posiciones en la sociedad por medio de actuaciones de educación. En segundo lugar, se reconoce el desarrollo y expansión de las evaluaciones nacionales, regionales e internacionales y el creciente uso de ellas dentro de la práctica, la política educativa y la burocracia. De este modo observamos que la tradición histórica en la educación de comparar y utilizar datos ha evolucionado hasta convertirse en una tecnología específica para la elaboración de la educación con una centralidad de números. Los números educativos vinieron como tal para ser transformados en representaciones de la educación y en la educación per se. Esto podría suceder debido a las conexiones históricas de la sociedad a razonar sobre la modernidad y la meritocracia, que fueron consideradas como un elemento central en el desarrollo del Estado y la sociedad. Porter (1995) plantea un argumento acerca de que la razón por la cual los números llegaron a ocupar un lugar central en el desarrollo de la sociedad tendría que ver con que los números son percibidos como «objetivo» y como algo «neutral», pero en realidad esto es, en muchos casos, una suposición falsa, e incluso contradictoria. En cambio, los números deben ser percibidos como una tecnología de dirección y gestión de la sociedad y del Estado, una tecnología basada en connotaciones de «objetividad», y también como una tecnología de la distancia y neutralidad. Afirmamos que las comparaciones y el uso de datos numerados para describir la educación depende de los procesos sociales paralelos, en la ciencia, la sociedad y el Estado, llegaron a transformarse en que los datos enumerados en la educación llegó a ser percibido como la educación per se. Este desarrollo puede ser descrito a partir de varios aspectos, pero nosotros estamos describiendo principalmente a través del énfasis sobre algunas comparaciones históricas y colaboraciones de agregación de datos dentro de la ciencia y de las organizaciones no gubernamentales y más tarde la importancia creciente de las agencias transnacionales y las evaluaciones internacionales, regionales y nacionales.

Palabras clave: Educación a través de los números; meritocracia; evaluaciones; comparaciones educativas; modernidad.

Recibido / Received: 11/01/2016

Aceptado / Accepted: 17/01/2016

1. Introduction

Our interest is on historically how numbers become a way to tell the truth about schooling, teachers, pupils and their relation to society. Numbers, Hacking (1983) argues are parts of systems of communication whose technologies create distances from phenomena by appearing to summarize complex events and transactions. The seemingly rigor and uniformity appear transported across time and space so as to not require intimate knowledge and personal trust. Numbers also appear to exclude judgment as the perennial struggles of sciences and policies struggle against subjectivity. Its mechanical objectivity implies personal restraint through following rules that project fairness and impartiality.

But there is what Ian Hacking considers that historical qualities' of numbers as the product of human activity that «acts» on the activities it has produced (Hacking, 1983, p. 123). The seemingly technical appearances of the numbers enter into cultural realms that are never merely numbers but codifications and standardization of what are to constitute reality and planning. Circulation of ideas and mass schooling in the 19th and 20th century led to an interest about other countries' educational processes. International missions, organization of exhibitions and the production of international encyclopedias (*cf.* Nóvoa, Yariv-Mashal, 2003) were all products of this interest. In parallel numbered data on education evolved for governing reasons (e.g. Lawn, 2013). Data on education were collected through measurements – first developed for internal usage –

eventually widened to include comparisons within and later on between nations (Landahl, Lundahl, 2013). The different projects inscribe rules and standards by which experiences are classified, problems located, and procedures given to order what is seen, thought about, and acted on. Numbers operate as road maps and «highways» to the desired future that is simultaneously about who is not desired – not only of school but of the family and society.

This text primarily discusses two phenomena – educational comparisons and data usage – and how they developed in parallel and inspired each other into a specific technology described as a technology of distance (Porter, 1995). As a first stepping stone in our discussion we claim that comparisons and data usage is also a specific style of scientific reasoning (Hacking, 1992a) that becomes inscribed in social planning in that numbers are never merely numbers. They also embody cultural theses about differences – creating for instance perceived differences in kinds of people. We start this discussion in a position of asking why the use of different statistics, tests, examinations and surveys in education has become so self-evident and why the practice of conducting them has become so naturalized. With this in focus we discuss two central knowledge problematics. First, the relation between education and a specific technology, framed by ideologies on modernity and meritocracy, understood as a selection to different and hierarchical positions in society by means of education performances. From this point of view education can be given a meritocratic meaning (*cf.* e.g. Bourdieu, 1971) that numbers become a way to assess and measure education performance. Second, the development and expansion of national, regional and international assessments and the increasing use of them within educational practice, policy and bureaucracy (*cf.* Forsberg, Pettersson, 2015).

2. Legitimating educational selections in Modernity

Horkheimer and Adorno (1944) are making an argument of that civil society tends to make the incommensurables comparable by reducing them into abstract quantities. This strategy is to a large extent based in a belief in numbers as more objective (Porter, 1995). Porter illuminates that strict quantification through measurement, counting and calculation is among the most credible strategies for perceiving objectivity. The strategy has enjoyed a widespread and growing authority for about at least two centuries in e.g. science and the organizing of the state. In education this strategy can be discussed in relation to reasoning that links political theories of government with notions of democracy and merit, which begin to appear in the 19th century on numbers that providing narratives about equality and social progress. In the following we will use this discussion for highlighting some historical trajectories of contemporary phenomenon.

The emergence of the merit tied to individual capabilities and qualities is an invention that replaces the manners and gentlemanly conduct as a way of thinking about truth and competency (Sapin, 1994). Reasoning about merit, however, is not unique to the modern societies although it was embedded in different systems of reason that did not have notions of individuality, agency and the temporality of progress. Historically and prior to the Enlightenments, for example, societies had trade-offs made between merits, seniority, heritage and divinity's given orders in organizing the social order (Neves, 2000). What the Enlightenment brought into view was a notion of modernity that gave individuals their own history and the capacity of development that enables the idea of merit. The French philosophers prior to its revolution spoke about needing a equal system of measurement if there was going to be an equal society (see, Popkewitz, 2008; also see Kett, 2013). In this argument one can begin to see the development of the ideas that numbers have autonomy from human activities yet to be applied in a manner that «acts» in social arenas as the procedures for correcting social wrongs and enabling human equality in organizing society.

In a critique of how liberal society constructs inequality through ideas of merit, the British sociologist Michael Young and his book *The Rise of the Meritocracy* from 1958 is a well-known example. The concept was although already used a few years before in an article by Alan Fox (1956) that gives merit a functionalism through discussing institutions and ideologies as reproducing and legitimating social stratification (Littler, 2013). In Fox's interpretation meritocracy is a societal concept in which the talented, the energetic and the ambitious are the favored not only due to their talents but also because of the interconnectedness between education, merits and social benefits.

What became apparent is that in the emerging modern society differences could no longer be legitimized with references to birth, rank or economic preconditions. What was also evident was that in the 19th century a reasoning evolved being suspicious of privilege and meritocracy was introduced as a safely elitist form of democracy (Porter, 1995). Hence, relationships between the individual and society had to be rewritten.

In modernity reasoning on meritocratic selection is normally justified with references to equality of life-chances. This is often interpreted in terms of individual's having the same talents and desire to make use of them should have the same chances in life. The only hierarchy that can be accepted is based on meritocratic ideas aggregated from evaluations of the performance of individuals. Consequently inequality is then accepted on who gets access to education and social positions, but only if this is based on merits. Meritocracy is as such not numb on inequality, but inequality is based on other prerequisites and so is also the defini-

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

tion of equality. Equality is to be staged through merits, but merits also lead to inequality. In other words, meritocracy is an ideology, but also a state sanctioned technology, that promotes elimination of a traditional heritage based inequality but at the same time it legitimizes inequalities based on individual performances. In fact Lemann (1999) makes a critique out of this stating that the American meritocracy in fact is a lie – socio-economic background and ethnicity is still the most dominant predictors, ahead of performances, when to foresee the future of individuals. The observation of Lemann is probably universal and not only applicable for just the American context, which can be seen in discussions on meritocracy. Meritocracy is as such in many ways a both problematic and complex ideology. Already Michael Young and many after him have pointed out problems associated with meritocracy – e.g. considering social and cultural heritage in terms of access to merits (Bourdieu, 1971). These descriptions are often tied to the many unspoken assumptions and styles of reasoning (Hacking, 1992a) that meritocracy rests on, such as the conceptualization of talent/intelligence, the ability to discern what is essential knowledge, skills and abilities, whether these are measurable - preferable by means of standardized tests – and if they can be made comparable[1]. Others have criticized the meritocratic technology in that it is not able to maintain the meritocratic ideal, that new hierarchies are established and that certain groups are systematically disadvantaged and discriminated (Bell, 1972). Thus, we here note a combination of meritocratic reasoning related to categorizations or taxonomies of individuals or groups in criticisms of educational systems at work.

Consequently, the meritocratic technology affects and regulates, almost as a gatekeeper, the entrance, the passages and the outputs in education and labor market (*cf.* Forsberg, 2006). Within the education system this can be highlighted through administrative and pedagogical systems designed for assessment, evaluation, documentation and comparisons of student achievement. In other words, meritocracy as a technology is designed as a combination of equality and competitive ideals. In this perspective, meritocracy as a just injustice or a just inequality can be highlighted (*cf.* Forsberg, Pettersson, 2015).

Today, the use of numbers and statistic comparisons are taken for granted as a way of understanding how society grow and schools respond to the social and political commitments associated with equality as expressed through ideologies of merits. Data from grades, exams, student performance on national tests and regional and international knowledge assessments are aggregated and are now widely used for establishing national results and to make comparisons

[1] «[…] most common definition of meritocracy conceptualizes merit in terms tested competency and ability, and most likely as measured by IQ or standardized achievement test» (Levinson, Cookson, Sadovnik, 2001 p. 436).

between them, establishing a trust in numbers (Porter, 1995) affecting the reasoning and discussions on education. A way to present these results is through school or country rankings which in turn often lead to discussions about reforms for achieving better performances in the rankings.

In the following sections we are emphasizing on the institutionalization of a meritocratic reasoning within a field of comparative educational assessments, which we consider as a specific technology for organizing education. But before that we have to describe how comparisons and data usage came into the field of education establishing this specific technology.

3. Educational techniques for establishing a trust in numbers

For understanding much of the development of educational research during the 20th century Lagemann (2000), from an American context, claims that «One cannot understand the history of education […] unless one realizes that Edward L. Thorndike won and John Dewey lost» (a.a. p. xi). Apart from agreeing or not with this claim one has to admit that research using quantitative, e.g. statistical techniques, gained in attention in the coming years. There is a lot of different, but often, interrelated factors responsible for this development – e.g. an acceptance of positivism as a dominant scientific reasoning, a fast institutional growth of educational institutions but also educational research with an interest in comparisons and data usage, socialization of upcoming researchers influenced by the use of statistics within psychology (Stigler, 1992; Hacking, 1992b), a supremacy of meritocratic values in modern societies and the constant need to legitimate these by «objective» and «neutral» research (Smyers, Depaepe, 2010; Porter, 1995).

Thorndike embraced the method of testing and the use of statistics was central, so was also the belief that everything can be measured. This was at the time expressed in terms like: «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 of education and a part of the broader educational context. The power and the efficiency connotation to statistics gave rise to a faith in measurement and metrics (Smyers, Depaepe, 2010; Porter, 1995). The growth of scientific statistics as a dominating reasoning creates beliefs in that the more data we gather and the more comparisons we make – the more will we know. This use of comparisons and data within statistics carries a number of presuppositions: that reality can be represented in numbers, that it can be controlled and that risks can be managed.

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

When educationalists embraced the new method of testing and statistics this was on a societal level nothing new – discussions and institutional manifestations based on the predecessors to statistics – comparisons and numbered data – was rather old for e.g. counting people and holdings in administrative purposes for military service and tax surveys. For western societies this can be dated back to at least the days of William the Conqueror and the administration of Domesday Book of 1086 (Igo, 2007)[2]. A tradition of societal comparisons developed historically in at least three different lines: i) comparisons between the contemporary and the history, ii) comparisons between countries, and iii) comparisons for implementing, what today is discussed in terms of, *best practices* (Pettersson, 2014).

During the 19th century a desire for internal educational reforms led to a quest for information and data concerning education in other countries. A variety of literature developed either written by travelers with an interest in education or by government officials collecting data for governmental reports. However, the belief in truth-telling capacity of numbers, in order to establish values about social and personal life, has not been the case before the 19th century. Prior, truth was established through the manners and rhetorical qualities combined with social status of the speaker (Poovey, 1998). In times of social turmoil moral, progressive and scientific campaigns appeared to ameliorate bad social conditions. In particular, this is envisioned within social science, which during this period took an *empirical turn* in order to understand general social processes. By turning into a more empirical approach social science could seemingly distance itself from the moral and progressive value laden, social activist roots and by that give itself an appearance of neutrality. Consequently, the empirical turn created an opportunity for social science to act as perceived neutral in a manner before more connected to natural science.

The empirical turn lead to the emergence of the new scientific branch of statistics. Statistics create *facts* about social life and became part of the societal change that traversed different sectors like economy, statecraft and culture (Poovey, 1998). Statistical comparisons also create recognition of differences between nomenclatures as a problem that must be eliminated. In doing so a grid can be constructed that appears to be valid and unresponsive of national contexts or time. Hence, information about contemporary taxonomies is preserved instead of dissolved. This view, discussed by Desrosières (1991), also marks a clear rupture with the more classical ways of social science where numbers were used to describe things that exist independently of the conventions establishing them. With the entrance of a new scientific and political way to use numbers in terms

[2] The Domesday Book is a survey written in Medieval Latin. Today the manuscript is held at the National Archives at Kew, London. In 2011 the Open Domesday site (http://www.domesdaymap.co.uk/) made the manuscript available online.

of statistical descriptions it becomes evident that coding creates equivalence. The act of coding came to construct equivalence classes between diverse objects, and the class more than the individual objects came to be judged and described. In that way objects, through the process of constructing equivalence classes, were made comparable. The effect was that the individual was lost in favor of overall descriptions and numbers were then used for describing overall emphasis more than an emphasis on the individual (Igo, 2007; Lemann, 1999). However, the coding to find equivalence also provides a tight link between political and cognitive dimensions. One of these tighter links can be observed in the 19th century when equity and equivalence appear as prescriptive and descriptive aspects of disciplines. Initially, comparability and equivalence between objects were less a question of knowledge and more of justice regulated in e.g. laws governing market exchange and as such more focused on quality aspects. Quantification through numbers emerged for separating the act of political management of people from the scientific management of things. In the evolution of this separation, transcendence is taking form concerning the contingency of particular cases and circumstances to things that hold together and displays qualities of generality and permanence. The development raises a lot of question and one of them is if these objects really are equivalent, but maybe a more appropriate standing is who decides to treat these objects as equivalent, and why? As such numbers can be seen as a technology of distance used as a claim of objectivity instantiated by moral and political discourses (Porter, 1995). All kinds of quantified knowledge are in that respect artificial through creating uniformity among different qualities of things, uniformity that gives social authority to the interrelation of science and policy.

Numbers became visualized as social facts whose objectivity was important in the making of citizens in the 19th century. Consequently, numbers are thought as a social technology that instantiate consensus and harmony in the world. The uniformity given by numbers brings as such order in social life by regulating relations (Rose, 1999). However, while the things of numbers «act» as real, they 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 for understanding qualities of governing we first need to consider numbers as defining a problematized space where subjects and objects are stabilized. Numbers seem technical, objective and calculable and embodying the idea of giving all equal chances and representation. Numbers standardized the subject of measurement and assessment but also the act of exchange so that they were no longer seen as dependent on the personalities or the statuses of those who performed the measurements or assessments. The faith in numbers in social affairs today is so markedly part of common sense that it is possible to

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

talk about «transparency» in governmental social affairs and even personal relations can be discussed through statistical charts and graphs. Making government «transparent» have become an act of democratic modes of acting for ensuring that everybody «knows» how decisions are made. In that sense, numbers have become part of a discourse about guaranteeing democracy.

4. Trajectories of educational comparisons and data usage

For understanding why educationalists came to take an empirical turn we have to discuss some trajectories visible within the history of education and educational research. We have above discussed statistics and numbers as important issues for understanding the appearance of a specific reasoning and a specific technology. What we discuss below is that the field of education was ready to embrace the empirical turn, statistics and the importance of numbers due to that there for quite a long time was an internal discussion about the importance of comparisons and aggregation of numbered data for developing education and educational performance.

The French scientist Marc-Antoine Jullien was one of the first to construct a methodology for comparisons using numbers that: i) separated the empirical field of observation into its constituent parts, ii) devised techniques of inquiry, and iii) used formal models of analysis in an attempt to explain how schools functioned (Gautherin, 1993). In 1817, Jullien tried to compare educational establishments throughout Europe by setting up a *Special Commission on Education* and an *Educational Institute* publishing *Educational Newsletters* (Brickman, 2010). As part of the project researchers were sent to different countries to investigate educational systems.

Apart from Jullien's ideas about a search for general principles and the creation of international agencies working with numbered comparisons, the nineteenth century was dominated by men appointed by their governments to develop internal education (Holmes, 1981). Another practice evolved simultaneously and parallel to governmental interest in numbering education into data which fostered a specific discourse and as such created a specific reasoning which can be recognized in different ways but one is an increasing interest in arranging international exhibitions and scientific congresses. This was possible due to the shift within science as a cultural practice from the Enlightenments notions of reason and science into reasoning more in terms of science as reason itself. Science as such came to be more organized as a domain in which a specific scientific reasoning could be consolidated. One example of this consolidation is the first *International Statistical Congress* which was held in Brussels in 1853 in which education

was recognized as an interesting area for statistical comparisons. However, many of the studies conducted in the late 19th century were national in nature although some did focus on international comparisons (Smyth, 2008).

In parallel to the scientific development the growing authority of data in governing education systems was mostly influenced by the use of data in the US. The *Department of Education* was created in 1867 and later reorganized as the *Bureau of Education*. The agency's main task came to be to collect and disseminate educational statistics although the data was also seen as a catalyst in the improvement of education (Lawn, 2013). During this period the most famous reports were those written by the Secretary of the Massachusetts Board of Education, Horace Mann. He had in 1838 started to publish reports about American schools, and in 1844 he published a report (Mann, 1844) in which the Massachusetts school system and schools in Great Britain, France, the German states and Holland were compared. Mann's report served as a model for such comparisons and had a major effect on other reports of different educational systems.

When an exhibition was held in Paris in 1878, the US was able to produce a variety of reports using educational data, which in turn influenced the future of comparisons (Lawn, 2013). Especially important was the numerical data visualized as graphs and diagrams. Graphs and diagrams could store large amounts of data and make explicit claims on data relations that were easy for people to grasp. These images were able to «speak for themselves», as Lawn (2013) puts it. The statistical data that was displayed in graphs and diagrams was also standardized, thus creating a new language and interpretation of society (Lawn, 2013). The practice of using statistics and presenting educational figures in graphs and diagrams spread rapidly and became part of the educational language. In Porter's (1995) words, using the «language of quantity» has the advantage of being «the technology of distance» (a.a. p. ix) and a decision based on numbers has at least the appearance of being fair and impersonal. Taking a big step in history, in the 1990s we could note another shift in emphasis in international discourses on education – from statements on input and enrolment such as presented in UNESCO *World Education Reports* to statements on outputs and learning and achievement gaps as presented by the OECD reports on *Education at a Glance* (see e.g. Lindblad & Popkewitz, 2001).

5. Spaces for educational techniques: compare-collaborate-learn

Going back to a more temporal order of things describing the comparative development – the interest in collecting educational numbered data in the late 19th century increased and led to systematic comparisons becoming relatively

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

common. Examples can be found in governmental organizations such as the U.S. *Bureau of Education* (1867), the *Musée Pédagogique* in France (1879), the *Office of Special Inquiries and Reports* in London (1895), and the *Zentralinstitut für Erziehung und Unterricht* in Berlin (1915) all collecting numbered data in order to develop education (Brickman, 1966).

In 1900, the comparative educationalist Michael Sadler delivered a speech called *How Far Can We Learn Anything of Practical Value from the Study of Foreign Systems of Education?* This signaled a new era of comparing educational systems – a period in which researchers like Isaac Leon Kandel, Friedrich Schneider, Nicholas Hans and others laid the foundations and built structures for the scientific field of comparative education. In the first decades of the 20th century monographs, yearbooks, statistical compilations, conferences, instructions, and discussions about practical applications of comparative education started to multiply. Even though statistical data was evident during this period the foundation of comparative education was mainly rooted within humanities, but eventually the scientific community started to make arguments for stronger educational planning and the development of means for reliable predictions (Noah & Eckstein, 1969). Hence, independent organizations were developed in order to compare education provided by different nations using available data, such as *Institute of International Education* in New York (1919), the *International Institute of Teachers College*, Columbia University (1923), the *Bureau of International Education* (IBE) in Geneva (1925) and the *Institut International de Coopération Intellectuelle* in Paris (1925) (Brickman, 1966). Connected to these institutes were also a variety of scientific journals, such as the *Educational Yearbook* from Teachers College, Columbia University, edited by Isaac Leon Kandel between the years 1925 and 1944 having a major impact in the areas of comparative and international education; the *Year Book of Education* (1932-1940) published by the University of London; the *Annuaire International de l'Education et de l'Enseignement* issued by the *International Bureau of Education* between the years 1933 and 1939 (Brickman, 1966). Of particular importance was also Kandel's seminal book *Comparative Education* (Kandel, 1933).

After Second World War, comparative education developed in somewhat different trajectories. Science at one level came to develop into what can be called «big science» more organized and funded by states and new collaborative international agencies. This development required large groups of scientists and thus produced a new kind of science that did not exist on this scale prior to the war. What became evident was that scientific facts produced within this environment further came to emphasize on comparisons and numbered data was collected on a more world-wide basis, creating science with explanatory values adopted for universal use. Some of the old institutions were revitalized, but new ones

appeared like UNESCO (1946), the *Institut für Vergleichende Erziehungswissenshaft* in Salzburg (1946-1953), the *Pädagogische Arbeitsstelle* in Wiesbaden, later in Bonn (1947), the *Hochschule* (later *Deutsches Institut*) *für Internationale Pädagogische Forschung* in Frankfurt-am-Main (1949), the UNESCO *Institut für Pädagogik* in Hamburg (1951), the *Center of Comparative Education* at the University of Ottawa (1954) and the *Research Institute of Comparative Education and Culture* at the University of Kyushu, Japan (1954) (Brickman, 1966).

In short it can be said that comparisons are a historically well-known fact. After 1900 the scientific field of comparative education strengthened and was able to address questions concerning methodology issues, institutional organization, publishing of journals and seminars discussing how and what to compare. The governmental collection of educational numbered data developed in parallel to the scientific process. At first it was mostly descriptive data relating to different educational systems for governance reasons although there was later a shift towards a more knowledge output-based discourse (Landahl & Lundahl, 2013). These parallel processes of scientific development, a shift into «big science», and governmental collection of numbered educational data (Lawn, 2013) intertwined and created a specific style of reasoning on education that made it possible to later discuss and stage large-scale assessments of student achievements.

6. The phenomenon of large-scale assessments

Performance in schools has been increasingly judged on the basis of effective student learning outcomes. Countries inspired by the importance of comparisons and data developed tools and techniques for evaluation and assessment as part of their efforts to improve student learning outcomes, this because education is characterized as a central requirement for national economic development and political democratization. A way to deal with this is through benchmarking, identified as: «[…] basis for improvement […] It is only through such benchmarking that countries can understand relative strengths and weaknesses of their education systems and identify best practices and ways forwards» (OECD, 2006 p. 18). Another slogan is the World Bank statements saying, *Examine, assess and compare* (World Bank, 2005). Statements like these exemplify international consensus in which assessments and comparisons are seen as a necessity. However, assessments are often linked to efforts to reform educational systems and are themselves stimuli for further reforms (Baker & LeTendre, 2005).

Since the end of the nineteenth century the production of numbered data and comparisons starts to be used for bringing new visions of the social and economic world. The new construction of epistemic references for defining «reality»

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

with the help of facts perceived by numbered data is linked to the creation and management of the self-defined «democratic» state. Numerical data also provided more than an «objective way» of seeing reality, it «instituted» reality by creating a «common cognitive space» that could be both observed and described through data (Lussi, Borer, & Lawn, 2013). This «common cognitive space» has been framed by e.g. the reasoning on different international, regional and national assessments.

7. International Assessments

Data was gradually considered as a more objective way to understand «reality» (Lussi, Borer & Lawn, 2013). One offspring was the creation of International Large-Scale Assessments (ILSA) of student learning outcomes. They were created using a vision that if custom and law define what is educationally allowable within a nation, the educational systems beyond national boundaries could suggest what is possible educationally (Foshay, Thorndike, Hotyat, Pidgeon & Walker, 1962). The argument was used in introducing a pilot study in mathematics not only describing the origins of an emergent field but also predicting an exceptional growth of comparative assessment studies (Owens, 2013).

7.1. *IEA: founder of a reasoning*

The first organization formally instituted for performing ILSA is the *International Association for the Evaluation of Educational Achievement* (IEA). The world was conceived as a natural educational laboratory, where different school systems experiment to obtain optimal results in the education of youths. They assumed that science could obtain evidence from different national education systems that would make education more effective (Pettersson, 2014).

Already in 1955 a group of researchers had met to collaborate and learn about what they considered to be common educational problems. A well spread notion among various states on a desire for educational change prompted cooperation beyond geographical borders. This, coupled with a desire to increase measurement data, led to a proposal being brought to UNESCO for an international study of intellectual functioning (Foshay *et al.* 1962). The study differs from previous comparative studies in that it seeks to introduce an empirical approach into the methodology of comparative education, a field initially said to rely on cultural analysis (Foshay *et al.* 1962). The IEA embarked on the task with great enthusiasm and managed a pilot study (beginning in June 1959 and ending in June 1961) which concluded in that cross-national comparisons of educational performance could be made with comparable 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 test and evaluate (Purves, 1987).

In 1961, researchers from twelve countries met to discuss the pilot study. The study was considered a success and plans for another study in mathematics took shape. From the outset it was agreed that the project should be a co-operative enterprise. The major purpose of the inquiry was to measure achievement in mathematics and to relate that achievement to the relevant factors in the home, school and society. In determining these factors the investigation had to rely on the findings of previous research. The project called the *First International Mathematic Study* (FIMS) was said to represent an attempt to assess the efficiency or productivity of different educational systems and practices (Bloom, 1969). The final results of FIMS were presented in a publication by Husén (1967). In addition to the main study, various reports were published (e.g. Keeves, 1968; Pidgeon, 1967; Kuusinen, 1967; Hultin, 1968). In the study it became evident that there is a difference between how a subject actually is taught in the classroom and how it is described in the curriculum and that this is a good predictor of the differences in student performance. FIMS also showed that there is a lack of equity between different groups of students on how they perform. After this study the IEA performed a variety of studies on different subjects, time spans and periodicity (see e.g. Lindblad, Pettersson & Popkewitz, 2015).

7.2. *PISA The Queen of ILSA*

Indeed, it was the IEA studies that lead to many assessments being undertaken in various countries. The *Programme for International Student Assessment* (PISA) study, a project of the *Organization for Economic Cooperation and Development* (OECD), was similar to the IEA studies in many respects. Although OECD primarily has been concerned with economic policy, education has become increasingly important due to the fact that over the last 40 years education has been framed by an economic discourse related to human capital and «knowledge economy» (Pettersson, 2008). Through statistics, reports and studies OECD has activated a «common sense» in political decision making by saying that scientific «proofs» are indisputable (Martens, 2007). Martens argues 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 (Grek, 2009). Novoa and Lord (2002) state that comparisons may not be regarded as a method; instead it can in fact be seen as policy. The policy is driven by an expert discourse that, by means of comparative strategies, tends to impose natural or common sense answers in national settings (Pettersson, 2008). While OECD serves na-

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

tional policymakers well with a comparable discourse in terms of statistics, it also provides them with a global policy lexicon concerning what education is and ought to be (Pettersson, 2014).

PISA provides comparisons of the competencies of 15-year-olds that are relevant to everyday adult life, rather than simply evaluating knowledge based on curriculum (OECD, 2001). It is also said that assessments that test curriculum only offer a measure of internal efficiency, and cannot reveal how schools prepare students for adult life (OECD, 2001). PISA has the ambition to be a platform for policy construction, mediation and diffusion at national, international and even global level (Rizvi & Lingard, 2006). It can be said that PISA both shapes an international discourse and at the same time PISA is shaped by the discourse.

PISA assessments have been conducted several times. In every assessment students' knowledge in reading, mathematics and scientific literacy is tested, together with interests and backgrounds. In addition, innovative domains are also assessed, for example what is called collaborative problem solving. The emphasis on «real-life» circumstances and the capacity to enter the labor market with the relevant skills has been said to shift PISA's focus away from less explicit educational aims that are more complicated to measure (Grek, 2009).

PISA also easily connects to the idea of the self-governance of active subjects, which expands governance into a system of individual self-regulation (Ball, 2003). Even though PISA is both constructed and operates under a clear policy framework that is designed to improve future results, it is therefore not just a testing regime. PISA should also be seen in light of its ability to improve and attract economic and human capital investments. For policymakers, PISA is hence 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 – in international policy discourse (Pettersson, 2008). As these two aspects are interwoven and strengthen each other, they can hardly be analyzed separately. Besides PISA, OECD has also staged and presented various other studies (see e.g. Lindblad, Pettersson & Popkewitz, 2015).

Since the introduction of PISA, discussions have been flourishing on how to describe and distinguish between surveys performed by the IEA and the OECD. One way to explain these differences is to describe IEA as more research-oriented and OECD as more policy-oriented (Pettersson, 2008). In a comparison between 22 participant countries in the 2003 PISA and TIMSS cycle it is concluded that despite differences of curricular and literacy intent, country results are comparable (Wu, 2009). Comparable is also the methodological construction and the applications for further research emanating from PISA and TIMSS publications (Hutchinson & Schagen, 2007).

8. Regional Assessments

Parallel, another kind of learning assessment also developed – regional assessments. In these assessments instruments were developed to assess and compare learning outcomes among nationally representative samples of students within a particular region. For Europe the most salient organization conducting these studies is the EU with e.g. its knowledge assessment ESLC (*European Survey of Language Competences*). In collaboration with UNESCO two regional organizations: the *Latin American Laboratory for the Assessment of the Quality of Education* (LLECA) and the *Southern and Eastern African Consortium for Monitoring Educational Quality* (SACMEQ) perform assessments. Besides, the *Conférence des Ministres de l'Education des Pays Ayant le Français en Partage* (COMFEMEN) managing the assessment PASEC (*Programme d'Analyse des Systèmes Educatifs de la CONFEMEN*) in Francophone countries conducts regional learning assessments. All these regional bodies spurred and developed important regional assessments with the support of international, regional and national experts, as well as with the help of national and international funding sources.

The regional organizations are very much interconnected to both the methodology and the personnel involved in developing and conducting international tests. One example of this is when SACMEQ was established it was through a meeting between the Zimbabwe's Minister for Education and the Director of UNESCO's *International Institute of Educational Planning* (IIEP). On 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 Ministry decision-makers. The project resulted in a report written by Kenneth N. Ross and Neville Postlethwaite in 1991. Both Ross and especially Postlethwaite were prominent contributors within the IEA where they e.g. had cooperated in the late 1980s in the *Reading Literacy Study* (Postlethwaite & Ross, 1992).

Starting in Zimbabwe the project eventually resulted in the establishment of a wider association with more countries participating under the acronym SACMEQ. The organization took on the challenge to develop cross-national cooperative activity and after the first report in 1995 the organization and the cross-national ambition was very well appreciated. The mission is to undertake integrated research and training activities that will expand opportunities for educational planners and researchers. This will be done through (a) receiving training 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

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

can be used by decision-makers to plan the quality of education and finally, (c) utilizing innovative information dissemination approaches and a range of policy dialogue activities in order to ensure that the results are debated, discussed and understood by stakeholders and then used as the basis for policy and practice. SACMEQ have until today conducted four major studies (I-IV).

Latin American countries have only sporadically participated in ILSA, but in two regional tests their desires to participate have been much more evident. In 1997 the LLECE carried out the *First International Comparative Study in Language, Mathematics, and Associated Factors in the Third and Fourth Grades of Primary Education* (Hanushek & Woessmann, 2012). The assessment tested the performance in mathematics and reading of representative samples of students in each participating country. In 2006 another test, said to be especially design for the Latin American countries, was launched in the region called the *Second Regional Comparative and Explanatory Study*, with the acronym SERCE. The second study also tested performances in mathematics and reading.

Another regional organization with a somewhat different history is CONFEMEN founded in 1960 by states within the francophone world. In 1991 it was stated that there was a need for bringing together quantitative and qualitative aspects of educational systems and identify the most effective educational strategies. During two decades PASEC's (the acronym for the assessment staged) mission to evaluate performances have resulted in 35 national assessments in more than twenty countries in Africa and Asia. Since 2012 PASEC also implements international comparative assessments. The so called international assessments are not international in an orthodox meaning; instead international is interpreted within a tighter frame of states within francophone cooperation. The main objectives are: (a) to measure student performance and identify factors of effectiveness and equity, (b) to provide national policy indicators for comparisons in space and time, (c) to continue the development of an internal and permanent evaluation, and finally (d) to disseminate international assessment results to contribute to the quality of education. PASEC measures achievements in French (and/or national language if that is the language of instruction) and mathematics, simultaneously with contextual, institutional, social, economic and cultural data.

Also, the EU has to be considered. The *European Commission* describes the purpose of ESLC as providing participating countries with comparable data on foreign language competence and knowledge about good practice in language learning. The test is intended to provide indicators for measuring progress in foreign language learning. The survey tested the two most widely taught European languages in each country among students in their final year of lower secondary

education. Assessments were made on listening, reading and writing. In addition to the test the students were asked to fill in a questionnaire about their language learning and other background factors, this for providing data on how demographic, social, economic and educational variables affect language proficiency.

9. National Assessments

Over the years the prominence of ILSA, especially the ones conducted by the IEA contributed to the spread of a new form of assessments – national assessments (Kamens & Benavot, 2011). These assessments of cognitive achievement, but sometimes also attitudes and behavior, were conducted under the auspices of the countries themselves, but also often with support from different NGOs and IGOs or international donor agencies. Also the help with educating and disseminating assessment personnel within the IEA is important (Pettersson, 2008).

Different national assessments have been conducted regularly in history, but they have virtually exploded during the last 20 years (Kamens & Benavot, 2011). This is due to that performance in schools is within the global educational «common sense» increasingly judged on the basis of effective student learning outcomes. Hence, countries are developing a range of tools and techniques for evaluation and assessment in school systems as part of their efforts to improve their students' outcomes (Rosenkvist, 2010). At a first glance, these national assessments may appear somewhat identical, but in fact there are often substantial differences in implementation, use and design. The differences mostly arise from the fact that assessments are political phenomenon reflecting the agenda, tension, institutional norms and the very nature of power relations between different internal political actors (Kellaghan, Greaney & Murray, 2009).

The main focuses in national assessments are normally to describe and evaluate the quality of student learning outcomes. During the last 40 years a number of industrialized countries established national assessments on a more regular basis. This heightened awareness about the importance was encouraged by the declaration of the *World Conference on Education for All*, held in Jomtien, Thailand, in 1990 which stated that providing students with access to education was meaningful only if the students actually acquired useful knowledge. The *Dakar Framework for Action* in 2000 reinforced this message and also stressed the importance of having a clear definition and accurate assessment of learning outcomes. These declarations were especially important for developing countries in staging national assessments (Kamens & Benavot, 2011).

Within the report *Strong Foundations: Early Childhood Care and Education* (Encinas-Martin, 2006) a preliminary overview of national assessment activi-

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

ties is provided. In a UNESCO follow-up study, Benavot and Tanner (2007) analyze this statistics more thoroughly, concluding that national learning assessments have become a common feature of national education systems around the world. Countries performing at least one national assessment have risen steadily over time, from 65 countries (1995-1999) to 111 countries (2000-2006). The subjects assessed in national assessments are said to be predominantly curriculum-based and subject oriented. Broken down into five major subject categories, the UNESCO report (Benavot & Tanner, 2007), shows that mathematics and language are by far the most prominent subjects evaluated. This not only reflects the strong emphasis on mathematics and language in national curricula worldwide, it also shows the centrality in international assessments and educational discourse. Half of the countries assess learning outcomes in science and almost two-fifths assess learning in social science. One-fifths in foreign languages and nearly the same is true for other subjects like art, physical education, problem solving, life skills, visual literacy, coloring, cognitive behavior and music. It is also shown that there has been little change over time in the prevalence of assessed subjects, except foreign language which have gained and social science which have lost the same prominence.

10. Comparative Knowledge Assessments: Numbers, Modernity and the Meritocratic claim

Our text is focusing on two central knowledge problematics. First, the relation between education and a specific technology, framed by ideologies on modernity and meritocracy, understood as a selection to different and hierarchical positions in society by means of education performances. Second, the development and expansion of national, regional and international assessments and the increasing use of them within educational practice, policy and bureaucracy is acknowledged. In doing so we note that an historical tradition within education to compare and use data evolved into a specific technology for framing education with a centrality of numbers. Educational numbers came as such to be transformed from representations of education into education per se. This could happen due to societal historical connections to reasoning about modernity and meritocracy, which were considered as central in the development of the state and society. Porter (1995) is making an argument about that the reason that numbers came to be central in the development of society had to do with that numbers are perceived as «objective» and as such «neutral», but in reality this is in many respect false, and even contradictory. Instead, numbers should be perceived as a technology of steering and managing society and the state, a technology based on connotations of «objectivity», but also as a technology of

distance and neutrality. What we are making an argument about is connected to Porters statements. We state that comparisons and the use of numbered data for describing education dependent on parallel societal processes, in science, society and state, came to be transformed into that numbered data on education came to be perceived as education per se. This development can be described in several aspects, but we are primarily describing it through emphasizing some historical comparative and data aggregative collaborations within science and governmental organizations and later the growing importance of transnational agencies and international, regional and national assessments.

We have exemplified different spaces for knowledge acquisition as well as the expansion of tests at all levels. In this we noticed that how the different knowledge assessments on different levels are linked to each others are not that well studied. Some researchers are highlighting the growth of international large-scale assessments as an explanation for the expansion of regional and national assessments (Kamens & Benavot, 2011). However, we believe that they should be considered as parallel processes embedded within a reasoning domiciled in today's use of numbers and how this is connected to how concept like modernity and meritocracy is conceived. The manifested partnerships and cooperation's at national, regional and international levels exhibit an almost rhizomatic structure which allows for a variety of usage of data, comparisons and assessments. At the same time there is closeness through the linking of spaces for cooperation and exchange of knowledge and experience that administrators, researchers and policymakers are involved in. Consequently, there is room for both borrowing and lending and with opportunities to learn from elsewhere and transfer of policies and programs for the measurement of performance (Steiner-Khamsi, 2004).

Different comparative knowledge assessments have different purposes. The international assessments, and to some extent also the regional assessments logic is primarily about comparisons between educational systems. With the results in these assessments the quality of education is thought to be improved. National knowledge assessments have a somewhat different focus more emphasized on monitoring performances for national quality assurance. In these tests there is more about controlling the individuals', both students' and teachers, but also schools, for guaranteeing quality. Focuses in national assessments have also affected curriculum with more detailed descriptions of knowledge and grading criteria (Forsberg & Pettersson, 2015).

Our analysis of comparisons and the usage of numbered data within education show interlinkage between research and policy. This interlinkage led to a specific formulation of a technology in which considerations like comparisons, use of numbered data, statistics, meritocracy and ideas about modernity play a

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

crucial role. The examination of this development was on the style of reasoning. The arguments we put forward is the principles to organize education but also how one should act and think of solutions as important to contemporary issues of education.

Changes in the relationship between individuals', state and society have paved the way for the so-called audit society and the increasing amount of inspections and evaluations in and of education. The objectives, content and design of education are determined by the evaluation of student achievement and in this especially different «achievement gaps» (Lindblad *et al*. 2015) are emphasized. With the impact of comparative knowledge assessments it is reasonable to raise the question if Bernstein's (1975) view on the relationship between curriculum, pedagogy and evaluation expressed in his work should be reformulated. It appears in the contemporary like the strength of the structure of the signaling system of evaluation will determine the strength of classification and framing rather than vice versa. Even though we have an empirical example in this paper on how educational numbers frame discussions on education, more examples are needed and further studies are in need to address the question. However, it is clear that the selection and arrangement of the content of education today are increasingly discussed and politicized on the basis of the results in knowledge assessments and that it also manifested itself into the practices of national education assessments.

Elsewhere (Forsberg & Pettersson, 2014), the shift in terms of knowledge, against internationally anchored competency-based knowledge discourses resident in the EU, the OECD and the UNESCO, in other words, the same spaces that contains the ideology of meritocracy and comparisons, have been demonstrated. If education is narrowed down to what is measured in assessments there will also be a narrowing down of the goals in education. We can see how the what (topics and content) that are tested and how they are tested (assessments) and how the results are disseminated and used restricts what is seen as desirable skills, norms and values of education. Through these practices practitioners are also formed concerning identities and ways of looking at oneself, others and the world around us.

However, a meritocratic ideology has not only brought about assessment practices that enable and promote some, but not others, educational activities. Meritocracy also sustains and legitimizes educational distribution of life chances for different individuals. It holds beliefs about individuals' talents and abilities and how performance can be used to determine who will take part of what and serves as a tool for equality, competition and differentiation. Thereby legitimized by society's separation of powers and what we can call the justifiable inequality. Consequently, the style of reasoning prevalent in the contemporary legitimatizes

a justifiable inequality based on an ideology of meritocracy, which in itself must be legitimized through comparisons and data usage. With the dominant position that presentations of international comparisons of school results has achieved in current political discourses, such an emphasis of education as meritocratic instruments for student employability and national economic competition is also dominating current reasoning on education for the individual as well as for society.

11. References

Baker, D., & LeTendre, G. K. (2005). *National differences, global similarities: World culture and the future of schooling*. Stanford: Stanford University Press.

Ball, S. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, *18*(2), 215-228.

Bell, D. (1972). On meritocracy and equality. *The Public Interest*, *29*(48), 30-68.

Bernstein, B. (1975). *Class, Code and Control. Volume 3. Towards a Theory of Educational Transmissions*. London: Routledge & Kegan Paul.

Bloom, B. (1969). *Cross-National Study of Educational Attainment: Stage I of the IEA Investigation in Six Subject Areas. Final Report. Volume I*. Chicago: University of Chicago.

Bourdieu, P. (1971). Systems of education and systems of thought. In Young, M. F. D. (Ed.), *Knowledge and Control: New Directions in the Sociology of Education.* London: Collier-Macmillan.

Brickman, W. W. (1966) Ten Years of the Comparative Education Society. *Comparative Education Review*, *10*(1), 4-15.

Brickman, W. W. (2010). Comparative Education in the Nineteenth Century. *European Education*, *42*(2), 46-56.

Desrosières, A. (1991). How to make things which hold together: Social science, statistics and the state. In *Discourses on society* (pp. 195-218). Springer Netherlands.

Encinas-Martin, M. (2006). *A Global Survey of Educational Evaluation: International, Regional, and National Assessments of Student Learning. Paper commissioned for the EFA Global Monitoring Report 2007*. Strong Foundations: Early Childhood Care and Education.

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.

Espacio, Tiempo y Educación, v. 3, n. 1, January-July 2016, pp. 177-202.
ISSN: 2340-7263

Forsberg, E. (2006). *International tests, national assessment cultures and reform history*. Research project. Stockholm: The Swedish research council.

Forsberg, E., & Pettersson, D. (2014). European Educational Transfer and Curriculum Displacement: The Swedish Case. In Nordin, A., & Sundberg, D. (Eds.), *Transnational policy-flows in European education: Conceptualizing and governing knowledge*. Oxford Studies in Comparative Education. East Greenwich: Symposium Books.

Forsberg, E., & Pettersson, D. (2015). Meritokratin och jämförande kunskapsmätningar. In Wärvik, G-B., Runesdotter, C., Forsberg, E., Hasselgren, B., & Sahlström, F. (Red.), *Skola Lärare Samhälle: en vänbok till Sverker Lindblad*. Göteborg: Göteborgs universitet.

Fox, A. (1956). Class and Equity. *Socialist Commentary*, *11*(13).

Gautherin, J. (1993). Marc-Antoine Jullien (Jullien de Paris) (1775-1848). In *Prospects: the quarterly review of comparative education*. Vol. XXIII, No. 3/4 (pp. 757-73). Paris: UNESCO.

Grek, S. (2009). Governing by numbers: The PISA 'effect'in Europe. *Journal of education policy*, *24*(1), 23-37.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science.* Cambridge: Cambridge University Press.

Hacking, I. (1992a). «Style» for historians and philosophers. *Studies in the History and Philosophy of Science*, *23*(1), 1-20.

Hacking, I. (1992b). Statistical language, statistical truth, and statistical reason: The selfauthentification of a style of scientific reasoning. *The social dimensions of science* 3, 130-157.

Hanushek, E. A, & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics*, *99*(2), 497-512.

Holmes, B. (1981). *Comparative Education: some consideration of method*. London: Allen & Unwin.

Horkheimer, M., & Adorno, T. W. (1948). *Dialektik der Aufklärung: Philosophische Fragmente*. Frankfurt: S. Fischer Verlag.

Hultin, M. (1968). *Skolans produktivitet*. Stockholm: Almqvist & Wiksell.

Husén, T. (Ed.). (1967). *International study of educational achievement in mathematics: A comparison of twelve countries. Volume I and II*. New York: John Wiley & Sons.

Hutchinson, D., & Schagen, I. (2007). Comparisons Between PISA and TIMSS – Are We the Man with Two Watches? (http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2006/Papers/IRC2006_Hutchison_Scha en.pdf)

Igo, S. E. (2007). *The Averaged American: Surveys, Citizens, and the Making of Mass Public*. Cambridge: Harvard University Press.

Kamens, D. H., & Benavot, A. (2011). National, regional and international learning assessments: trends among developing countries 1960-2009. *Globalisation, Societies and Education*, *9*(2), 285-300.

Kandel, I. L. (1933). *Comparative Education*. Boston: Houghton Mifflin.

Kellaghan, T., Greaney, V., & Murray, S. (2009). *Using the results of a national assessment of educational achievement (Vol. 5).* World Bank Publications.

Kett, J. F. (2013). The History of a Founding Ideal from the American Revolution to the Twenty-First Century (American Institutions and Society). Ithaca: Cornell University Press.

Keeves, J. P. (1968). *Variation in Mathematics Education in Australia*. Sydney: ACER.

Kuusinen, J. (1967). *School Systems and Level of Achievement: Some Notions on the Basis of the IEA Study in Mathematics*. Helsinki: University of Finland.

Lagemann, E. C. (2000). *An elusive science. The troubling history of education research.* Chicago: University of Chicago Press.

Landahl, J., & Lundahl, C. (2013). (Mis-)Trust in Numbers: shape shifting and directions in the modern history of data in Swedish educational reform. In Lawn, M. (Ed.), *The Rise of Data in Education Systems: collection, visualization and use.* Oxford: Comparative Histories of Education, Symposium Books.

Lawn, M. (2013). The Internationalization of Education Data: exhibitions, tests, standards and associations. In Lawn, M. (Ed.), *The Rise of Data in Education Systems: collection, visualization and use*. Oxford: Comparative Histories of Education, Symposium Books.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy.* New York: Farrar, Straus & Giroux.

Levinson, D., Cookson, P. W., & Sadovnik, A. R. (2002). *Education and Sociology: An Encyvlopedia*. London: Taylor & Francis.

Lindblad, S., & Popkewitz, T. S. (2001). *Education governance and social integration and exclusion: Studies in the powers of reason and the reasons of power. A report from the EGSIE project.* Uppsala: Uppsala Reports of Education 39, Department of Education, Uppsala University.

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.

Littler. (2013). Meritocracy as Plutocracy: Marketising of Equality! Under Neolibealism. *New formations: a journal of culture/theory/politics*, *80*(81), 52-72.

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.

Mann, H. (1844). *Seventh Annual Report of the Secretary of the Board*. Boston: Dutton & Wentworth.

Martens, K. (2007). How to become an influential actor – The «comparative turn» in OECD education policy In. Martens, K., Rusconi, A., & Lutz, K. (Eds.), *Transformations of the state and global governance*. London: Routledge.

McCall, W. (1922). *How to measure in education?* New York: The Macmillian Company.

Neves, L. M. P. (2000). Putting Meritocracy in its Place. The Logic of Performance in the United States, Brazil and Japan. *Critique of Anthropology*, *20*(4), 333-358.

Noah, H., & Eckstein, M. (1969). *Toward a Science of Comparative Education*. London: Macmillan.

Nóvoa, A., & Lord, M. (Ed.). (2002). *Fabricating Europe: The formation of an education space.* Dordrecht/Boston/London: Kluwer Academic Publishers.

Nóvoa, A., & Yariv-Mashal, T. (2003). Comparative Research in Education: a mode of governance or ahistorical journey? *Comparative Education*, *39*(4), 423-438.

OECD. (2001). *Knowledge and skills for life: First results of Programme for International Student Assessment*. Paris: OECD.

OECD. (2006). *OECD (Organization for Economic Cooperation and Development). 2006. Education at a* Glance, 2006. Paris: OECD.

Owens, T. L. (2013). Thinking Beyond League Tables: a review of key PISA research questions. In Meyer, H-D., & Benavot, A. (Eds.), *PISA, Power and Policy: the emergence of global educational governance.* Oxford Studies in Comparative Education, 23:1. Oxford: Symposium Books.

Pettersson, D. (2008). *Internationell kunskapsbedömning som inslag i nationell styrning av skolan*. Acta Universitatis Upsaliensis. Uppsala Studies in Education, 120. Uppsala: Uppsala universitet.

Pettersson, D. (2014). Three Narratives: National Interpretations of PISA. *Knowledge Cultures*, *2*(4).

Pidgeon, D. A. (1967). *Achievement in Mathematics: A National Study in Secondary Schools*. NFER: Upton Park.

Popkewitz, T. (2008). *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*. New York: Routledge.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, New Jersey: Princeton University Press.

Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society.* Chicago: University of Chicago Press.

Postlethwaite, N., & Ross, K. N. (1992). *Effective schools in reading: implications for educational planners. An exploratory study.* IEA: The Hague

Purves, A. C. (1987). The Evolution of the IEA: A Memoir. *Comparative Education Review*, *31*(1), 10-28.

Rizvi, F., & Lingard, B. (2006). Globalisation and the changing nature of the OECD's educational work. In Lauder, H., Brown, P., Dillabough, J., & Halsey, A. H. (Eds.), *Education, globalization and social change*. Oxford: Oxford University Press.

Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge: Cambridge University Press.

Rosenkvist, M. A. (2010). *Using students test results for accountability and improvement: a literature review. OECD Education Working Paper No.54.* Paris: OECD.

Sapin, S. (1994). *A social history of truth. Civility and science in seventeenth-century England*. Chicago: University of Chicago Press.

Smyers, P., & Depaepe, M. (2010). Representation or Hard Evidence? The Use of Statistics in Education and Educational Research. In Smyers, P., & Depaepe, M. (Eds.), *Educational Research: The Ethics and Aesthetics of Statistics.* Educational Research 5. New York: Springer.

Smyth. (2008). The Origins of the International Standard Classification of Education. *Peabody Journal of Education*, *83*(1), 5-40.

Steiner-Khamsi, G. (2004). *The Global Politics of Educational Borrowing and Lending*. New York & London: Teachers College. Columbia University.

Stigler, S. M. (1992). A Historical View of Statistical Concepts in Psychology and Educational Research. *American Journal of Education*, 101.

World Bank. (2005). *Expanding opportunities and building competencies for young people.* Washington DC: World Bank.

Wu, M. (2009). *A Critical Comparison of the Contents of PISA and TIMSS Mathematics Assessments*.(https://edsurveys.rti.org/PISA/documents/WuA_Critical_Comparison_of_the_Contents_of_PISA_and_TIM SS_psg_WU_06.1.pdf)