THE PRODUCTION CYCLE OF PISA DATA IN BRAZIL: THE HISTORY OF DATA BEYOND THE NUMBERS

MARIALUISA VILLANI
marialuisavillani@hotmail.com | Independent Researcher, France

ABSTRACT

In this paper we analyse the process of the production cycle of the OECD's Program for International Student Assessment (PISA) data in Brazil for 2012 and 2015 editions. We present the qualitative data collected in 2016 through semi-structured interviews conducted with members of the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [National Institute for Educational Studies and Research] (INEP) who worked for PISA in Brazil for 2012 and 2015 editions. We made an ethnography of Brazilian PISA data to show the level of social embeddedness of statistics. In this paper, we propose a mindful approach to statistical data, particularly when results of international large-scale assessments (ILSA) are used to describe an educational system as complex as the Brazilian one. It is possible identify several levels of governance in PISA data production. That produces gap and misalignments that have an impact on quality of information disseminated by PISA data in Brazil.

KEY WORDS

PISA; INEP; ILSA; data production; reference model.

SISYPHUS
JOURNAL OF EDUCATION
VOLUME 6, ISSUE 03,
2018, PP.30-52

DOI: https://doi.org/10.25749/sis.15100

O ciclo de Produção dos Dados PISA no Brasil: A História dos Dados Além dos Números

MARIALUISA VILLANI

marialuisavillani@hotmail.com | Investigadora independente, França

RESUMO

Neste artigo analisamos o processo de produção dos dados do programa de avaliação internacional PISA (Program for International Student Assessment), dirigido pela OCDE, no Brasil nas edições 2012 e 2015. Apresentamos dados qualitativos coletados em 2016 por meio de entrevistas semiestruturadas realizadas com a equipe do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) que trabalhou no PISA no Brasil em 2012 e 2015. Realizamos uma etnografia dos dados do PISA para mostrar a incorporação dos aspectos sociais na produção dos dados estatísticos. Propomos uma abordagem consciente dos dados estatísticos, especialmente quando os resultados deste tipo de estudos são usados para descrever a complexidade dos sistemas educativos brasileiros. É possível individuar diferentes níveis de governança da produção dos dados estatísticos do PISA no Brasil. Isso produz uma distância e um desalinhamento que têm consequências na qualidade das informações difundidas através dos dados do PISA sobre o sistema escolar brasileiro.

PALAVRAS-CHAVE

PISA; INEP; ILSA; produção dos dados; modelo de referência.

SISYPHUS

JOURNAL OF EDUCATION

VOLUME 6, ISSUE 03,

2018, PP.30-52

DOI: https://doi.org/10.25749/sis.15100

The Production Cycle of PISA Data in Brazil: The History of Data Beyond the Numbers

Marialuisa Villani

> Instead, I ask, what world do we prefer to live in: the world without education measurements and comparisons, with its issues of imprecise goals and lack of standards, or the world of explicit targets, indicators, and statistical measurements, with its problems of unwarranted assumptions, bad measurements, misplaced goals, and inappropriate incentives? If I lived in Scandinavia or, for that matter, Boston, places where the quality of education is beyond question, I would probably choose the first, but I live in Brazil, and have always chosen the second.
>
> Schwartzman, 2013, p. 270

INTRODUCTION[1]

The aim of this article is analysing the process of PISA data production in Brazil. PISA (Program for International Student Assessment) is an International Large-Scale Assessment program created and developed by OECD since the end of the 90'. The first edition of the program was in 2000. The assessment happens each three years and evaluates literacy, numeracy and sciences competences of 15 years old pupils. Brazil is the only Southern-American country, which is not a member of OECD state that participates to PISA since the beginning of the program in each edition.

In this article are investigated and studied the processes of design, gathering and analysis of PISA data in the transition from the central level, managed by PISA consortium to the national level, managed by INEP (Instituto Nacional de Pesquisa Educacionais Anísio Teixeira. In this work I develop a mindful approach to statistical data, realising an "ethnography" of data produced by PISA in Brazil. The principal aim of my research is presenting the social, political aspects that characterise the production cycle of PISA data in Brazil, and how they influence the representation of Brazilian scholar system.

The article is organised in four parts. The first part describes the theoretical framework of the research and it is composed by the first two paragraphs of the text. The second part, represented by the third paragraphs, describes the methodology used in the research and how the qualitative data were collected and presented in the text. The third part from the fourth paragraph to the seventh paragraph, presents the results and the issues emerged from the research. Therefore the fourth part, composed by the eighth paragraph, is the final part of the article. In this part I present some conclusions,

1 This research was funded by REFEB fellowship 2016 of French Embassy in Brazil.

but more important I develop new questions that can gain a link in the analysis between statistical data, social aspect of it and relationship with educational policy analysis production in Southern-American Countries.

NEW PUBLIC MANAGEMENT, EVIDENCE-BASED POLICIES AND STANDARDISATION PROCESS IN EDUCATIONAL FIELD

The standardisation process in educational field in Southern-American countries has been developing and diffusing for the last twenty years (Rivas, 2015). This process is in part a result of demand made by local, national and international public institutions, and by policy makers and designers. Globally, the link between statistical data and policies suffered changes in the 1980s, namely with the introduction of 'Evidence Based Policy' (EBP) in United Kingdom when "New Public Management" (NPB) became the framework to design policies. This transformation was followed by a deregulation process in several policies sectors. This context also led to an important change in policy making process (Head, 2008).

Segone and Pron (2008) describe a transition to a new model of policymaking, and this change was influenced by statistics. The "old model" of policy making was based on "opinion-based policy" approach, today we are moving toward "influence-based policy". According to Segone and Pron (2008) in this process the stage of "evidence-based policy" is fundamental. Davis et al. (2000, p. 5) define "evidence-based policy as a tool that helps people make well informed decisions about policies, programs and projects by putting the best available evidence at the heart of policy development and implementation". Meanwhile, according to Scott (2005, p. 1): "Evidence-based policy-making in a democratic context means that, wherever possible, public policy decisions should be reached after an open debate which is informed by careful and rigorous analysis using sound and transparent data". The use of evidence-based policies became necessary to implement and realise the accountability process (Strassheim & Kettunen, 2014). According to some policy scholars, the role of statistics in policy design is crucial due to its "objective" nature (Scott, 2005).

However, the relationship between the production of statistical data and policy making is not as linear as one imagines. Authors inspired by the work of Foucalt (2004) on the relationship between power, policy making and technical tools, have shown that in the last thirty years, statistics became one of the "technologies of power" (Lascoumes, 2004). According to this point of view the quantification process of social phenomena revolves also around performance indicators, accountability, and other tools established by NPM (Desrosières, 2014). Use of statistic shaped by NPB framework produces a discontinuity with the old way these tools were used by governments. According to Desrosières, statistics indicators designed by NPM model produce a retroactive effect on choices, behaviours and actions of political and social actors (Desrosières, 2014).

The Foucauldian analysis of governmentality technologies compels us to consider the social scope of statistics and the way they are immersed in a "embeddeness" process. The embeddedness is a concept developed by Polany and Granovetter (Granovetter, 1973; Polany, 1977) to explain the link between economy, society and

particularly to highlight how economic development is "embedded" with the social aspect of the phenomena. We can translate this concept in policy analysis using the approach of Strassheim and Kettunen (2014). They analyse the "reflexive politicisation" of evidence-based-policy. The discussion around EBP is split about what we need to consider in knowledge production of public policy (Schütz, 1976). Selection mechanisms of knowledge production evolve in a context characterised by power asymmetry. This mechanism influences statistics production giving a social connotation to their design process.

OECD: A TRANSNATIONAL ACTOR OF EDUCATIONAL GOVERNANCE TROUGH PISA

According to Clara Morgan (2011), OECD with PISA program positions itself as a leader in statistical data design and production in educational field:

> By creating the PISA, the OECD has positioned itself as a leader in the international statistical and data gathering and production infrastructure that exists within a global architecture of education. This global structure includes several key international governmental organizations such as the World Bank, the International Monetary Foundation (IMF), and the United Nations. Regional organizations such as the European Union (EU) are also connected to it. In addition, there is the OECD's 'rival' in educational measurement – the International Association for the Evaluation of Educational Achievement (IEA) which was created under the auspices of the UNESCO Institute for Education (UIE). (Morgan, 2011, p. 4)

Thus, we can define OECD as a transnational actor through PISA also due to its expansions in several dimensions (Carvalho, 2012). First of all, the sixth editions of assessment were realised at same time (time expansions) in all the continents of the world (over the years we moved from 32 countries to 72 participant countries), geographical and political expansion. Secondly, we also register an expansion of skills evaluated in PISA assessment, such as inclusion of ICT[2] competences or financial competences in numeracy. Thirdly, this expansion produces a link between PISA and other OECD assessment programs.

According to Mons (2007), PISA program develops a "universalistic conception" of comparison in educational field with an educational system analysis that aims to demonstrate the isomorphism process and how all countries chose the same goals to achieve. Consequently, a comparison realised with the same analysis and representation criteria seems justifiable at methodological level.

Clara Morgan (2011) views a supranational institution as OECD with its assessment programs as having a goal of measuring the human capital. One of the first PISA publications defines the evaluation paradigm as follows:

2 Information and Communication Technology.

> The indicators are designed to contribute to an understanding of the extent to which education systems in participating countries are preparing their students to become lifelong learners and to play constructive roles as citizens in society. (OECD, 1999, p. 7)

Morgan affirms that the underlying goal of OECD countries and other non-member participant countries in PISA is measuring international competitiveness of their labour force in a society based on knowledge-economy. Bottani and Vrignaud (2005) meanwhile note that the normative force and influence of PISA in political agenda of participant countries is due to nature of OECD, in other words *"political organ"*.

It is very important to highlight that OECD does not have any legal tools to influence policymaking, and despite this fact it succeeded to develop and diffuse a policy and politic discourse based on international assessments and rankings (Pettersson, 2014). PISA shoulders two functions in the construction of international policy discourse: economical and educational (Pettersson, 2008, 2014). According to Luís Miguel Carvalho:

> Together with the inquiry, relevant face-to-face activities take place (in meetings, workshops, seminars, etc.). Likewise, multiple publications – apart from the survey's main reports – are generated and have a worldwide flow. All these activities involve a great variety of social worlds – public and private research centres, experts, OECD professionals, policy makers, bureaucrats and technicians from many countries – and multiple kinds of knowledge, interests and perspectives. The program is, from this point of view, a complex of activities, objects and actors that generates diverse resources for social action in various social spaces. (Carvalho, 2012, p. 173)

When we analyse the design and the implementation process of PISA we need to consider the "expertise monopoly" produced by the program. There are, in private and public organisations, a group of 10-20 experts from several OECD fields that participate in different ways in coordination (surveys, exchanges, publications). In this coordination process these experts mobilize, also, several actors of PISA implementation and internal and external agencies (Carvalho, 2012).

Felouzis and Charmillot (2012) point out that the analysis and reception process of PISA data is the result of the appropriation model of each participant country, and is also influenced by the state of public statistics of each country. The researchers assert that countries such as United Kingdom or France where statistics in educational field are very developed do not respond with concrete actions to PISA results. On the contrary, countries where PISA is the only evaluation tool, the program assumes the role of guidance.

Following Felouzis and Charmillot (2012) classification, and considering the institutions and agencies (in addition to INEP) that produce educational statistical data in Brazil, I guess that is possible collocate Brazil in an hypothetical mapping of educational data production on the border between countries with a strong statistical tradition and those who use OECD data as reference. This is one of the reasons it is important to analyse Brazil as a case study of PISA program. Thus, data collected through the

interviews at the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [National Institute for Educational Studies and Research] (INEP) with researchers in charge of PISA for 2012 and 2015 edition give us the possibility to understand the impact of the "PISA model" on design and implementation of standardised assessment at national level. 3

When we analyse the influence of PISA in Southern-American countries we also need to consider other factors that shape this process. Firstly, Southern- American countries that participate to PISA have a very unequal distribution of economic and social resources. Despite very significant changes, these elements remain important factors of educational inequalities for those contexts (Rivas, 2015).

> These promising trends cannot overlook large social and education debts in the countries in the region. The increase in financing and education rights was a trend that ran parallel to the structural continuity of unequal societies and was not enough to re-vert large needs at schools, especially in rural areas in the poorest countries. (Rivas, 2015, p. 8)

One of the major problems in Brazil educational system (as in other Southern-American countries) is the access to compulsory education for a large population of school-age children. The question of representation and of pupils in Brazil scholar system is very important to analyse the state of art of this country. Other issues that are specific of Brazil and Southern-American countries will be then presented through our data.

THE RESEARCH

It is very important to describe the political context of research that had an impact on its realisation. The research was realised during March and August 2016 when Impeachment of former president Dilma Rousseff was happening in parallel with my arrival to Brazil. This element has a strong impact on political-institutional system of the country that produced changes of governance, therefore on INEP as well. These facts created significant difficulty in establishing contact with the organisation and the conduct of interviews with the PISA team. Due to the nature of the topic, and to respect the anonymity of interviewed researchers, their identity or role are made anonymous. I meet 8 researchers for the two editions considered (PISA 2012 and PISA 2015)[4] . The PISA team at INEP is small (introduce number of members), the Institution externalise a big part of assessment administration. The history life cycle of PISA data in Brazil, for the two editions studied, was influenced by a lot of changes within the research team. These changes had, also an impact on organisation and administration of test.

For each edition of PISA, we can distinguish two parts of the program: the centralized part managed by the PISA consortium, and the national part that is managed by a local

3 Furthermore, Brazil was the only Southern-American country that has participated in PISA from the beginning. Since 2013 Brazil has been member of the governing board of PISA (PGB) and its representatives have occupied a vice chair position.

4 From this point on, we refer to interviewees with using an int# and a number from 1 to 8. All the excerpts of the interviews were translated by the author from the original text in Brazilian Portuguese.

research group and supervised by consortium. In this process other actors participate in PISA. In Brazilian case for each edition that we studied, the administration of the test in the school was on the charge of private companies. The scheme in figure 1 summarise the production of PISA data and in particular the transition from central level to national level of the program.

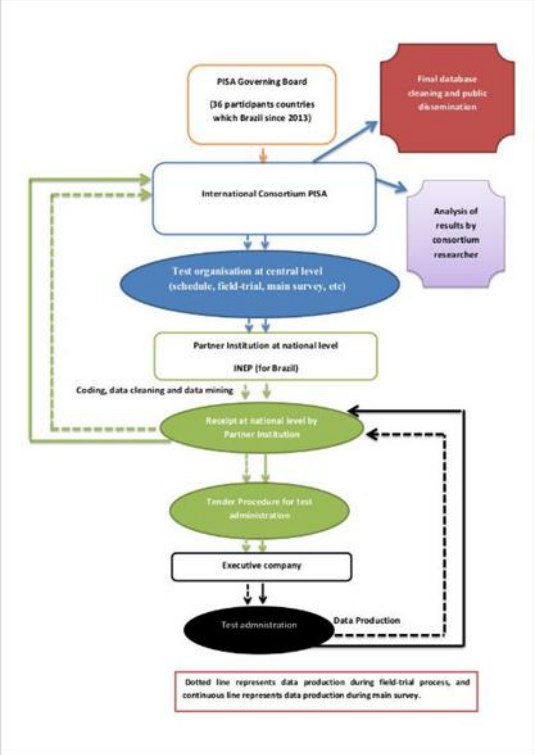

Figure 1. The cycle life of PISA data production.

FROM THE CENTRAL TO NATIONAL LEVEL

Based on the technical report and our data, we could notice that the design of test is managed at a central level by PISA consortium. The first step of the program is the experts' meeting at the central level in order to decide which items are kept from precedent edition, and which new items will be tested for the field-trial process. In each edition, PISA assessment focuses on literacy, numeracy or sciences. According to the OECD, the choice to maintain a stable corpus of items for each edition of assessment allows to compare pupils' performances with a diachronic approach (Le Donné, 2015). It is important to highlight that PISA data are not designed as longitudinal tools. The population assessed changes with each edition, consequently in order to maintain a statistic rigour it is not possible apply econometric duration models or develop analysis in a longitudinal approach on PISA database.

In this part of the process, some participant countries can propose new items to improve the assessment. In the Brazilian case, despite the fact that the country participates in the assessment since the beginning, it had the opportunity to propose some items just for 2018 edition. That happened for several reasons, but it is important to highlight two in particular: 1) the entrance in 2013 of Brazil in the governing board of PISA, that positioned Brazil in a decision-making place; 2). As has been presented above, the several changes of management within the team, with the consequence that INEP researchers have not been able to participate. The interviewee#2 describes us that he was not able to participate in this first part of the assessment, and contribute to the choices made:

> Int #2: The Inep was in a heavy institutional crisis within general management because of a leak regarding a national test. PISA was at drift after 8 months of stalemate...they call me...to try…to implement the program, everything was done within a short timeframe, I worked a lot to respect the schedule, the implementation of field-trial in September... Anything was ready, all tasks to realise, the sampling were blocked because of an institutional crisis during which several managers who succeeded one another and also three presidents of Inep followed one another...during my three years in the program.

After the first stage the national team started to design the proposition of sampling and they received the first draft of test to translate. Translation is made from French and English test in the official language of the country.

THE SAMPLING STAGE

The sampling design is a very important part of the life cycle of data, because the choices of stratification variables determine the sample size, and define the kind of information we can obtain of it. Following the size of the country and the size of sample, the implementation process of PISA assessment become more complex due to application standards, imposed by OECD, to achieve. The Sample of Brazil is one of the biggest among the participant countries.

In 2012 editions in Brazil more of 20.000 pupils were assessed, and the final sample was made by 19.204 pupils. In 2015 more of 24.000 pupils were assessed and the final sample was made by 23.141 pupils. In 2012 Brazil (a small sample of other participant countries) during the main survey realised for a subsample of 5.506 a computer-based assessment (cba). This was a trial in anticipation of cba assessment implementation in

2015 PISA edition. In 2015, 53 countries out of 65 administered a (cba)[5] test, and Brazil was one of them[6].

The 2012 edition of PISA in Brazil has experienced interruptions of implementation process during 8 months until the new team take charge of the program at national level. In 2010 at the arrival of new team, the calendar of the assessment process had not been observed. The sampling and field-trial stages have been held trying to catch up with the schedule. The Brazilian sampling for the two studied editions was designed using two step method of stratification (OECD, 2014, 2016a, part 4).

Table 1
Stratification variables

PISA edition	Method	Explicit stratification variables	Implicit Stratification variable
2012	2 stages	State (27); Maintenance (3); Certainty Selections	Administration (3); DHI Quintiles (6); ISCED level (4); Urbanicity (2)
2015	2 stages	State (27) Modal grade (2); Certainty selections	Funding (5); HDI quintiles (5); ISCED level (3); Capital/Interior (2); Urbanization (2)

Source : OECD, 2014, 2016a.

For 2012 edition they choose 81 explicit stratification variables[7], and 55 for 2015 edition.

We can analyse the sampling design following the keys, technical and political, as our interviewees described. Since the first edition of the program were excluded from the sample: Indigenous schools or all schools whose first language was not Portuguese; as international schools, or "quilombola school"[8]; North-east rural school on account of logistic implementations problems of test administration. Furthermore, in some case these schools have one multilevel-grade class, that do not allow the certitude of preparation of pupils (MEC/INEP, 2013).

5 In 2012, 35 pupils passed the paper test, if the school administrated also the cba test 18 of the 35 passed both of the assessments.

6 In 2015, 42 pupils for school participated to the assessment. The administration of cba test has been done by logging on offline platform based on a flash drive. For school questionnaire, the principals had to respond to online survey by logging on specific platform (OECD, 2016 a). A paper test was designed for 12 participant countries (Algeria, Argentina (Ciudad Autonoma of Buenos Aires), FYROM, Georgia, Indonesia, Jordan, Kosovo, Lebanon, Moldavia, Puerto Rico(USA), Romania, Trinidad and Tobago).

7 The comprehensive national list of all eligible schools is called the school sampling frame. Prior to sampling, schools in the sampling frame can be assigned to a predetermined number of explicit strata (mutually exclusive groups of schools which together cover the whole school sampling frame) and/or implicit strata (variables for sorting schools in the explicit strata, or the whole school frame if no explicit strata are used). If explicit strata are used, the school sample is allocated over the explicit strata in proportion to the PISA students in each stratum. A minimum of 150 schools will be selected in each country, with the requirements of national options often requiring a somewhat larger sample. Note that the minimum of 150 schools is selected with the expectation that there will be at least 150 participating schools, once field exclusions, ineligibility and nonresponse are accounted for. As the schools are sampled, replacement schools are simultaneously identified, should they be needed to replace eligible, non-excluded and non-participating sampled schools. (OECD: https://www.oecd.org/pisa/pisaproducts/PISA2012MS-SamplingGuidelines-.pdf)

8 Quilombola school is a school situated in a Quilombo, where first language is not Portuguese. Quilombo is a In Brazil a remote settlement originally founded as a refuge by fugitive slaves or their descendants. The Brazilian Constitution of 1988 for the first time recognized the right of descendants of slave-era quilombos to receive lands from the state (Art. 68).

For 2012 edition, after the administration of the PISA test, the INEP chose to drop out of sample two others category[9]: All rural settlement school of the country; Pupils of 7th grade of education, due to scholar system reform implemented in 2006 with federal law 11.274. This reform changes the age start of compulsory education in Brazil from 7 years old to 6 years old. Consequently, compulsory school in Brazil provides for 9 education years.

The exclusion of rural pupils was justified as a willingness of more rigour to achieve the OECD standard (because since the beginning any north-rural schools were assessed, they chose to exclude all the other rural settlement school). This procedure demanded a political intervention from INEP to PISA governing board:

> int #1: Concerning the sample the problem is (with is growth) ...is error growth, error rise with rate of pupils who are represented...and also school strike. These two things were a problem…but they understood (reference to OECD), and they accepted.
>
> MV: Why did they accept? I would like to know if at political level there were problems between OECD and Brazil (INEP), the sample issue, did it create any difficulties?
>
> Int #1 : About that Int#2…..knows very well diary problems, all decisions that we had to take and the impact that produced, as the exclusion of rural school pupils, before we administered the test to rural school pupils, but we decided to exclude all rural school pupils due to problem access in these places, the complexity of the process…..For that we had to write a technical note, everything passed by political sphere.

We can note how technical sphere of statistics and political world is nowadays crucial in data production. Brazilian case shows that these elements influence each other; on one hand politics is at service of statistics to INEP sample change is accepted by OECD, and on the other hand, the INEP establishment, according to interviewee #2, obtained validation by OECD. In this case, it seems that statistic is at service of politics. In this part, we find all the elements related to governmentality process described by Foucault (2004).

AGE TOPIC

For Southern-American countries the age issue is important. In the sample design the way we consider age of pupils has an impact on the quality of data as shown by Prais (2003) or Klein (2011), who produce work with a focus on Brazil context. Rubem Klein explores the comparability issue of South American educational systems and Western educational systems. In this article, he analyses how the use of age as continuous or discrete variable has an impact on pupils' performances. He studies the results of three Northern countries (USA, Mexico, and Luxembourg) and those of three South-American countries (Chili, Argentine, Brazil). He shows that during three different editions and in

9 For these changes on final Brazil sample INEP presented a technical note to PISA consortium that was accepted in one of consortium meeting in 2013. During our interview we had the possibility to have access a copy of this document, which is part of our data.

the countries the definition of scholar age changes. The month and the year of pupil determine in which grade he should be enrolled. According to Klein the change of month of the test administration, and consequently, the change of definition of pupils' age, has an impact on definition of population evaluated and on their performances. The assessment bias would happen, as reported by Klein, due to school age used as reference, but also as a result of the period during which the administration test happens. In his article, he explains how three of studied countries have been positive developments between 2000 and 2009 (specifically Luxembourg, Chili, and Brazil) they had changed during these three editions the date of test administration and the way to define age of pupils (Klein, 2011). In line with Klein, the solution will be to select 15 years old pupils and assess them after an identical number of month teachings, depending on the beginning of school year for each participant country.

Table 2
The Administration test period PISA 2000-2015 in Brazil

Month and age of the test administration	Eligible Pupils
October 2000	Born from 1st July 1984 and 30th of June 1985 enrolled 7th grade/8th years of compulsory school.
August 2003	Born from 1st May 1987 and 30th of April 1988 enrolled 7th grade/8th years of compulsory school.
August 2006	Born from 1st May 1990 and 30th of April 1991 enrolled 7th grade/8th years of compulsory school.
May 2009	Born from 1st January 1993 and 31th of December 1993 enrolled 7th grade/8th years of compulsory school.
May 2012	Born from 1st January 1996 and 31th of December 1996 enrolled 7th grade/8th years of compulsory school.
May 2015	Born from 1st January 1999 and 31th of December 1999 enrolled 7th grade of compulsory school.

Source: MEC/INEP, 2016.

During the interviews, Klein's (2011) article was mentioned by interviewee #1 and interviewee #2. The first used this analysis to support the technical choices made by INEP regarding the sample. The second has made reservation regarding the analysis proposed by Klein (2011). We found references to this article also in the Brazilian national report of PISA2015, to support the choices made by INEP team on sampling process.

At this stage, we find a harmonisation process of Brazilian educational system in line with Western educational systems that bring the years of compulsory education to nine. These changes happened in 2006[10], and we started to see the effects on 2012 PISA edition, that corresponds to a very particular moment of Brazilian history in PISA. In

10 Law 11.274/2006.

October 2013 Brazil become the 35th full member of PISA governing board. Since this moment, the position of the country in the program changes. Its role grows within the PISA. Brazil had the opportunity to present national issues at PISA Governing board.

The 2015 Brazilian sample was designed following the 2012 model. In line with OECD, for this assessment, INEP excluded from their studied population these categories: indigenous schools, North-East rural schools, schools placed in rural settlement, quilombolas schools, and schools situated in sustainable protected area. INEP declares that the rate of excluded school respect the OECD standard. In 2015, 6431 schools were excluded from the sample, an estimated rate of 2,2% calculated on Censo[11] escolar 2013 (MEC/INEP, 2016).

After this stage, if the sampling plan is accepted by the OECD consortium, countries begin the organisation of field trial and the translation of the test. The field trial as important to verify the god fit of sample design, and also to identify the dodgy[12] items to drop out of the test.

In this phase INEP, for both editions studied, produces a tender procedure for the translation of the test, and the administration of trial field and main survey. That creates a market where external actors participate in data collection. In the 2012 edition, the tender procedure generates bureaucratic problems:

> Int #2
> Until field trial stage these are… these are the steps…translation of assessment tools, tools assembly, sampling of the field trial, organising the database, cleaning and codification the database, and all these steps are still part of the field trial stage...Until the moment we send the database, we finished the work and we started to work on PISA2009 results, that were not exploited by INEP, therefore we were able to write a research report in this period, and we prepared the adaptation for the main survey, then the new tasks for the main survey sampling design…..there were all the items that were administrated, we needed to translate and adapt some linguistic terms, some elements that were not accurate, and the consortium sent us adjustments for the main survey administration….We outsourced with four contract with privates companies for the field trial as the main survey we realised the administration process in a rush…..For the field trial I knew about the delay, for main survey I did not expect to be so late….in August I sent the plans to be validated, and to implement the bids for the administration test, bid is a process that stipulates transparent contracts…what happened is that the process… I started in August has delayed with a back and forth of documents….when February came after six or seven months…Then we did manage the test administration of PISA2012 with delay, we had the whole month of May to administer the test and I think we realised the test administration only during the last two weeks, we had specific characteristics related to Brazilian context that do not exist outside the country….That they (OECD) do not

11 Censo Escolar is a collection of statistical-educational data at national level. In this survey, we found information regarding: mainstream education (from primary to the end of secondary school), Educação Especial [special needs education], EJA - Educação Jovens Adultos [Education for young adults].

12 If a source element had been altered by the Consortium, or if it was "dodgy", it worked in an unexpected way during the Field Trial at national or international level, the National Centre was required to provide the Main Survey adaptation. All flagged items are considered to be dodgy items either nationally if a problem occurs only in a particular country, or internationally if the same problem occurs in many countries (in more than 50% of cases).

> understand….so during the test administration, for instance, we had a lot of teachers strikes, so a lot of things have been impacted, for example, to draft the list of participant pupils. Some schools were on strike or closed…so they did not deliver class since the end of previous scholar year.

The translation and adjustment of the tests from the English or French original tool to national language (in our case in Portuguese) is a very important moment that has an impact on pupils' performances. Several studies have drawn attention to the cultural and linguistic bias produced by translation process (El Masri, Baird & Graesser, 2016; Prais, 2003; Robin, 2002). For these reasons, in this stage, national team have to confront the consortium in order to design an assessment tool adapted to the local context. For the 2015 editions the int#7 describes the adjustment difficulties that INEP team faced. Some items presented bias related to their significance in Portuguese (Prais, 2003; Robin, 2002). To succeed in changing the question the INEP operated a negation at technical and political level.

> Int #7: there was some items that had translation problems, science items specifically, I talk about science item…….I remember of an item that had name of animals which did not have meanings…..which was mistranslated….and they (OECD) accepted our suggestion, we had a openness to send back (to the consortium) the items and ask for a change.

About this issue, interviewee#7 asserts that OECD consortium was more open to accept changes on open-ended items. That was a result of negotiation of several Lusophone participant countries (as Portugal) that had commitment to change the biased items. The int#7 followed a cycle of training organised by the PISA program (these courses happened concurrently the field trial and the main survey), in these occasions researchers of national teams had the possibility to meet and exchange experiences of PISA implementation process. During these meetings, the Lusophone countries have been successful in convincing the OECD consortium to accept changes on the Portuguese test.

THE ADMINISTRATION STAGE

In this part of the cycle of PISA data production new actors play a role in the process. In Brazilian case for both of editions private executing companies, took charge of the test administration, for field trial and main survey stages. In 2012 two different companies administered the test (one for the field trial, and one for the main survey). In 2015 one executing company managed both of the stages. Considering the size and the geographic structure of Brazil, the administration stage can be a difficult process. In addition, the sample size can be producing bias during the administration stage that has an impact also on the quality of the PISA data.

> Int#2: Another thing I fixed and that bothered me was that in one school the assessment was administered below a tree...this was the most critical point…This expansion of sample made in Brazil plays against Brazil performances, because you do not have a lot of control. You are not able to bring the school reality to know if you could administer the test, reach the school, you talk about the PISA relevance, in 2015 we got improve some aspects, I left a folder for the new team with indications.

Our data show that there exists a specificity of Brazilian educational system which influenced both of the considered editions: the teachers or students strike. This element has a high impact on the test administration due to one of standards imposed by OECD, i.e. the obligation for participant pupils to have had at least six weeks of regular class upon test participation.

> Int#2: You have Brazilian specificities that do not exist out of the country, they (OECD) do not understand,…so during the test administration, for example, in Brazil we have a lot of teachers strikes, and a lot of things that have an impact on (the process)….for example, to collect the participant pupils list….they (the school) did not give lesson since the previous school year…ok you obtained the lists and finally you can administer the assessment...in other schools you find others situations...in the next school the strike was starting during the administration period….so you have a lot of difficulties ….in PISA you have a rule, the pupil needs to have had 6 weeks of lesson to participate in the assessment …so you are not able to administer the test because just the 30% or 40% (of school) had given classes….because the school year started in March, and then two weeks of vacation, and the strike…..so we do not have six weeks….the other school is striking, the teachers invite the pupils to participate, but the pupils were not attending the lessons…..so these standards are very good on paper, but when you struggle with Brazil reality, it is another situations.

In 2015 school strikes impacted again the progress of the edition, but in this case even more importantly. For one of Brazilian states INEP registered a sample representation problem because schools were often closed during the administration of the PISA test. When I delivered my interviews the results for PISA2015 were not yet released, consequently INEP researchers were not able to produce analysis about the influence of the strike on survey.

> Int #6: During the administration process I tried to pay attention on school participation, if there were on strike and there were closed, if it was necessary exclude a school and replace it….. because a school with a participation rate below 25% of population was not considered in the final analysis…..I tried to make this analysis during the test administration but since the process was realised in the rush, sometimes the executing company did not give us forthwith information….I just saw rapidly with the idea to realise analysis in a second time, but it was difficult because at any time (the school status) could

> change regarding participation, so started, but as I saw that it was not good feedback, I have left it to the end…I waited to see how many schools registered a good participation rate, and as I so it was acceptable...We need to hope that it did not have an influence, and according to international consortium (the participation) was sufficient, in this way we did manage the survey.

According to interviewee #6 the strike issue influenced the test administration, he decided to analyse the rate of school participation or other elements that have produced a bias on data. This was a strategy adopted by INEP throughout the program. The int#6 announced during the interview that 30.000 pupils and 964 schools participated in PISA assessment. Nowadays we know that PISA database for Brazil is composed of 23.141 cases for a number of 841 schools.

The introduction of computer-based assessment added logistical problems for both of considered editions. In 2012 INEP faced of problems related to the incompatibility of operational system used for cba test, the lack or insufficiency of laptops in several participant schools. To overcome these issues in PISA2015 INEP decided to rent laptops (distributed to all participant school in the country). Despite that, INEP faced the difficulty of access to schools, and the assault risk.[13]

The administration stage is a delicate process. The outsourcing of the process creates an ulterior level of governance in data production. We define this stage a sort of *"black box"* where INEP does not have all the control of all the process and that can produce bias on data (fig 1).

After the administration process test are graded by professors trained and controlled directly by the INEP team. We find new actors that participate in data production; also, we register a new level of *"uncertainty"* and *"discretion"* of the process.

> Int#2:Another problem is represented by codification….So PISA for the codification process has a book of 300 pages….if a pupil responds "I do not know, but maybe it can be that"….when you need to note this item you will find an example of another country that already has noted it, it asked(the country) to OECD consortium and consortium already did an answer…So anything you do, any point you give you do not do that with discretion, you have standards and rules, you have the rules if this answer is good, partially good or false…you give 0 point, 1 point or 2 points….more or less does not exist….you codify the pupil answer….you have all the historical….in a case of doubt you need to consult the international consortium…If you cannot solve with answers codified on the book by other countries, moreover our codification process happens after the other countries, you have almost all the answers ready…..for this pupils you give 2 points,….this is 21 code, code 22 for this question….and there was people that had work in PISA that said "look in case of doubt the merit is to the pupil"…so in this way the codification was realised until 2009, was realised in a way does not follow all the PISA rules….so in a case of doubt the point is for the pupil, because if you give the point for the pupil then the point is for Brazil….I

13 In this paper we show the process of PISA data production, other works analysing the effect of digital divide and pupils' performances (Jerrim, 2016; Komatsu & Rappleye, 2017), and the effect of cba tools on the assessment.

made a lot of correction in this direction...PISA said…look for these booklets you need to codify four times, because you will know if this booklets….you will take the proof-reader errors of this booklet, if you disseminate that on all the sample…So here you correct the first time, the second time, and if you have a disparity you give one(point)….so there were things do not follow the international psychometrics standards imposed by PISA.

After the correction, and the creation of a national database, these data are sent at central level of OECD. The consortium collects the entire national database, it makes the database cleaning, including a second check on dodgy items, and starts the analysis.

Generally, in this moment at national level, research team starts the work for the next edition of the assessment. OECD produces the main analysis on the final database, and usually it releases in the August of the successive year of assessment realisation the database for the national research team. In December of the successive year of PISA realisation the database and the analysis are publicly released.

In addition to the OECD comparison, analysis and ranking, each participant country realises a national report and develops its own analysis model. From the interviews, arises a distance from the research interests of OECD and those of Brazil as a country. What we are facing here is a process of *"misalignment"*, as Carvalho describes (2012), between the analysis produced by OECD and those developed in Brazil. We have on one part the question of *"immigration"*, a phenomenon that impacts (until now) the Western educational system, but which has no impact on the Brazilian; and secondly the dimension of *"pupils' exclusion"* that is barely present in OECD analysis.

Int#2: There is an international debate about that (the pupils' exclusion), for example in the Vietnam case. Vietnam is positioned among the top first 15th countries, and from 11th countries in math, when you look at the Vietnam sample, you have 40% of pupils at school and 60% out of school. So when OECD places Vietnam on top of the ranking, you are placing a country with high levels of pupils exclusion from school on the top…Shanghai is an example as well (on pupils exclusion), where there is a debate made by an American professor named Tom Loveless[14]: he takes the numbers of Shanghai population, (which are more or less those of Portugal population), but the number of students enrolled at school is higher than Shanghai….he starts (the American professor) his analysis by describing how Shanghai excludes pupils from their educational system, while OECD places them on top. I say that (my position is that), considering that Brazil's performances worsen ….we have a rate of pupils out of school as well, but when in Vietnam's case that advantages it, same as in the Shanghai case…they (OECD) do not discuss the sample issue (on its measurements), it is only an annex in the technical report….so this is a shame….that you do not develop a debate on that ….as these issues exist….they influence the results…. when you're placing them (the Vietnam and Shanghai's case) as an educational model…I am not saying that without OECD standards the performance of Vietnam changes…Vietnam will always

14 https://www.brookings.edu/research/lessons-from-the-pisa-shanghai-controversy/

> have a better performance than Brazil…because as Brazil also excludes pupils…and you need to consider also the representation issue in the OECD analysis…They (OECD) do not apply that, and is not for lack of attention…..they do not do that because (for them) it is not an issue that should be raised…but I think it should.

Analysing the PISA report, we note that the representative issue is mentioned only in the 2012 analysis. In this report, it is underlined the element that in most of participant countries the number of out-school children is low (OECD, 2013, Volume II, p. 53; OECD, 2016b, Volume I, p. 44). For this reason, this variable is not included in estimation model regarding performances. This factor has had an impact on few countries:

> In most countries and economies that participate in PISA, the proportion of 15-year-olds who are not enrolled in school is very small; thus, including proxy results for them would have limited impact on an education system's mean performance. For example, one could assume that students outside the education system would score at the lowest level of performance identified in PISA (the lower end of Level 1 in the mathematics performance scale, or 358 score points) and weight these students by their proportion in the population of 15-year-olds. As Table II.2.12 shows, most countries would show no or very small changes in average performance after taking into account the performance of those outside the school system. Of the six countries and economies where the changes in performance are largest (more than 15 score points; Hong Kong-China, Macao-China, Mexico, Shanghai-China, Turkey and Viet Nam), only Viet Nam would drop a significant number of places in its relative rank (up to 24 places in the ranking, followed by Macao-China, with a loss of up to 10 places). (OECD, 2013, Volume II, p. 53; OECD, 2016b, Volume I, p. 44)

It is important to highlight that Shanghai is one of the countries with the best performances and the not consideration of the inclusive variable improves the average score of 24 points. We consider that this is as an important issue, therefore is necessary to debate and analyse this in a public space at an international level. Tom Loveless (2014) analyses this controversy describing *"the hukuo"*[15] phenomenon and the impact on inequality of access to education. According to Loveless the impact of *hukou* on the 15-year-old population in Shanghai is easily seen in PISA performances.

As on Carvalho (2012) and Normand (2015) studies, PISA is a vector to design and disseminate new sets of tools for knowledge production. Our interviews show an example of this phenomenon. In 2016 INEP started a project titled "A Hora do PISA"[16] [PISA TIME] that was an interactive platform where teachers and pupils could find instruments and tests to train on PISA assessment. That creates:

15 The Chinese hukou system is unique in the world. Started by Mao Zedong in 1958 as a tool for controlling internal migration from rural to urban areas, hukou is a household registration system that restricts rural migrants' access to urban social services, including education.

16 A hora do PISA [Pisa Time] was a pedagogical platform designed for teachers, to be familiar with PISA items and the idea of competences, and students to train PISA test.

- New pedagogic tools that improve teacher's instruments and allow familiarising with other kinds of assessment;
- The production offers a knowledge space to stablish strategies to pass the assessment, in other words the program privileges learning for success more than learning for acquiring competences and knowledge.

What we find here is also a translation of the assessment nature. The program becomes a *"subject"* of the educational system debate and a goal to be achieved. We can see how the data translates from an "instrument" to a "subject" on educational process.

CONCLUSIONS

This reconstruction shows some of the social, cultural and political aspects of a standardised statistic process. We have several levels of governance in the process, with different public and private actors.

Using a Foucauldian paradigm, we can identify, as Morgan (2011) describes three level of relationship between PISA and the participant countries: a relation aimed at developing technical competences, relationship of communication, and relationship of power. This relationship changes among the participant countries. In Brazilian case, we find traces of the relationship in developing technical competences (the introduction within the INEP team the new assessment approach), our interviewees assert that their experiences in PISA training and workshop had improved their knowledge and skills in a very good way. We registered also a link between politics and statistics in Brazilian case that has an impact on PISA implementation.

Our data show that there exists a distance between Southern-American countries and OECD stakes in educational field. We consider this as a result of the *"reference model"* used by the OECD to design, implemented the assessment and analysed the results of it.

The program is settled on the *"Western educational system"* and we found this at the methodological and analytical level. At the methodological level, for example the variables chosen for the sample stratification do not include a lot of the educational reality, as in the Brazilian case. When we use the Brazilian PISA data, we describe mainly the *"Brazilian pupils of urban centre"*. If we do not have the knowledge of the Brazilian context, we can produce biased analysis, or represent only a part of the pupil population as a generalisation. The PISA standards that regulate the assessment process do not consider the local problems that can have an impact on data production. As we show in the Brazilian case, the school strikes have affected the two analysed editions. The size of the country, and access difficulties of several regions of it, also produced logistical difficulties for the implantation of the assessment.

Another element is represented by the period when the test is administrated. The choice of administration produces a lot of differences between pupils from Western countries and Southern-American pupils.

An official letter [17] sent from the representative of educational Ministry of MERCOSUR[18] countries to PISA deputy director Andreas Schleicher is an example of this matter. In this letter, the MERCOSUR members focus on four issues that influence the assessment process in their countries.

The first element was related to the gap in the school path of pupils and their age. Southern-American educational systems have differences based on school tracks. What they propose is a separate analysis between: population of 15-year-old pupils enrolled in their *"theoretical graduation-age"* and population of 15-year-old pupils that have a repetition background. According to MERCOSUR members, this differentiation allows the collection of more information about their educational systems, and limits the differences with countries where the educational system is organised on a common track in compulsory grade. The issue of age also emerged in the interviews as related to sampling process. Secondly, MERCOSUR members do not agree with OECD and PISA analysts about showing their results based on ranking and countries comparison. They suggest developing another way of focusing on regional issues, in order to give more importance to the socio-economic and cultural context. We need to find dissemination approach not based on the relative position of the countries and that considers the structural differences of the countries. Thirdly, they proposed to include in PISA items that better represent social-cultural context of MERCOSUR countries who participate in PISA. Fourthly, in order to facilitate digital assessment in countries where digital divide in school is a real issue they suggest using free software for delivering the assessment. Our data show that other issues related to statistics, policy and politics can be important in the production cycle of PISA data in Brazil.

In this paper, we find the *"misalignment"* illustrated by Carvalho (2012) that need to be developed as an issue in international large assessment. That debate is necessary to design data that consider all these contextual factors. Furthermore, we consider as important a *"database ethnography"* based on historical and social analysis of educational data design, in order to produce secondary analyses that consider the nature and quality of it. Considering all the elements presented, to continue and improve this debate, we are presented with several questions: How can Southern-American Countries, such as Brazil, play the assessment game when the rules are not adapted to them? Is it possible to use *"a reference model"* in design of data, settled not only in Western educational systems? How can we make secondary analyses which consider contextual factors that influence data design, using a mindful approach?

REFERENCES

BOTTANI, N., & VRIGNAUD, P. (2005). *La France et les évaluations internationales*. Rapport pour l'haute Conseil de l'évaluation de l'école. Paris : Ministère de l'Éducation Nationale.

17 We had the possibility to access a copy of this letter that is part of collected data.

18 Mercosur is an economic and political bloc comprising Argentina, Brazil, Paraguay, Uruguay, and Venezuela. Mercosur was created in 1991 when Argentina, Brazil, Paraguay, and Uruguay signed the Treaty of Asuncion , an accord calling for the "free movement of goods, services, and factors of production between countries." Mercosur's one-year suspension of Paraguay in 2012 and indefinite suspension of Venezuela in 2016 for violating its rules on democracy have revealed fractures within the group.

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

DAVIES, H. T. O., NUTLEY, S. M., & SMITH, P. C. (2000). *What works?: evidence-based policy and practice in public services.* Bristol: Policy Press. doi: https://doi.org/10.1111/1467-9302.00144

DESROSIERES, A. (2014). *Prouver et Gouverner*. Paris : La Découverte.

EL MASRI Y. H., BAIRD J., & GRAESSER A. (2016). Language effects in international testing: the case of PISA 2006 science items, Assessment. *Education: Principles, Policy & Practice, 23*(4), 427-455. doi: 10.1080/0969594X.2016.121832

FELOUZIS, G., & CHARMILLOT S. (2012). *Les Enquêtes PISA*. Paris : Puf.

FOUCAULT, M. (2004). *Sécurité, territoire, population. Cours au Collège de France (1977-1978)*. Paris : Seuil.

GRANOVETTER, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology, 78*(6), 1360-1380. doi: https://doi.org/10.1086/225469

GRENET, J. (2008, February 8). *PISA : une enquête bancale*? . *La vie des idées*. Retrieved from http://www.laviedesidees.fr/PISA-une-enquete-bancale.html

HEAD, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration, 67*(1), 1-11. doi: https://doi.org/10.1111/j.1467-8500.2007.00564.x

JERRIM, J. (2016). PISA 2012: *How do results for the paper and computer tests compare*? *Department of Quantitative Social Science Working Paper* No. 16-02, February 2016, (16), 1-38.

KLEIN, R. (2011). Uma re-análise dos resultados do PISA: problemas de comparabilidade. *Ensaio: Avaliação e Políticas Públicas Em Educação*, *19*(73), 717-768. doi: https://doi.org/10.1590/S0104-40362011000500002

KOMATSU, H., & RAPPLEYE, J. (2017). Did the shift to computer-based testing in PISA 2015 affect reading scores? A View from East Asia. *Compare: A Journal of Comparative and International Education*. doi: https://doi.org/10.1080/03057925.2017.1309864

LASCOUMES, P. (2004). Présentation - La gouvernementalité : de la critique de l ' État aux technologies du pouvoir. *Le Portique, Revue de Philosophie et de Sciences Humaines, 14*(13-14), 1-15. doi: https://doi.org/https://leportique.revues.org/625

LE DONNE, N. (2015), *Inégalités sociales de compétences (scolaires) et organisation du système éducatif. Études comparées à partir des enquêtes PISA de 2000 à 2009*. (Unpublished doctoral dissertation). Science Po, Paris, France.

LINDBLAD, S., PETTERSSON, D., & POPKEWITZ, T. S. (2015). *International Comparison of School Results. A sistematic review of a research on large scale assessments in education*. Stockholm: Swedish Research Council. doi: 10.13140/RG.2.2.23176.01286

LOVELESS, T. (2014). *How well are American students learning. With the section on the PISA-Shanghai controversy, Homework and common core*. Washington, DC: Brookings Institution—Brown Center. Retrieved from http://www.brookings.edu/~/media/research/files/reports/2001/9/education/09education.pdf

MEC/INEP. (2013). *Relatório nacional PISA 2012: Resultados brasileiros*. Brasilia. Doi: https://doi.org/10.1787/9789264207486-11-de

MEC/INEP. (2016). *Brasil no PISA 2015: análises e reflexões sobre o desempenho dos estudantes brasileiros*. Brasilia. Retrieved from http://download.inep.gov.br/acoes_internacionais/pisa/resultados/2015/pisa2015_completo_final_baixa.pdf

MONS, N. (2007). *Les nouvelles politiques éducatives*. Paris : Puf.

MORGAN, C. (2011). Constructing the OECD Programme for International Student Assessment. In M. Pereya (Ed.), *PISA under examination: Changing Knowledge, Changing Tests, and Changing Schools* (pp. 47-59). Rotterdam: Sense Publisher. doi: https://doi.org/10.1007/978-94-6091-740-0_4

NORMAND, R. (2015). L'enquête PISA : un objet-frontière des politiques éducatives. *Administration & Éducation*, *145*(1), 121-126. Retrieved from https://www.cairn.info/revue-administration-et-education-2015-1-page-121.htm.

NÓVOA, A., & YARIV-MASHAL, T. (2003). Comparative Research In Education: A mode of governance or a historical journey? *Comparative Education, 93*(4), 423-438.

OECD (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD Edition.

OECD (2013). *PISA 2012 Results: Excellence through Equity Giving Every Student the Chance to Succeed Volume II*. Paris: OECD Edition.

OECD (2014). *PISA 2012 Technical Report*. Paris: OECD Edition.

OECD (2016a). *PISA 2015 Technical Report*. Paris: OECD Edition.

OECD (2016b). *PISA 2015 Results (Volume I): Excellence and Equity in Education.* Paris: OECD Edition.

PETTERSSON, D. (2008). *International knowledge assessments: an element of national educational steering*. (Doctoral dissertation) Uppsala: Acta Universitatis Upsaliensis. Uppsala Studies in Education No 120.

PETTERSSON, D. (2014). Three Narratives: National Interpretations of PISA. *Knowledge Cultures, 2*(4), 172-191.

POLANYI, K. (1977). The Economy Embedded in Society. In K. Polany (Ed.), *The Livelihood of Man* (pp. 47-56). New York: Academic Press.

PRAIS, S. J. (2003). Cautions on OECD'S Recent Educational Survey (PISA). *Oxford Review of Education*, *29*(2), 139-163. doi: https://doi.org/10.1080/0305498032000080657

RIVAS, A. (2015). *Latin America after PISA: Lessons Learned about Education in Seven Countries.* Buenos Aires: CIPPEC.

ROBIN, I. (2002). L'enquête PISA sur les compétences en lecture des élèves de 15 ans: Trois biais culturels en question. *VEI Enjeux*, *129*(1), 65-91.

SCHÜTZ, A. (1976). Tiresias, or our knowledge of future events. In A. SCHÜTZ (Ed.), *Tiresias, or our knowledge of future events* (pp. 277-293). Den Haag: Martinus Nijhoff.

SCHWARTZMAN, S. (2013). Uses and abuses of education assessment in Brazil. *Prospects,* 4*3*(3), 269-288. doi: https://doi.org/10.1007/s11125-013-9275-9

SCOTT, C. (2005, January). Measuring Up to the Measurement Problem The role of statistics. *Evidence- based policy-making*, London School of Economics. Research in Economics, United Kingdom. Retrieved from https://www.paris21.org/sites/default/files/MUMPS-full.pdf

SEGONE, M., & PRON, N. (2008). *Bridging the gap. The role of Monitoring & Evaluation in Evidence-based policy making*. UNICEF.

STRASSHEIM, H., & KETTUNEN, P. (2014). When does evidence-based policy turn into policy-based evidence configurations, contexts and mechanisms. *Evidence and Policy, 10*(2), 259-277. doi: https://doi.org/10.1332/174426514X13990433991320

TURNER, R., & ADAMS, R. J. (2007). The Programme for International Student Assessment: An Overview. *Journal of Applied Measurement*, *8*(3), 237-248.

*

Received: September 25, 2018

Final version received: October 22, 2018

Accepted: October 25, 2018

Published online: October 31, 2018