

THE “THIN DESCRIPTIONS” OF THE SECONDARY ANALYSES OF PISA

RADHIKA GORUR¹

ABSTRACT: The heavy hammer methods of OECD and PISA in influencing policy through the rankings and through its policy advice are well documented. This speculative paper explores the more subtle and perhaps deeper implications of the development of the PISA database, and of the secondary analysis that is performed using this database. Speculating with concepts from Science and Technology Studies, this paper suggests that PISA deflates “ontologically luxuriant objects” into “ontologically impoverished objects” through standardization and simplification. Freed from their moorings and translated into inscriptions, these ontologically impoverished objects are promiscuous, freely combining with other such objects across spaces and times in different ways to produce lessons for policy and practice. In this paper, I suggest that, while these promiscuous relations may produce mathematically defensible assertions, such findings may be ontologically absurd. Using data from interviews with measurement and policy experts, as well as published secondary analyses, this paper ventures some speculative ideas about how we might understand the PISA database and the use of this database in secondary analysis. The paper argues that secondary analysis is not merely a mathematical or technical exercise but a sociotechnical one, and that, given its influence and reach, it attempts to open up the black boxes of the PISA database and the practices of secondary analysis, and make them available for wider sociological and philosophical examination and critique.

Keywords: Large-scale databases. Secondary analysis. Science and Technology Studies. PISA.

AS “DESCRIÇÕES FINAS” DAS ANÁLISES SECUNDÁRIAS DO PISA

RESUMO: Os métodos da OCDE e do PISA para influenciar a política por meio dos *rankings* e do aconselhamento político estão bem documentados. Este artigo é especulativo e explora as implicações mais sutis e talvez mais profundas da evolução da base de dados do PISA, e da análise secundária que é realizada utilizando-a. Com base em conceitos dos *Science and Technology Studies*, este artigo sugere que o PISA reduz “objetos ontologicamente luxuriantes” em “objetos ontologicamente empobrecidos” por meio da padronização e simplificação. Libertos das suas amarras e traduzidos em

¹Deakin University – Australia. E-mail: radhika.gorur@deakin.edu.au
DOI: 10.1590/ES0101-73302016166211

inscrições, esses objetos ontologicamente empobrecidos são promíscuos, combinando-se livremente, de diferentes formas, com outros tantos objetos, através de espaços e tempos, tendo em vista a produção de lições para as políticas e as práticas. Neste artigo, sugiro que, embora essas relações promíscuas possam produzir afirmações matematicamente defensáveis, esses resultados podem ser ontologicamente um absurdo. Utilizando dados de entrevistas com especialistas da avaliação e com políticos, bem como análises secundárias publicadas, este artigo introduz algumas ideias sobre como podemos compreender o banco de dados do PISA e o seu uso em análises secundárias. O artigo argumenta que a análise secundária não é um exercício meramente matemático ou técnico, mas sociotécnico, e que, dada a sua influência e o seu alcance, intenta abrir as caixas negras do banco de dados do PISA e as práticas das análises secundárias, disponibilizando-as para uma análise e uma crítica sociológica e filosófica mais ampla.

Palavras-chave: Bases de dados de larga escala. Análise secundária. Estudos Sociais em Ciência e Tecnologia. PISA.

LES “DESCRIPTIONS FINES” DES ANALYSES SECONDAIRES DES DONNÉES PISA

RÉSUMÉ: Les méthodes lourds de l’OCDE et de PISA pour influencer la politique à travers les ‘rankings’ et les recommandations politiques sont bien documentés. Cet article est spéculatif et explore les implications plus subtiles et peut-être plus profondes de l’évolution de la base de données de PISA, et l’analyse secondaire qui est effectuée en utilisant cette base de données. En spéculant de concepts des *Science and Technology Studies*, cet article suggère que l’enquête PISA vide “des objets ontologiquement luxuriantes» dans des «objets ontologiquement pauvres» grâce à la standardisation et la simplification. Libérés de ses amarres et traduites en des inscriptions, ces objets ontologiquement pauvres sont sujets à la promiscuité, en s’associant librement, de différentes manières, avec autant d’objets à travers l’espace et le temps, afin de produire des leçons pour la politique et des pratiques. Dans cet article, je suggère que, bien que ces relations de promiscuité puissent produire des assertions mathématiquement défendables, ces résultats peuvent être ontologiquement un absurde. En utilisant des données à partir d’entretiens avec des experts de l’évaluation et de la politique et des analyses secondaires publiées, cet article introduit des idées spéculatives sur la façon dont nous pouvons comprendre la base de données du PISA et son utilisation dans des analyses secondaires. Cet article soutient que l’analyse secondaire n’est pas un exercice purement mathématique ou technique, mais sociotechnique, et que, compte tenu de son influence et de la portée, a l’intention d’ouvrir les boîtes noires de la banque de données PISA et des pratiques des analyses secondaires, en les fournissant pour une analyse et une critique, plus large, sociologique et philosophique.

Mots-clés: Bases de données à large échelle. Analyse secondaire. Sciences, Technologies et Société. PISA.

A few years ago, while doing research on contemporary practices of evidence-based policy in education, I interviewed a number of Programme for International Student Assessment (PISA) experts about the origins, development, and influence of PISA. Many of the interviewees said that countries focused too much on superficial information such as the rankings, which were really not very useful for making policy decisions, and that too little attention was given to the wealth of information that could be got from secondary analysis of the PISA database. They felt disappointed that, even though the PISA database was freely available, it was not being adequately exploited to yield important and useful understandings. One PISA expert said:

[T]he OECD's view is 'we collect this data, we provide an initial report, and each country will provide a country report on each country, we make the data available for secondary researchers to analyse it, but my sense is that there haven't been enough people to actually do that secondary analysis. Personally I think there are more interesting stories you can get or benefits you can get from looking at the data more closely. There are some attempts – and you get to hear them at some of the IEA [International Association for the Evaluation of Educational Achievement] conferences – but the big reports can only tell limited stories. (PISA analyst, interview transcript, 2008)

In recent years, there has been a lot more emphasis on secondary analysis. The Organisation for Economic Co-operation and Development (OECD) has instituted the Thomas J Alexander Fellowship¹ to encourage scholars to use the PISA database for secondary analysis. The OECD itself produces thematic reports and other documents based on its own secondary analysis of the database. There are now a number of researchers rummaging through the PISA database and producing papers with such titles as *Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 Science and School Socio-economic Composition and Student Outcomes in Australia: Implications for Educational Policy*. Secondary analysts of PISA are publishing in a range of journals including *Multivariate Behavioural Research*; *Studies in Educational Evaluation*; *International Journal of Science Education*; and *Comparative Education Review*.

Interest in secondary analysis of large-scale data is not unique to the OECD and PISA. As large data sets have accumulated, there is an increased appetite, globally, for exploiting them through secondary analysis, which involves reanalysis of existing data sets to answer new questions (GLASS, 1976), or answering old questions with new methodologies and theories using existing data. In the UK, the training of doctoral students and early career researchers in secondary analysis became a national priority (ESRC, 2011). The American Educational

Research Association runs annual institutes, where about 500 researchers have been trained in the use of the large data sets, supported by grants from several federal agencies (AERA, 2012).

The practice of secondary analysis of large databases in education is thus becoming well established. There are, today, a recognisable body of scholarship; a set of practices with which those in the field are expected to be familiar; courses in how to do secondary analysis being taught; journals in which to publish; and conferences where scholars can exchange ideas, share their findings, critique each other, and advance the field. It is also becoming an industry, as various think tanks and for-profit as well as not-for-profit organizations are commissioned to produce these analyses.

Advocates of secondary analysis argue that it allows temporally and geographically distributed and theoretically and methodologically diverse researchers, and researchers with modest funds, to produce findings of significance and for targeted and specific purposes (SMITH, 2006). Critics cite methodological and conceptual challenges in secondary analysis, and warn against false conclusions. Rutkowski *et al.* (2010), for example, provide a comprehensive account of the pitfalls that may confront the unwary or inexperienced in analyzing the large databases of PISA, Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literacy Study (PIRLS), and offer pointers on how to use these databases in a technically defensible way. There are debates – sometimes bitter – about the validity of a technique or the relative merits of different approaches. These discussions, however, remain in the technical realm, raised in journals for statisticians and quantitative analysts. Such ‘insider’ debates are somewhat limited; they lack the perspective that those outside the paradigm might offer.

Drawing from the interdisciplinary field of Science Studies (also called Science and Technology Studies, or STS), I want to explore this epistemic practice as an outsider to statistics and secondary analysis. STS is a relatively new, interdisciplinary field, the origins of which are often traced to Kuhn’s seminal work *The Structure of Scientific Revolutions* published in 1962. It began as a new approach to historical and social studies of science, in which scientific facts were studied not as what objectively corresponded with nature, but as constructed, and based on social and institutional conditioning and a network of practices that supported particular epistemic cultures. Latour and Woolgar’s (1979) seminal *Laboratory Life*, for example, traced the day-to-day practices and work of scientists in the Salk Laboratories at San Diego to produce an account of scientific practice that those in the Laboratory found somewhat disconcerting and even unrecognisable at times. However, the Head of Salk Laboratory wrote in his introduction to *Laboratory Life*:

Whatever objection may be raised about the details and by the author’s [Latour’s] arguments, I am now convinced that this kind of direct examination of scientists at work should be ex-

tended and should be encouraged by scientists themselves in our own best interest, and in the best interest of society.... (Jonas Salk, in the Introduction to *Laboratory Life*, LATOUR; WOOLGAR, 1979, p. 13)

In this speculative paper, I want to move the internal debates about secondary analysis into a wider arena, where they can be examined as practical and philosophical matters. STS offers concepts that are well suited for such an enterprise. It is not my intention to debunk these analyses or challenge them on technical grounds. Rather, viewing PISA as a sociotechnical enterprise, I want to examine the ontological status of the database itself and its consequences for secondary analysis.

Three main arguments are offered in this paper:

1. there is an inherent handicap engendered by the distant gaze of PISA's global comparisons which favor structural analyses and the search for principles and limit more useful and meaningful understandings;
2. the objects that PISA generates to populate its database are so abstract and so flattened in translation that they cannot be reanimated into connecting with their original selves in all their complexity and diversity; i.e., the "circulation of reference" crucial to the practice of good science (LATOUR, 1999) is compromised; and
3. these flattened, abstract objects are so removed from the contexts of their production that they become mobile and promiscuous, traveling across times and spaces speedily and combining freely and without restraint with other similarly displaced objects, to produce knowledge that may be mathematically defensible but perhaps ontologically absurd.

Like the scientists in Salk Laboratory, secondary analysts might find this account of their practices strange or disturbing and perhaps even inaccurate. Even so, it could provide a way to think differently about these practices. Embarking on this unusual and risky adventure, I have adopted the somewhat whimsical style of STS writing and analysis, exemplified by Latour and very apparent in the writing of Serres, who predated STS as a discipline, to persuade secondary analysts to give up, momentarily, their familiar understandings of their own practices and accompany a stranger on her journey into their world.

Although this paper is a philosophical thought experiment, it arises from a research study that included an analysis of approximately a hundred published journal articles and reports based on secondary analysis of PISA, and semi-structured interviews with 30 experts: secondary analysts of PISA, measurement experts, psychometricians and statisticians at the OECD, Education Testing Services (ETS), the Australian Council for Educational Research (ACER), and in

various universities, as well as policy officials who used such analyses. The study focused on the US, Australia and New Zealand.

Serres, Latour and the PISA database

My theoretical speculations are developed from concepts elaborated by Latour, particularly in his (2009) analysis and explanation of the philosophy of Tarde on the practices of quantification, and in his (1986) paper *Visualistaion and Cognition: Drawing Things Together*. Explaining Tarde's views, Latour says that natural scientists are hampered by the distance between themselves and the objects and "societies" they seek to study. For Tarde, explains Latour, stars and bacteria and other objects of scientific interest were all "societies" (or, in actor-network theory parlance, "assemblages"). Whether it is the astronomer gazing at the skies or the biologist peering through a microscope, the numbers in the natural world are so impossibly large that natural scientists are forced to think in terms of structures and aggregates and develop generalized laws and principles. This is the handicap of the natural scientists – they are forced to neglect the individual in preference for the group or collective, in the process accepting a distinction between the two, which is not appropriate, Latour argues, particularly in the social sciences:

The distinction is an artifact of distance, of where the observer is placed and of the number of entities they are considering at once. The gap between overall structure and underlying components is the symptom of a lack of information: the elements are too numerous, their exact whereabouts are unknown, there exist too many hiatus in their trajectories, and the ways in which they intermingle has not been grasped. (LATOURE, 2009, p. 148)

The social scientist, on the other hand, deals with smaller numbers and does not need to sacrifice the "individual" for the "society" — individuals and societies can be studied at once, together, as co-constituents. This is appropriate and necessary "individuals" are made up of their society — they reflect and respond to the society of which they are part, and "society" is constituted by individuals — an idea neatly expressed as the hyphenated "actor-network" in actor-network theory.

One of Serres' central points with regard to quantification is that we understand very little about individual behavior from the rules and principles we might derive from studies of structures and systems as a whole. On the other hand, he believes that by studying individuals closely, and by accumulating many such accounts, it is possible to develop "types" or tentative

principles of behavior; but, because individuals and societies are always interacting and influencing each other, these “types” and principles are constantly co-evolving. These “rules” or “types” created by the aggregations of individual accounts do not supersede or sit above the individual cases, dictating accounts of individual behavior.

The second point he makes is about the influence of our research instruments and constructs on our findings and understandings. “Society” (or structure) is created by statistics, and the “societies” created by statistics change as statistical knowledge grows and changes.

It is the distorted idea that natural science is “real” science and social science is not adequately rigorous, argues Latour, that the very methods that the natural scientist is forced to make do with to overcome the handicap of distance and overly large numbers have come to be imitated and valued in social science. Porter also argues that the “thick descriptions” of the ethnographer do not gain as much trust and power as the “thin descriptions” of those who deal with numbers:

Enthusiasts of ethnography might ask, Why, in a thick world, do economists stride with heads high through the corridors of power, while cultural historians pass along their possibly profound insights to one another? Why, in the world of business and administration, are lengthy reports with all their uncertainties circulated among underlings, while the “executive summary,” purged of ambiguity and detail, goes to the people at the top? Thinness is, if not the natural state of things, an appealing modern project. It beguiles us with its terse, muscular economy. (PORTER, 2012b, p. 212)

Large-scale international comparative assessments such as PISA, whose surveys include about 70, mostly middle- and high-income nations, are rather like the astronomer’s gaze at distant stars and galaxies. It has to make do with abstract, standardized “classes” of objects at the cost of the luxuriant nuance of individual actors that a close-up look provides. Nor can they see how various elements interact with each other or how they travel and intermingle.

Two sets of questions arise from these understandings. First, what kinds of objects are “seen” and recorded and back by PISA from its forays into the far corners of the globe to hold in its central database?

The second set of questions, which form the crux of my speculation, concern the practices of secondary analysis of such data. When secondary analysts sit at their computer and open the files containing these abstract, aggregated and mask-like objects, are they looking “up close” or from a distance? The PISA database contains a mix of raw and aggregated data. What kinds of

distortions and magnifications appear, and with what consequence? Does the illusory nearness of the data in the database encourage a misunderstanding that the objects themselves, rather than impoverished, distant images of these objects, are being apprehended?

The PISA database

To address the first set of questions: “What kinds of objects does the PISA database hold? What is their ontological status?” we must make a brief foray into some of the methods of PISA and examine the database itself: the objects that comprise it, as well as the processes by which they end up in the database.

Flattened objects

PISA aims to examine the readiness for life of 15-year-old students, based on their ability to apply what they know. The surveys comprise three main components — the tests of reading, mathematical and scientific literacy; the student background survey that aims to measure “advantage”; and the school background survey that is used to create descriptions of schools and schooling systems.² Each of these components requires relentless translation and simplification in order to work across a wide range of systems and cultures. For example, individual test items are first passed through the sieve of a framework that ensures they are suitable for the purpose of eliciting students’ ability to apply what they learn. The items are then tested across a range of contexts to ensure they behave in the same way across cultures.

The items are then categorized according to difficulty level. The difficulty level is calibrated such that the difference between one level of difficulty and the next is standardized and even. This is an interesting notion, since the difficulty experienced by different students with the same problem can hardly be uniform. In PISA, “difficulty” is detached from students and attached to the test item, by standardizing the student based on ability to answer a particular question. The questions can then be defined in “logits” by difficulty level. This abstraction ensures that replacing a question by another of the same difficulty level has no effect on the outcome of students. The test is thus made independent of the actual test items — whether a particular test item appears in the test or not makes no difference to student outcome. As a result of these detachments, student scores can be made predictable using such techniques as Item Response Theory. So, paradoxically, student’s performance levels only become calculable when students and test items are detached from each other (GORUR, 2011).

Students themselves are abstractions in the PISA database. As samples, they represent the larger group of 15-year-olds in the system.³ Students are described in the sample on the basis of some universal attributes such as sex, age, and socioeconomic advantage so that students in Germany or Australia or Poland or China or Turkey are all indistinguishable from each other — or at least they can be described in the same terms. To overcome the challenge of the time, it would take for a single student to complete a test with enough items to make it valid, a single “test” is distributed between several students. Thus a “student” in the PISA database is far removed from an actual student who completes the surveys in real a classroom.

The PISA laboratory thus processes ontologically complex entities into ontologically impoverished ones to facilitate large-scale commensuration and calculation. Parsimonious abstractions replace the luxuriant ontologies of children, schools, families, communities, and nations. Similarly, households, families, and schools come to be defined in standardized terms. PISA surveys go out into the world every three years and bring back more and more impoverished, flattened objects to store in the database.

In itself, this flattening of objects is not problematic — it is inherent in all processes of research. Where things start to become problematic with the PISA database, I venture, is in the difficulty of tracing back from the flattened objects to the original three-dimensional (3D), real-life actors they represent. The chain of reference that should facilitate the translations backwards and forwards (LATOUR, 1999) appears to be broken in the case of PISA. Once the various actors have been translated into their impoverished forms and have taken up residence in the PISA database, the possibilities of their reanimation into full-fledged, complex objects appear diminished. The impoverished selves appear to replace their luxuriant, complex selves, rather than merely stand in for them temporarily. In part, I suggest, this is because not only are actors detached from their contexts, but the contexts themselves are standardized and universalized — described parsimoniously in terms of a few universal details.

Mobile inscriptions

A great benefit of these standardized, flattened objects is that they can now be translated into inscriptions. Latour, in his studies of scientific practice (e.g., LATOUR, 1987; LATOUR; WOOLGAR, 1979), is struck by the dependence of scientists on inscriptions. He talks of how quickly squealing, bloody lab rats are abandoned in favor of smears on slides under microscopes, which are in turn translated into readings on a table. For PISA scientists as well, it is a great relief to move from 15-year-olds and all the complexities of school and national politics and the diversity of classrooms and teaching styles and values

and the host of issues that mediate student performance, to the inscriptions that replace them. Once these translations occur, the multitude of students and their complexities can be exchanged for digitally recorded numbers that can easily be transported from the distant corners of the world to the secure “center of calculation” in Paris, where all manner of manipulations can be performed without distraction from the “real world.” The advantages of moving from 3D objects to inscriptions are manifold:

Scientists start seeing something once they stop looking at nature and look exclusively and obsessively at prints and flat inscriptions. In the debates around perception, what is always forgotten is this simple drift from watching confusing three-dimensional objects, to inspecting two-dimensional images which have been made less confusing.... Lynch, like all laboratory observers, has been struck by the extraordinary obsession of scientists with papers, prints, diagrams, archives, abstracts and curves on graph paper. No matter what they talk about, they start talking with some degree of confidence and being believed by colleagues, only once they point at simple geometrized two-dimensional shapes. The “objects” are discarded or often absent from laboratories. Bleeding and screaming rats are quickly dispatched. What is extracted from them is a tiny set of figures. This extraction ... is all that counts. (LATOUR, 1986, p.15-16)

Inscriptions provide ways for PISA scientists to talk in a common language. Given that PISA is nearly global, this is no small advantage. Numeric inscriptions in particular have the knack of “jumping linguistic boundaries and displacing local knowledge and native informants” with ease (CULLATHER, 2007, p.337). Numbers also hold the various actors stable across space and time. The many complex actors — students, schools systems, principals, teachers, parents, and society itself — may undergo seismic shifts in real life — but, once translated, they remain stable as inscriptions in the database. These also “domesticate” the scientists who study these actors: as Latour explains, confronted by inscriptions, researchers can no longer speak variously or subjectively about matters — they are all subjected to the force of the inscriptions.

The actors in the PISA database are thus both stable and mobile — they have become, in Latour’s terms, “immutable mobiles.” Freed from their tethers, they can travel across space and time with relatively little distortion and at little or no cost. As each round of PISA gathers these survey data, the database swells with more and more abstract entities, measured in easily translatable metrics and combinable in multiple ways with other, similarly abstracted entities.

Importantly, the numbers, figures, graphs, and tables that make up the PISA database are not mere representations of the world — they are also a presentation of the world-view in which these representations make sense. As Latour (1986; 2009) puts it, they not only tell us what to see but also how to see. The instruments that are used to generate the data render them visible in particular ways and sensible within particular framings. Such inscriptions encapsulate the world in which these entities are rendered sensible, and at the same time, describe the entities that make up that world. Secondary analysts who think they are simply using some kind of neutral and unmediated data would be mistaken — the data are teeming with Trojans in the forms of methodology, models, assumptions and world-views.

Promiscuous relations

The abstraction and mobility of the objects in the PISA database renders them uninhibited, so that they may promiscuously relate and combine with other objects that, in their less impoverished and more luxuriant forms, they might not have engaged with so readily. Students and schools and governance systems in Azerbaijan and Australia, for example, can be brought together in relations of comparison in their ontologically impoverished formats, whereas in their luxuriant “real” forms, such a comparison might have appeared absurd or have required a range of qualifiers and caveats. Stripped bare of the contexts of their production, these objects can be coupled together in myriad ways, combining and recombining in secondary analyses to describe states of affairs, provide explanations, articulate policy problems and solutions, identify “best practices”, project futures, paint utopias, and detect threats. The PISA database brings distant objects near — sitting in my study today, I can access data from all over the world in seconds. I could, if I wished, compare the correlations of particular school system features with student performance in Australia with the correlations in Japan and Canada and Mexico and Sweden, all on the computer screen. At the same time, this database also makes the nearby distant — so easy it is to use the impoverished objects and shuffle them as I please to make my claims and assertions, and so amenable are these data to my mathematical calculations, that I have no need to actually step outside my house to look at actual schools, teachers, and students, in all their complexity and ontological stubbornness, even to learn about schools in my own neighborhood.

These three attributes — flatness, mobility, and promiscuity — render the PISA database highly amenable to secondary analysis, but these very features are also the ones that can easily lead analysts astray.

Rummaging through data: making thin descriptions

With each survey, the PISA database accumulates more and more abstract, standardized entities that are combinable in multiple ways, like Lego pieces, with other, similarly abstracted entities. Most crucially, the PISA database has created “optical consistency” (LATOURE, 1986). Whether they are institutions, societies, or individuals, all the actors in the PISA database, translated into 2D inscriptions, are rendered in the same form, and can be seen in the same way wherever one is standing in relation to it. In other words, whether a researcher sitting in Paris is looking at data on French students or Indigenous students in Australia, the perspective does not change. For the secondary analyst, this affords unlimited possibilities, just as optical consistency affords an artist a multitude of possibilities:

[R]eal objects can be drawn inseparated pieces, or in exploded views, or added to the same sheet of paper at different scales, angles and perspectives. It does not matter since the “optical consistency” allows all the pieces to mix with one another. (LATOURE, 1986, p.8)

It is precisely this freedom to exaggerate parts, shift angles, and perspectives and combine indiscriminately that makes secondary analysis such a fraught process.

Indiscriminate mixing and matching

An immense investment is made in creating “optical consistency” so that systems with disparate political and social cultures, teaching practices, schools, children, and family life appear commensurate. This domestication of diversity, as well as the easy access to an abundance of data, has resulted in a range of studies in which a phenomenon is investigated using data from a large number of diverse countries.

Paradoxically, rather than anticipating that comparisons across vastly different countries may diminish the validity of data, in secondary analysis, there appears to be more faith in calculations involving data from a large number of countries. One such study is *Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs* (AFONSO; AUBYN, 2006), which appeared in the *Journal Economic Modelling*. This study is described as follows:

We address the efficiency of expenditure in education provision by comparing the output (PISA results) from the educa-

tional system of 25, mostly OECD, countries with resources employed (teachers per student, time spent at school). We estimate a semi-parametric model of the education production process using a two-stage procedure. By regressing data envelopment analysis output scores on non-discretionary variables, both using Tobit and a single and double bootstrap procedure, we show that inefficiency is strongly related to GDP per head and adult educational attainment. (AFONSO; AUBYN, 2006, p. 476)

Using data from countries as diverse as Finland, Korea, and Indonesia, and based on the “output” (performance in PISA) and “resources employed” (“number of teachers per student” and “time spent in school”), they first “derived a theoretical production frontier for education.” Countries’ efficiency scores were calculated as the distance of their performance from this frontier. The first part of the paper “determines the output efficiency score for each country, using the mathematical programming approach known as DEA [Data Envelopment Analysis], relating education inputs to outputs” and taking the nation as the “decision-making unit” (DMU) (AFONSO; AUBYN, 2006, p. 478). Then two “environmental factors” — parents’ education and the students’ wealth (using the nation’s GDP as a proxy for wealth) — are factored in to temper the efficiency calculations. The authors take all the required steps to defend their modelling and to make the calculations transparent. The authors report:

Results from the first-stage imply that inefficiencies may be quite high. On average and as a conservative estimate, countries could have increased their results by 11.6% using the same resources, with a country like Indonesia displaying a waste of 44.7%. (AFONSO; AUBYN, 2006, p. 489)

However, they conclude, when the effects of the “environmental factors” of wealth and parents’ education are factored in, the efficiency scores and rankings of nations change substantially.

What I find striking in such studies is the care taken to explain, make transparent and defend every mathematical move, whilst making very little effort to look outside this “numbers world” to explore whether “inputs” such as national per capita expenditure on education and the number of teachers were in fact the right or sufficient “input” factors and whether reckoning them at a national level was sensible. Modeling the calculations of resources on the basis of “number of teachers per student” makes little sense if one does not factor in the structure of schools. In most countries, the distribution of teachers is not even – there may be a concentration of teachers in urban areas and in private

schools. National averages mask these very consequential differences. Whether there are para-professionals such as teachers' aides or parent volunteers in schools, whether schools are inclusive or exclusive, and the kind of pedagogies employed are all factors that would have very different outcomes for the same teacher-pupil ratio. Culture plays an important part — Asian schools often have large class sizes, but the students' disciplined behavior and deference for the teacher means that time taken up to maintain discipline is much less than in an inner city school in a big US city.

Comparing per-capita expenditure on students is equally fraught when viewed in aggregate. Within Australia, for example, where nearly 40% of the students study in private schools, and there is a complex mix of federal, state, Church, and private sources of funding, can "per capita" expenditure on students capture this complexity? What about the expenditure of parents on private tutoring and coaching, as, for example, in Korea? In the simplified world of the PISA database, these questions are dissolved. Using GDP as a proxy for student wealth is similarly confounding, since wealth is almost invariably very unevenly distributed within countries.

Analysts who take the impoverished objects of the PISA database as starting points for research may be seeing their objects of study, the data in the PISA database, just inches away, on their computer screens, but they might as well have been as far away as Latour's stargazers, for all the difference it makes to their ability to grasp reality. Ignoring within-country differences and taking the aggregated values, and using data which are blind to the dynamic inter-relationships between actors to perform careful, sophisticated calculations may lead to mathematically defensible methodologies, but the conclusions may be ontologically nonsensical.

Broken chain of reference

One set of difficulties in secondary analysis ensues from a break in the "chain of reference" (LATOURET, 1999), i.e., objects in the PISA database cannot always be traced back to their original, pre-translated selves. This happens for several reasons. Most of the data in the PISA database are gained from surveys. Each survey is based on a set of assumptions that are not always right, or not right across varied contexts, creating a disjuncture between the object of interest in real-life, and the object that represents it in the data. While these assumptions maybe contested when debating PISA's methodology the secondary analyst deals with objects that mediated by the assumptions that are possibly no longer visible, as in the example below.

In the era of devolved responsibility and market-based logic in governance, one policy idea promoted by the OECD is school-based autonomy. This topic has been studied extensively through secondary analysis, in which attempts have been made to link school autonomy with student performance. One such study is *The Effect of School Autonomy and School Internal Decentralization on Students' Reading Literacy* by Maslowski, Schreerens, and Luyten (2007). In their literature review, they examine the various theories that underlie different studies that have examined this question, and find that some of the studies contradicted each other. They suggest that some of the strong, positive associations found in earlier studies, which had led to strong policy advocacy for increased autonomy in schools, may have been flawed. They attribute these flaws to the modelling that underpinned the calculations. Their own research included student and school data from PISA 2000 from 28 OECD systems, involving 5,269 schools and 137,526 students. They concluded that:

... schools with autonomy on personnel management issues have, on average, higher mean reading literacy scores than schools with lesser autonomy in this domain. For autonomy on financial resources, student policies, and curriculum, no significant effects on students' reading literacy were found. (MASLOWSKI; SCHREERENS; LUYTEN, 2007, p. 314)

But the wide variation in the results of studies of school autonomy effects might well arise from a flaw that goes much deeper, involving the objects in the database itself, as one interviewee explained:

Well, the basis of school autonomy is Principals' answers to the questionnaire in which they are asked who has more or less influence on different things. Our structure [in the US], which is schools, then school districts, then states – is not represented in the questions. So what looks like local school control is really school district level control. So we don't have what they are talking about. School principals, I don't think, feel like they have a lot of autonomy over things like their budget and not even the hiring and firing of teachers, much, and I think it probably varies a lot within country because of the way the districts relate to school and the size of districts and the number of schools in a district. So we don't think that it reflects the US well, and we think that probably other countries also – that the question doesn't fit. (Policy expert, US Government, interview transcript, 2013)

In the translation from world to survey, a distortion is built in. Each subsequent translation simply magnifies the distance between the world and the object by which it is represented in the PISA database.

Any problem you have with large-scale quantitative data necessarily transfers to secondary analysis. There is a transferability problem, which is when you use questions that are not suitable to your purpose – you're forcing ... to call 'autonomy' when you are actually measuring something else. So that transferability – forcing a construct to become something else – that negotiation – that could be a problem. (Policy official, US Government, interview transcript, 2013)

Distortions may be caused by PISA's methodologies as well:

There's one thing that kind of makes me worry a little bit, maybe – that we force the distribution of the indices to have an average of zero and a standard deviation of 1 in OECD countries - that is kind of our practice, but by doing that maybe the latent construct has maybe a very small variability just like that [gesturing to show a small gap between thumb and forefinger] but we force it to have a standard deviation of 1 across OECD countries on average. So you may have a very, very small index that is actually not saying anything special, but you're forcing it to say something - it becomes a thing. One of them is school leadership, for example. Personally, I don't think that index is very useful but because we forcing it to have variability where there is none – no meaningful variability in that underlying construct.... So one worry I would have is that people don't know that we are forcing there to be variability when in fact there might not be. (PISA expert at OECD, interview transcript, 2013)

To ensure good science, one has to be able to travel back and forth along the chain of reference, and not merely engage with the translations and simplifications:

[i]f you are minimally responsible [as a secondary analyst], then you say, okay, school leadership, we want that, they think that's good, let's relate it to performance or reading literacy or whatever, then you find nothing, then okay, we just think a little bit, taking it on face value the index may give you a problem but then but if you do the small step that comes afterwards, you're kind of okay – you're guarded against that possibility of being reckless. (PISA expert at OECD, interview transcript, 2013)

OECD may delegate recklessness to the secondary analyst, but that is a bit disingenuous, because the OECD has strongly promoted a focus on leadership and school autonomy.

Spatial and temporal distortions

The “optical consistency” of the PISA database facilitates the collating together of data without needing to be too fastidious about scale, perspective, or time. One of the issues that my interviewees identified had to do with temporal disjunctures. PISA collects data about 15-year-olds, who are usually in Junior High or High School in the US. It also collects data about the schools these 15-year-olds attend, through the Principals’ survey. The bulk of student learning would occur in a different school structure — an elementary P-6 or P-8 school, for example. Yet PISA links students’ performance at age 15 to the institution they have only attended for a small fraction of their school lives to make inferences about the kinds of systems and structures that might enhance learning, as one interviewee explained:

...PISA is an assessment of everything you’ve learned in the first 15 years of your life, right, and learned it inside of school and outside of school and so if you wanted to have independent variables that affected your performance, you’d want to measure them over those 15 years – it’s not just this year, and PISA is just measuring this year. And in the United States, that means - because we collected PISA data in October – so that means we are collecting all this information about your school in October (schools start in September) as if that’s going to explain all 15 years of your learning... (Interview transcript, policy official, US Government, 2013)

It is a great temptation for the secondary analyst to piece things together into patterns and pictures because every piece appears to fit every other piece. There are no jutting out, pointy bits that warn analysts that they could be on the wrong track.

Trapped in the data world

One of the most impactful findings from PISA has been with regard to teacher quality and its impact on student performance. Teacher effects are surmised on the basis of isolating other factors statistically, using regression analysis. However, this kind of “isolation” of the “variables” is only possible statistically — in actual classrooms, the variables are present simultaneously and interact with each other in complex ways:

So we say, okay, teachers matter a lot [all else being equal] – after accounting for or controlling for socio-economic status – but does that really happen – is there something as such as a

school where all students are equal, or can you compare a very low performing public school which serves a very disadvantaged population with a private school serving the rich community and [say that] *IF* they had the same socio-economic kids they would have comparable results... (PISA Expert at OECD, interview transcript, 2013)

The PISA database appears to offer so complete and attractive a world that it sometimes encourages researchers to ignore what is not in the PISA database:

[T]o me ... the largest problem is that there is so much unmeasured about countries that matter to kids' achievement... there is all this other information that you don't have, that unmeasured, about countries that may really, really matter. For instance the time – you go in here and you see that Korea doesn't spend a lot of time on education, or you'll see that Korea has really large class sizes They'll talk about 'look at these countries and how much they spend on education and then let's look at what they buy with their spending. Andreas will often compare Luxemburg with Korea... and he will say 'this is how much per student that they spending – they're both spending about the same, but what Luxemburg buys with its spending is small classes. And the way they are able to pay for small classes at that price is that they don't pay the teachers that much and they have the teachers spend a lot of time teaching. And then he'll look at Korea and say, they spend about as much per student, but what they're spending their money on is teachers – higher quality teachers and more preparation time. The way they are able to afford it is really large class sizes. And so you come out of that and say 'I want to be like Korea!' and therefore I can increase my class sizes, but I don't know – are you like Korea, that you could respond to 40 kids in a class – I don't know... that's unmeasured. So if you just went into these analyses and assumed all countries were the same, except for everything that is measured in this database, then you'd be fine! But we know that there is a lot unmeasured culturally and other ways that's outside it. (Interview transcript, US policy official, 2013)

Curiously, it appears that the world as described by PISA numbers is so compelling and convincing that analysts ignore what they know about the real world. Most people have heard about the cram-schools in Korea, in which students spend almost as much time as they do in regular school. People who have lived or visited there talk about the cult status of some famous tutors in these schools. Certainly, Andreas Schleicher, the man in charge of PISA at the OECD, could not be ignorant of these cram-schools. Yet the numbers are pre-

sented as a world complete in themselves, ignoring the realities that are before them in the real world.

Mathematically defensible but ontologically absurd?

Databases made up of ontologically impoverished objects provide a kind of surface infrastructure (as opposed to a strong foundation) on which researchers can skate with speed and efficiency, and create apparently solid science through defensible calculations. This infrastructure provides the framework within which the logic of these studies work — but the same logic may not — and frequently does not — work when translated into policy and introduced into the world.

Thin descriptions and their apparent objectivity can also draw trust and resources away from other types of research, so that it becomes harder to challenge them. As Porter (1995, p. 168) suggests, “This kind of objectivity, when reason is reduced to an algorithm, can stand in the way of truthful knowledge.” Not only are numbers trusted more but more numbers are trusted more — i.e., large-scale analyses, despite all their limitations, are valued more:

[T]he term “large-scale” suggests completeness, while ease of collection and analysis suggest that little else need be done. Both tend to crowd out other interpretations; hence understanding their limits should be of the utmost concern. (BUSCH, 2014, p.1727)

Importantly, as I have also argued elsewhere (GORUR, 2015a; 2015b; 2016), contesting these thin descriptions is important because such descriptions are also forms of intervention:

The thinness of the testing regime, as such, is not its most troubling feature. What matters above all is its capacity to thin out programs of instruction and learning, to drink up the sea. (PORTER, 2012b, p. 225)

We have come to believe so strongly in the impoverished objects of the PISA database, that instead of reanimating them to validate the assertions they make, we are impoverishing the luxuriant, real-life objects and recreating them in the image of the PISA objects (GORUR, 2016; cf. SCOTT, 1998). These objects in the PISA database are no longer content to merely represent their real counterparts — they are taking over and replacing them.

Societies, as earlier argued, are created by statistics; but, instead of using this knowledge to shape statistics to create the kinds of societies that would

produce equitable and sensible societies, we appear to be willing to allow statistics to dictate what kinds of societies we create.

As scientists, we appear to be losing our ability to challenge secondary analysts and, more generally, the producers of large numbers, except on their own terms of mathematical certainty and precision. Nor are we able to come up with more suitable alternatives:

Experts on schools are increasingly outspoken on the problems of thin indicators and can even demonstrate quantitatively some of their shortcomings. Scarcely anyone argues that numbers lack any important role for understanding the problems of schools. Designing a satisfactory measurement regime, however, is a labor of Sisyphus, especially when officials in charge may find advantage in superficiality. (PORTER, 2012b, p. 226)

By bringing these issues outside the realm of the exclusively technical, I hope to interest more researchers to join in the critique from various disciplinary perspectives and, even more importantly, to work together in the labor of Sisyphus to which Porter alludes above.

Notes

1. <http://www.oecd.org/edu/thomasjalexanderfellowship.htm>
2. Some countries choose some additional optional components, such as a parent survey.
3. The two-step sampling system used in PISA is complex and presents a problem for the less-expert secondary analyst, because it is easy to make erroneous extrapolations from PISA samples unless one is really well-versed in this system. I have not emphasized these aspects in this paper since my interest is to move away from the “technical” critique which is quite extensive among statisticians and psychometricians.

References

AFONSO, A.; AUBYN, M.S. Cross-Country Efficiency of Secondary Education Provision: A semi-parametric analysis with non-discriminatory inputs. *Economic Modelling*, v. 23, p. 476-491, 2006.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION – AERA. *Professional Opportunities and Funding*. 2012. Available from: <http://www.aera.net/ProfessionalOpportunitiesFunding/FundingOpportunities/AERAGrantsProgram/MORE/tabid/10900/Default.aspx>. Cited: May 30, 2012.

BUSCH, L. A Dozen ways to get lost in translation: inherent challenges in large-scale data sets. *International Journal of Communication*, n. 8, p. 1727-1744, 2014.

CULLATHER, N. The Foreign Policy of the Calorie. *American Historical Review* (April), v. 112, n. 2, p. 337-364, 2007.

ESRC – ECONOMIC AND SOCIAL RESEARCH COUNCIL. *Delivery Plan 2011-2015*. ESRC, 2011.

GLASS, G.V. Primary, Secondary and Meta-Analysis of Research. *Educational Researcher*, v. 5, n. 10, p. 3-8, 1976.

GORUR, R. ANT on the PISA Trail: Following the statistical pursuit of certainty. *Educational Philosophy and Theory*, v. 43, S1, p. 76-93, 2011.

_____. Assembling a sociology of numbers. In: HAMILTON, M.; MADDOX, B.; ADDEY, C. (Orgs.). *Literacy as Numbers – Researching the Politics and Practices of International Literacy Assessment*. London: Cambridge University Press, 2015a. p. 1-16.

_____. The Performative Politics of NAPLAN and My School. In: THOMPSON, G.; SELLAR, S.; LINGARD, R. (Orgs.). *National Testing and its Effects: Evidence from Australia*. London: Routledge, 2015b. p. 30-43.

_____. Seeing like PISA: A Cautionary Tale about the Performativity of International Assessments. *European Educational Research Journal, Online First*, 2016.

LATOUR, B. Visualisation and Cognition: Drawing Things Together. In: KUKLICK, H. & LONG, E. (Org.) *Knowledge and Society - Studies in the Sociology of Culture Past and Present: A Research Annual*. V. 6. Greenwich: Jai Press, 1986. p. 1-40.

_____. *Science in Action - How to Follow Scientists and Engineers through Society*. Cambridge: Harvard University Press, 1987.

_____. *Padora's Hope - Essays on the Reality of Science Studies*. Cambridge: Harvard University Press, 1999.

_____. Tarde's Idea of Quantification. In: CANDEA, M. (Org.) *The Social After Gabriel Tarde: Debates and Assessments*. London: Routledge, 2009. p. 145-162.

LATOUR, B.; WOOLGAR, S. *Laboratory Life - The Construction of Scientific Facts*. Princeton: Princeton University Press, 1979.

MASLOWSKI, R.; SCHEERENS, J.; LUYTEN, H. The Effect of School Autonomy and School Internal Decentralization on Students' Reading literacy. *School Effectiveness and School Improvement*, v. 18, n. 3, p. 303-334, 2007.

PORTER, T.M. *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton: Princeton University Press, 1995.

_____. Measuring What? *Measurement: Interdisciplinary Research and Perspectives*, v. 10, n. 3, p. 167-169, 2012a.

_____. Thin Description: Surface and Depth in Science and Science Studies. *Osiris*, v. 27, p. 209-226, 2012b.

RUTKOWSKI, L. *et al.* International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, v. 39, n. 2, p. 142-151, 2010.

SCOTT, J.C. *Seeing like a state: how some schemes to improve the human condition have failed*. New Haven and London: Yale University Press, 1998.

SMITH, E. *Using Secondary Data in Educational and Social Research*. Maidenhead, Berkshire, and New York: Open University Press - McGraw Hill Education, 2006.

Received in July 05, 2016.

Approved in August 29, 2016.