

---

# Towards a Sociology of Measurement in Education Policy

**RADHIKA GORUR**

*Victoria Institute, Victoria University, Australia*

**ABSTRACT** The Organisation for Economic Cooperation and Development (OECD) has developed impressive machinery to produce international comparative data across more than 70 systems of education and these data have come to be used extensively in policy circles around the world. In many countries, national and international comparative data are used as the bases for significant, high-stakes policy and reform decisions. This article traces how international comparability is produced, using the example of equity measurement in OECD's Programme for International Student Assessment (PISA). It focuses on the construction of the objects of comparison and traces the struggles to produce equivalence and commensurability across diverse and complex worlds. Based on conversations with a number of measurement experts who are familiar with the OECD and PISA, the article details how comparability is achieved and how it falters and fails. In performing such an analysis, this research is not concerned with 'exposing' the limitations of comparison or challenging their validity. Rather, based on the work of Steve Woolgar and other scholars, it attempts to mobilise a 'sociology of measurement' that explores the instrumentalism and performativity of the technologies of international comparisons.

## Introduction

That Germany went into 'PISA shock' at its poor performance when the Programme for International Student Assessment (PISA) 2000 results came out is well known. But perhaps less well known is the fact that Australia also was shocked to find itself 'inequitable' – it was declared 'high quality, low equity' in PISA 2000. However, happily, by 2003, Australia had shifted into the 'good quadrant' of 'high quality, high equity'.

I was excited to learn this – inequities are so hard to shift, and if Australia had made such a positive difference in a short span of time, I wanted to find out how. What, I asked PISA experts, had Australia done to make such progress? What reforms had enabled it to become a 'high equity' nation? The answer was both surprising and disappointing. The actual situation of equity in Australia's education system may not have actually changed, I heard. Rather, the shift into the 'good quadrant' was attributed to the slight variation in the way 'equity' was measured in PISA.

I was intrigued. Until then, I had never given any thought to how the abstract notion of 'equity' was measured at all. Having lived in very different countries – Nigeria, India, Oman, Australia – I knew that there is great variety in inequities, since inequities are complex assemblages of economic, social, cultural and historical factors. How did the Organisation for Economic Cooperation and Development (OECD) manage to measure levels of equity and inequity, and do so in a comparative way, across scores of diverse locations, and translate these comparisons into dots on a two-by-two matrix?

Curiosity around this question led me to a number of conversations with assessment and measurement experts with intimate knowledge of PISA. What follows is an account of the complex and protracted struggles to establish comparability and refine indicators in the pursuit of accuracy and currency in international comparative measurements of equity. Opening up this black box, I

trace the productions of comparability and understand 'how objects of comparison are constructed, how they travel and how they are modified ... how the complexity of local worlds becomes involved in producing comparability; and, of course, how comparability fails' (Niewöhner & Scheffer, 2010, p. viii). I want to convert such measurements from a routine activity to a *momentous* one, an *event* around which we might have purposeful deliberations and interdisciplinary conversations (Stengers, 2011). Rather than treat such comparisons as a matter of routine application of methodologies or focus on their accuracy or validity, I want to focus on the practical productions of the worlds that are brought into being through such measurements.

### **Sociology of Measurement**

Although international comparisons in education are not a new phenomenon, their use and influence have been greatly ramped up with the advent of PISA. Correspondingly, critique of international comparisons has also proliferated. Technical critique of comparison has involved the examination of PISA's validity and of methodology. A recent special issue of *Research in Comparative and International Education* (vol. 8, no. 3), for example, focuses on large-scale international comparisons and raises a number of issues with regard to the warrant of the assertions made. Of particular interest for the present discussion is an article in that special issue by Rutkowski and Rutkowski (2013), in which they examine the accuracy of the measurement of equity in PISA. They demonstrate, using mathematical arguments, that the Home Possessions Index, while suited to OECD countries, might not work as well in developing economies. Indeed, given PISA's widespread influence and use, there is a vast literature that critiques aspects of its methodology (for example, Hopman & Brinek, 2007). These technical critiques are focused on the accuracy of PISA's representations and descriptions. In these critiques, the world is already present, to be described and measured, hopefully with increasing accuracy, through the application of rigorous scientific methodologies.

Some sociologists have focused on PISA and other types of international comparisons as symbols and examples of the globalisation of policy ideas and as technologies of power that epitomise the neoliberal agenda (see, for example, Rizvi & Lingard, 2010; Stronach, 2010). In this critique, PISA is situated against a background – or within a context – of an increasingly borderless or transnational policy space. Here PISA and other such measurement efforts are contextualised within the spread of surveillance and accountability practices in neoliberal times. The collapse of nation states and the emerging topologies of transnational spaces and globalised ideas and practices are highlighted (Ball, 2009; Lawn & Grek, 2012).

More broadly, within and beyond the education policy literature, there is recognition of the mobilities that characterise contemporary societies – mobilities brought about by a range of new technologies, large-scale migration, and the transnationalisation of desires and fears (Appadurai, 1996, 2006). These mobilities have encouraged a reconceptualisation of what is 'social', and how we might study so fluid a concept (for example, Beck, 1992; Appadurai, 1996; Hannam et al, 2006, and also Landri & Neumann, this issue). Along with the movement of people and ideas, there is also focus on the spread of new threats such as pandemics, terrorism, and fundamentalist forms of nationalism, which are associated with the rise of uncertainty and vulnerability to unexpected and unpredictable events that cannot be easily controlled or even predicted. The recent 'global financial crisis' is an example of such a threat, where actions of distant others had significant effects on widely dispersed individuals and nations.

Callon et al (2009) make a distinction between the notion of uncertainty and that of risk. Uncertain situations are ones in which the outcome is not known. Risk, on the other hand, requires the ability to imagine a range of possible scenarios and their outcomes. Managing risk involves making decisions about which scenarios are most likely and then building strategies to contain their negative consequences and promote positive ones. An important technology in managing risk is the establishment of standards, benchmarks and other regulatory mechanisms which redistribute and institutionalise trust and responsibility, and make feasible greater accountability and transparency (Timmermans & Berg, 2003; Busch, 2011). With transnational organisations of various kinds importing, exporting and outsourcing goods and services across the globe, standardisation has proliferated and has been greatly scaled up (Lampland & Leigh Star, 2009;

Gorur, 2012). In education, this has taken the form, among other things, of standardised assessments and comparisons to aid policies aimed at gaining an upper hand in a competitive knowledge society, and preparing for the predicted threats to this ambition from the rising economic powers of Asia.

This article – and my research more broadly – is situated in the explorations of how we, as researchers, might engage with and respond to these new mobilities, particularly as they concern education policies. Following a Latourian approach (Latour, 2004, 2005), I have sought to make phenomena such as globalisation and the mobility of ideas and practices my objects of study, with a particular emphasis on the production of global spaces of commensurability and calculability. Through a body of work that has examined a variety of policy assemblages – the development of PISA (Gorur, 2011a); the marketisation of Australia's education space and its impact on evolving productions of 'equity' (Gorur, 2013); and the struggles to technicise and systematise education policy (Gorur & Koyama, 2013); and more generally, the rise of 'evidence based policy' in education (Gorur, 2008, 2011b), I have been developing a line of inquiry in education policy that may be described, to use a term coined by Steve Woolgar, as a 'sociology of measurement' (Woolgar, 1991; Derksen, 2000). In this article, I perform a similar 'sociology of measurement' to study how international comparability is affected in PISA's equity comparisons.

To make his case for a sociology of measurement, Woolgar's (1991) focuses on the use of citations as a measure of research quality, and the persistence of this practice despite a range of criticisms highlighting its flaws and inadequacies. Woolgar points out that most of the criticism is based on the imprecision of such measures, and that this type of criticism is easily dismissed by citationists. Criticisms might be noted and then brushed aside as merely technical matters, which could be fixed in due course, as techniques of measurement improved. It might even be implied that in raising technical issues, critics are being just petty and impractical. Citationists might argue that having *some* basis, even an imperfect one, is better than not having any basis at all.

Woolgar's quest is to find the kind of criticism that would invite and encourage citationists to engage with it. Noting that the institutionalised practices of measurement are 'constructive' rather than 'descriptive', he asserts that raising issues of accuracy only reinforces the 'mundane ontology' of representation that is assumed in such measurement. Using citation measures as an example of the increasing use of performance measures more generally, he argues that '[i]t is timely to ask how and why different forms of assessment gain (or lose) credibility' (Woolgar, 1991, p. 319).

[T]hese actors operate within an institutionalised set of beliefs and practices such that what others (ontological relativists) might allege to be 'profound deficiencies' are seen by them as no more than 'technical difficulties'. From the point of view of the citationists, looking through these institutional spectacles, the continued appeal to mere difficulties appears as persistent niggling about difficulties of which they are already aware. (Woolgar, 1991, p. 322)

Woolgar suggests that it is useful to think of citations (and in our case, international comparisons) as a *measurement technology*. By using the term 'technology', he draws our attention to 'instrumentalism' that guides such practices. Drawing upon studies of the production of knowledge in laboratories, he uses the double meaning of 'instrumentalism': first, the focus on practicalities – on 'making things work'; and second, the apparently neutral applications of methodologies and the use of instruments and devices. He elaborates this constitutive nature, or performativity, of measurement technologies as follows:

[C]itations certainly measure whatever they measure very well. The results can be seen to be consistent, cross-correlated and so on. The only problems arise when we suppose that these citations reflect something called quality and/or influence. Or rather, the problems arise when we suppose that they are measures of the thing we always thought was quality/influence before the citation analyses came along. (Woolgar, 1991, p. 324)

Taking an ontologically relativist stance facilitates understanding that 'the measurement of a phenomenon can have a profound effect in redefining an existing phenomenon' (p. 324). So what is at stake here are the ways in which different groups and institutions might appropriate the right to define phenomena such as 'research quality' (or, in the case of this article, 'equity') through the institution and promotion of different technologies of measurement.

What do these understandings bring to the role of the critic? Woolgar suggests:

[W]e need to move beyond the simple technicist view that citations are mainly faulty indicators. We should eschew attempts to arbitrate on the accuracy of citation analysis, in favour of a sociological analysis of its institutionalised practice. This means, in particular, that we try to anticipate the circumstances under which the application of this measurement technology to the concept of quality will result in modifications (or redefinitions) of what counts as quality. This has especially important implications as different agencies become involved in the struggle to appropriate the measurement technology for their own specific purposes. (p. 325)

This, then, is the task I am attempting here – to perform a ‘sociology of measurement’ to understand how international comparisons of equity are effected, and to understand the worlds brought about by such calculation. Taking such comparisons as *performative* rather than merely representative (Pickering, 1995; Latour, 1999), I am eschewing engagement in arbitrating whether these measures are accurate or not. Instead, I focus on the instrumentalism by which ‘equity’ comes to be reconstituted as something that can be measured and described through the particular technologies of comparison. In tracing the instrumentalism that runs right through the practices of this comparison, I highlight the constitutive role of this measurement technology.

### **Equity and Inequity in Education Policy**

Inequity has been an issue in education policies worldwide. Inequities have been variously addressed using a variety of concepts, such as poverty, multiculturalism, disadvantage, equality, social justice and inclusion, each reflecting a slightly different concern and emphasis. My interviewees explained how they grappled with the slipperiness and complexity of ‘equity’. All interviewees are anonymous, but a descriptor of their position – ‘policy advisor’ or ‘PISA expert’, for example – is used to provide some understanding of their background. The interviews spanned several related projects and occurred between 2008 and 2012. A popular conception is that of equity of opportunity:

To me equity is about meeting the necessary conditions for everyone to learn, and to make available appropriate and high quality learning for all. I mean this just as much for the high-performing students as the low-performing ones. It means having well-equipped schools with good teachers in all areas. (Senior policy advisor, Australia. Interview notes)

Equity of *outcome* might be a desired goal, but is often seen as impossible to achieve. Other valued ideals, such as providing people with ‘choice,’ might also complicate notions of equity. One former minister explained:

[I]t can be grossly inequitable to impose equal provision on everyone. Because everyone is not the same – everyone has equal rights to be themselves and that means that diversity is essential as part of the concept of equity in provision because if you haven’t got the diversity in provision then you haven’t got equitable provision because it is unfair to the very diverse student group that the provision has to be for and I don’t think there is a lot of grappling with that concept. A lot of the discussion is too shallow. (Interview transcript, former Minister of Education)

‘Equity’ may also be in opposition with the individualistic ethos of competition and maximisation of profit for self, and there is some confusion as to what ‘equity’ means in relation to democratic ideals, as the former minister went on to explain:

Look, I don’t know – it’s a term – and I don’t think the concept behind that term is stable... I think the democratic idea itself has difficulty in grappling with this notion of ‘if everyone is equal, what does it mean?’ What is the inalienable right that people have that should be equalised. And I have tried to grapple with this myself, and what I think it ultimately has to mean is the equal right of every person to live a life which achieves their values provided of course that they don’t harm others. (Interview transcript, former Minister of Education)

To complicate matters, several understandings of equity – equality of opportunity and access, positive discrimination, need-based provision, equity of outcome – circulate simultaneously in policies. Some of these are in fact philosophically opposed to each other – for example, equity as equality – sameness – goes against positive discrimination. This lack of clarity about what equity is gets in the way of making equity measurable and internationally comparable. To make equity

calculable, the fuzzy notion of equity must be given a much more stable and recognisable definition. As a policy-focused agency, the OECD must go for messages that are clear and easy to express and understand. A policy analyst, who has worked extensively on equity issues at the OECD, explains:

[I]t is a very philosophical issue, and we are OECD – an *economic* development organisation – [we did not want to get] involved in a huge philosophical discussion on what is equity and what it is not, so we went for definitions that can be defined with our indicators and that are simple to understand and clear....After looking at the research, the philosophical research literature, we decided that we couldn't go in there. It was a very difficult time ... at that time, the EU had done a big effort of indicators for equity that ended up being 140 indicators. It was crazy! This was in 2004. So we said, do we do a compound of indicators ... that's too complicated – compounded indicators don't work – there's always a mistake – compounded indicators are a synthesis of a number of indicators – so which ones do we pick, and ... we spent a long time [trying to decide] ... So we said no, we better go for something simple, transparent ... and this is why we opted to remain [with our own indicators]. In the first No More Failures, we had a little box on what had been the arguments. (Senior OECD policy analyst, Interview transcript)

Through such instrumentalism, the confusing philosophical debates and the variety in possible approaches were quite effectively – and literally – boxed up and bypassed by the OECD. The OECD's decision to persist with their own indicators demonstrates how previously developed calculations and established practices tend to persist even when they are seen as not being ideal.

Equity is different from equality. It is associated with broader ideas of justice and fairness, sometimes with 'equality of opportunity' and sometimes with 'equivalent treatment'. There is a great deal of philosophical literature on this. This report will not add to that literature, but will instead work pragmatically with two dimensions of equity: *fairness and inclusion*. (OECD, 2007, p. 29)

It is through such pragmatism and instrumentalism that the OECD attempts to contain the complexity of 'equity' in order to work with it. By containing the concept of 'equity', it becomes possible to conceptualise how it can be measured.

### Measuring Equity in PISA

PISA was developed in the late 1990s by the OECD in response to a growing demand from member countries for reliable educational output data to supplement their extensive input data. Since the first survey in 2000, PISA has been conducted every three years, and the list of participating systems has been growing steadily. Sixty-five systems participated in the most recent PISA survey in 2012. PISA is focused on outcomes, so its equity measures, put simply, are based on patterns of differences in outcomes for students of different groups. It measures inequity in terms of how closely correlated factors such as gender, socio-economic and cultural factors, ethnicity and migration are to student achievement. A close association implies an iniquitous system. While gender, migration and other factors are considered, and data are gathered, analysed and reported on these dimensions, the most emphasised factor, the one that appears to be of concern consistently across nations, is that of socio-economic background.

Measuring the equity of school systems in PISA involves the gathering of three separate sets of data:

- Performance data collected through the PISA tests in reading, mathematical and scientific literacy. The nature and composition of the test is the result of a long and complex negotiated process which involves decisions about which domains to cover, what kinds of questions are suitable, and how to organise the questions so that the short, two hour test elicits enough information to make reasonable projections from a sample of test takers (for a detailed account, see Gorur, 2011a).
- Student background data collected through a questionnaire. Here, questions range from education and occupation of parents to possessions such as books and DVD players and a desk and a quiet place to study at home. There is some scope to customise questions for each country.

- Data on school organisation and structure which is gathered through a questionnaire which the principal fills out. This gives an account of the school features such as size, location and student-teacher ratio.

It is the first and second of these three parts that are used to calculate equity. While the first produces measures of *performance*, the second, the student questionnaire, provides some of the components to calculate the *Economic, Social and Cultural Status* (ESCS). It is the relationship between performance and the ESCS that produces the comparisons of equity. In this article I focus on the calculation of the ESCS.

#### *The PISA ESCS Index*

The ESCS is PISA's version of SES – socio-economic status – that is more commonly used in a variety of calculations. 'Socioeconomic status' is a kind of short-hand way of describing the capacity to create and consume valued goods:

Socioeconomic status is typically used as a shorthand expression for variables that characterise the placement of persons, families, households, census tracts, or other aggregates with respect to the capacity to create or consume goods that are valued in our society. Thus, socioeconomic status may be indicated by educational attainment, by occupational standing, by social class, by income (or poverty), by wealth, by tangible possessions—such as home appliances or libraries, houses, cars, boats, or by degrees from elite colleges and universities. At some times, it has also been taken to include measures of participation in social, cultural, or political life. (Sørensen, 1994, p. 4)

Calculating ESCS is not straightforward. It is a measure of 'advantage' – the 'capital' that children bring to school based on their family background – factors that are thought to affect how they perform at school. For an idea as abstract as 'advantage' to be measured, it has to be translated into components that can be measured. As the PISA Technical Report (OECD, 2012) clarifies:

The PISA 2009 context questionnaires included numerous items on student characteristics, student family background, student perceptions, school characteristics and perceptions of school principals. In 14 countries (optional) parent questionnaires were administered to the parents of the tested students.

Some of the items were designed to be used in analyses as single items (for example, gender). However, most questionnaire items were designed to be combined in some way so as to measure latent constructs that cannot be observed directly. For these items, transformations or scaling procedures are needed to construct meaningful indices. (OECD, 2012, p. 280)

The measures of ESCS in PISA 2000 included five indicators: 'highest occupational status of parents (*HISEI*), highest educational level of parents (in years of education according to ISCED), family wealth, cultural possessions and home educational resources ...' (OECD, 2012, p. 21). So PISA combines established indicators such as the ISCED (International Standard Classification of Education) and the ISEI (International Socio-Economic Index of Occupational Status) with new indicators developed specifically for PISA, to calculate ESCS. The accuracy and validity of the ESCS in turn depends on the accuracy and validity of ISEI and the ISCED. The ISEI and ISCED are themselves a result of protracted struggles, and were accomplished through negotiation, compromise, agreement, argument, speculation, and, often, 'making do' – as I show below.

#### *The ISEI*

One item that forms a proxy for economic advantage is parents' occupation. The PISA student questionnaires (OECD, 2009) ask students to indicate details of parents' occupation through such questions as:

- What is your mother's main job? (e.g. school teacher, kitchen-hand, sales manager)
- What does your mother do in her main job? (e.g. teaches high school students, helps the cook prepare meals in a restaurant, manages a sales team)
- What is the highest level of schooling completed by your mother?

- Does your mother have any of the following qualifications?
- What is your mother currently doing?

Parents' occupation, as declared by students on a questionnaire, is charted on an internationally agreed classification of occupational status, the ISEI, which scales occupations by the average level of education and average earnings of job holders. The ISEI itself regularly undergoes revisions, as updated databases become available and as new occupations arise or social changes occur (see, for example, Ganzeboom & Treiman, 1996; Ganzeboom, 2010). Some of these revisions can be significant. For example, one major difference between the ISEI-88 and ISEI-2008, is that the new one is based on data for men *and women*, whereas the earlier one used only data on men to estimate the scale, with adjustments for the differences in the levels of part-time work between men and women. The ISEI-2008 webpage [1] explains why data from women was used in the new version, and also details the issues with this inclusion:

The construction of the scale on men only has created feelings of discomfort among many users and raised suspicions that the scale may be invalid or less valid for women (mothers).... The reasons to move now to a scale estimated on women and men combined are of a pragmatic nature. First, it expands the available data and thus produces more reliable estimates. Second, this is all the more true for heavily female dominated occupations (such as nurses and preprimary teachers), for which the previous scales could produce only estimates using the rare men that are found in these occupations (who may be not so representative of these occupations to begin with). (Extract from the ISEI-2008 webpage)

These changes in measures change the understandings of social situations. Using data on men *and women*, the ISEI-2008 derived lower scores for some occupations:

The inclusion of women in the construction of the scale is responsible for the somewhat lower scores that some female-dominated occupations now obtain.

Thus the decisions about what to include, and how to count, are critical to the picture that emerges. And decisions about what to include are not that easy to make. Even after years of being used, they remain precarious. Notably, measurement seems to involve attention to 'feelings of discomfort' and raised suspicions and is a matter of experts convincing each other as much as cold and dispassionate calculation.

Of course, using occupations could only ever be an approximate proxy for income at best – people of the same occupation can have widely varied income levels, particularly when these measures are applied across a range of nations. And even if measurement experts keep tinkering to make the measures more accurate, are 15-year-olds interested in, and informed about, their parents' occupations? Can adolescents be trusted to provide accurate information about their parents' jobs? This is an empirical matter and can only be known if other measures, such as parent questionnaires, are introduced as a check. But parent surveys would only work as a check if parents provided reliable responses – and apparently this cannot be taken for granted:

When reviewing the validity studies it should be kept in mind that 15-year-old students' reports of their parents' occupations are not necessarily less accurate than adults' reports of their own occupations. Individuals have a tendency to inflate their own occupations, especially when the data are collected by personal (including telephone) interviews. The correspondence between self-reports of husbands and wives is also imperfect. In addition, official records of a parent's occupation are arguably subject to greater sources of error (job changes, changes in coding schemes, etc.) than the answers of 15-year-olds. (OECD, 2002, p. 62)

It appears that gathering accurate data on parents' occupations is not that easy whoever the informant.

### ISCED

Another feature of student background that is seen as relevant to a student's ESCS is the education of the parents. But there is a great variation in nomenclatures and structures of education around the world. These differences would make for a confusing scenario, even if children were able to name their parents' qualifications accurately. To facilitate comparability of qualifications across the

world, PISA works with another set of internationally agreed classifications, the International Standard Classification of Education (ISCED). The troubled history of ISCED is well documented (see, for example, Heyneman, 1999; Sauvageot, 2008; Sharma, 2011); it took decades to reach any kind of agreement in standardising this classification.

Claude Sauvageot was closely associated with the development of indicators and chaired the restricted group of experts [2] charged with the task of revising ISCED. Explaining the many reasons why the defining of entities, so crucial to the construction of such classifications, is difficult, Sauvageot describes the deliberations and negotiations that went into the processes of revising the indicators:

During the meetings of the restricted group and other meetings of experts, the discussions on definitions were lively and passionate, with every member trying to use or validate the opinions or organisation of their own country. This highlighted differences in the interpretation of basic education (different from primary education and compulsory education), university education (which is not necessarily provided in a university), vocational education (which is not necessarily provided in an education institution), completed level (always a delicate notion if there is no examination at the end of the programme concerned), guidance towards a determined curriculum (initial definition of the programme's objective which can largely differ from the actual career of the pupils who undertake this programme ... . Had there been no details or criteria on all these subjects, the countries represented in the groups of experts would have included very different programmes under these terms and comparisons would have been inaccurate. (Sauvageot, 2008, p. 214)

Again, it is passion, discussion and persuasion, rather than cold calculation, that are characteristic of this practice of producing commensurability. Nothing is simple or straightforward, as Sauvageot recounts:

I thought everything was simple. All we had to do was define a duration of studies for each level of education and the whole problem would be solved. Except that nobody is actually capable of comparing the content of a school year in a given country with that of another. Should we consider the total number of teaching hours, teachers' qualification, the educational equipment available to the pupils (books for example)? Many other elements such as Internet access can now be added to this list. In short, a school year is never identical within the same country, let alone from one country to the next. (Sauvageot, 2008, p. 220)

So gathering data about the level of education of parents in an internationally comparable way is itself challenging, and while instruments such as ISCED make it possible to assign values and perform calculations, there is a measure of instrumentalist pragmatism here. The PISA 2009 Technical Report also reports on this:

Like occupation, the collection of internationally comparable data on parental education poses significant challenges, and less work has been done on internationally comparable measures of educational outcomes than has been done on occupational status. The core difficulties with parental education relate to international comparability (education systems differ widely between countries and within countries over time), response validity (students are often unable to accurately report their parents' level of education) and, especially with increasing immigration, difficulties in the national mapping of parental qualifications gained abroad. (OECD, 2012, p. 2)

So classifying and standardising pose ongoing challenges, confounding the effort to measure precisely. Correlating these data with student performance is even more fraught. Decades of research using these approximate data shows that parents' education is even more strongly related to student achievement than material wealth and family income. However, this is again to be qualified contextually, because the effect of material resources is likely to be higher in developing nations where even basic resources are not available to all (Marks et al, 2006). Moreover, this corroboration through triangulation is itself based on the 'mundane ontology' of representation (Woolgar, 1991).

It is significant to note here that the PISA Technical Report itself highlights the inadequacies of the indicators it uses. But making such declarations in technical reports, rather than in the main



PISA report which is meant for policy makers, also serves to reinforce such measurement struggles as *technical* matters rather than political or ontological ones, and restricts the spaces of criticism to technical experts with statistical knowledge.

### *Social and Cultural Capital and Family Wealth*

The variety in expressions of wealth and social and cultural capital across the world must be enormous – how is this variety tamed so that it can be measured in a standardised way across the world? Understanding what items contribute towards ‘social and cultural capital’ and what items might be symbolic of ‘wealth of the family’ is no easy matter.

According to the PISA Technical Report, ‘[h]ousehold assets are believed to capture wealth better than income because they reflect a more stable source of wealth’ (OECD, 2012, p. 10). So, information about possessions stands proxy for wealth. In PISA 2009, this part of the questionnaire invited information on 13 household items, in addition to three country specific optional items (OECD, 2012).

The ability of parents to support their children’s learning is also a form of capital. Again, this was not so straightforward to operationalise into a measure. One interviewee explained this process in detail, illustrating how ideas were dropped if they ‘don’t work’ in empirical field trials:

Index of possessions – we worked hard earlier on to get social capital, cultural capital and financial capital. It was done a little bit differently in PISA – it was about how much are the parents involved in the education of their child, talk to you about your schoolwork: so what kind of resources are available to you outside of school; that didn’t work because if your parents are asking you a lot about your schoolwork, guess what the problem is – you are not a good performer, right? So often there is a negative relationship. If you look at kids who say ‘my parents always talk to me about school,’ they are the ones who don’t go very well. So that didn’t get used. (Interview transcript, PISA expert)

We might readily imagine that the involvement of parents in their child’s education varies widely from culture to culture (Asian ‘tiger mums’ have now become a widely discussed phenomenon), though of course there is wide variation within cultures too. It would also depend on the levels of education of the parents and their ability to be involved. In some countries, parents’ education may matter less if they are able to afford tutoring or coaching classes outside of school, as is common in many East Asian countries. When the meaning of such involvement begins to vary widely from country to country, these items are discarded from core questionnaires, even if they might be thought, conceptually, to be significant. However, countries can include additional questions to elicit information specific to their situation.

To determine social and cultural capital and wealth, PISA student questionnaires ask several questions about possessions – about DVD players, the number of books in the home, the number of mobile phones and so on (OECD, 2012). Decisions about which items to include are settled by field testing:

The cultural capital – whether you own a violin or a piano or go to concerts and stuff – that didn’t get used – but the possessions did get used. And the books in the home is always there – it is very odd, I haven’t understood why books in the home is such an important thing, it is obviously some kind of indicator of the environment that fosters an interest in education, but we have never actually unpacked what it is about the books in the home. (Interview transcript, PISA expert)

Even though lifestyles have changed drastically in recent times, the questions relating to equity have remained largely stable over time. Indeed, there is a bit of conservatism with the equity items:

These questionnaires are a little bit different to assessment items. With assessment items, you are always wanting to be creating something that hasn’t been done before. ... With questionnaire items there is not so much impetus to be having suddenly a new question on equity. (Personal communication, PISA expert)

Such items as the number of books in the home may change if there is a shift to soft copies of books – but, the PISA expert tells me, they have not faced this as yet.

Although a wealth of detail is collected, not all of it is stable across nations. Some of the data is only valid for analysis at the national level, and is not used in international analysis, because the possessions do not have the same meaning between countries as diverse as Azerbaijan and Australia.

[I]n international comparisons you can't use things like the number of bathrooms – they [country level analysts] use all that for their own analysis. That is why we've relied almost totally on education and occupation because across countries, it is the best comparison because it means something in each country. (Personal communication, PISA expert)

But agreement between experts about which indicators to use and how to weight them is not always achieved:

Basically, Doug Willms, the Canadian colleague, did the analysis for the OECD.... What he did, I think he gave much more weight to home possessions because he basically had five indicators – occupation, education, then three separate indicators based on home possessions and we [ACER researchers] found that actually it is a bit too much weight to indicators that are not very discriminatory, actually, in many countries. In Iceland or Norway, everyone has a DVD player or whatever you might want to use, whereas if you go to Mexico or Tunisia, you would still find that they discriminate. So we wanted to give less weight [to home possessions in later surveys]. (Personal communication, PISA expert)

My interviewee explained: 'I am not very positive about home possessions as a good indicator, quite frankly. In [another study] we dropped it after field trials and we used just books in the home as the measure'.

In a conference paper, Wolfram Schulz (2005) recounts the analysis done on PISA 2000 and 2003 data, in which the 2003 indicators were retrospectively applied to the 2000 data to measure the consistency across the two sets of indicators. These types of empirical tests as well as methodological experiments are part of the measurement community's processes of testing the robustness of their methodologies and of refining them:

It needs to be recognised that there are some methodological problems with the collection and validation of reliable student background data in international educational research. Reviewing the stability of the three components ... indicates that there is some larger than expected variation across cycles, in particular with regard to (non-ICT related) household items. Some variation may be due to (minor) format changes across the two cycles and care should be taken to avoid even smaller modifications in future surveys. (Schulz, 2006, n.p.)

With lifestyles changing quite rapidly in some countries, measuring the ESCS is likely to remain challenging and unstable with each round of PISA. But of more interest to the present discussion is the way correlation, triangulation and field trials are used to validate indicators and calculations. There is a circularity in such validation – a concept about a phenomenon is developed, and then measures are developed based on that concept. Technical discussions then centre around the fit between the measure and the concept, even as such measurement serves to 'fix' the particular concept of the phenomenon.

### **The Instrumentalism and Performativity of International Comparisons**

Tracing these processes of calculation, it is apparent that each element in the cocktail that makes up the ESCS index – ISES, ISCED, and PISA's own indicators of wealth and cultural capital and family possessions – are uncertain, tentative and loose agreements achieved with a mix of information, expertise, passion, suspicion, field testing and pragmatism, in addition to the application of complex mathematics. Moreover, the weight attached to each of the elements also involves negotiation. To what extent do each of these elements – parents' education, parents' occupation and family wealth – stand for 'advantage' in different contexts? How do they affect each other? Are they differentially influential at different intensities? These questions may come to be tentatively settled through negotiation and may stabilise and become methodological conventions over time.

If coming up with the calculation is complex and uncertain, ESCS runs into more uncertainty with how it is understood and used. There are a few things to remember about the equity data in

PISA, my interviewees tell me. First, the SES correlation data are based on performance in the major domain subject, so comparisons from year to year are not exactly like to like. In other words, PISA 2000 looked at the correlation between ESCS and how students performed in *reading* literacy whereas the 2003 data looked at the correlation based on performance in *mathematical* literacy.

What you need to be careful of is reading this graph over time. And you have to be sure you know which subject it is – because reading, mathematics and science have been the focus of analysis in separate surveys of PISA. So suddenly you will see [the graph for] Reading and [think] ‘oh, Australia has a very high SES effect’, and then you go to Mathematics, and [think] ‘oh, it is not so high’. You just have to be very careful that you know you are comparing like to like and I think some people have made that mistake. (Interview transcript, PISA expert)

Moreover, with so many approximations, and with one set of composite indicators nested within other composite indicators, and with each of these being revised and updated constantly, the measurements obtained can change significantly even with small changes in methodology. One expert explains:

You have to be careful what you take as a measuring stick for inequity. SES measurement is not 100% accurate. Small differences in effect coefficients may be due to the changes in measurement. In the first few cycles we were unhappy – we were experimenting and trying to refine – the first cycle was one big experiment. Never underestimate the effect of all that. (Personal communication, PISA expert)

PISA’s age-based sampling also presents a problem that is particularly significant in some countries, in that large populations of 15-year-olds [3] are missing from the survey, since they are no longer in school:

It was okay in the beginning [when only the OECD member nations were involved] to have an age-based population. In Mexico, 40% of the 15-year-olds are no longer in school ... in half the [PISA test-taking] countries unbelievably large populations of students are no longer there [in school]. Is it still good to compare in that way? (Personal communication, PISA expert)

So the size and shape of ‘equity’, and the extent to which it is seen as a policy problem, is very much determined by the decision to survey 15-year-olds rather than, say, 11-year-olds.

There is also a tension between the need to retain the same structure and content in order to generate data that is comparable over time, and the need to adapt and update surveys and methodologies to reflect new developments and understandings. In addition, to calculate how equitable a system is, these uncertain ESCS measures have to be combined with the literacy surveys which come with their own limitations and uncertainties (see Gorur, 2011a).

Although the uncertainties and imperfections of these measures are well known to the measurement experts, persisting with them and trying to refine them is seen as a practical decision. The ‘forms’ or formats that undergird these measurements require considerable resources to develop (Thévenot, 2009), and displacing them can be very costly. I described above how both ISEI and ISCED have been developed through the mobilisation of enormous resources. OECD’s decision to continue with the indicators with which it was already familiar despite the knowledge that they did not capture the complexity of the equity issue demonstrates how and why these forms persist. The more enmeshed a measure is in multiple networks, the greater the investment in that form, and the harder it is to change it. ISCED and ISEI, for example, are used internationally by a range of agencies and are embedded in a variety of calculations. Any shake-up or destabilisation of accepted indicators and methods would result in undoing all the calculations in which the indicators participate, which would be very costly.

I am not suggesting here that there is some conspiracy to deliberately conceal the flaws in measurement. Rather, such instrumentalism is simply a part of the routine epistemic practice of comparison – nothing out of the ordinary that needs to be highlighted or reported, as far as the measurement community is concerned. Moreover, since these details are seen as *technical* rather than political decisions, they are reserved for technical debate, and may appear in technical reports and get debated in specialised journals. Indeed, making these details available to the measurement community is a way to *gain* rather than lose credibility within the measurement community.

Cautionary footnotes are often used to draw attention to the technical uncertainties and the methodological choices involved.

While the measurement community might value insights into the messy details of the complex challenges of such comparisons, hesitant and tentative science is not currently valued by politicians and policy makers. As Jasanoff (2011) explains, there is a difference between the registers of rigour and truth that are required for policy knowledge and for academic science. For this reason, technical reports are produced separately to the analyses of PISA prepared for the consumption of policy makers. Separating some aspects of PISA as ‘policy-relevant’ and others as ‘technical’ neatly classifies ‘politics’ and ‘science’ as two separate activities, engaged in by two separate groups. This move accomplishes several important actions:

- It presents policy makers with unambiguous and ‘objective’ data that they can use with assurance
- It distances the technical experts from politics so that they are not tainted with the accusation of subjectivity or partisan interests, and can be seen as providing dispassionate advice
- It preserves the messiness – anathema to policy makers but markers of good science to technical experts – to justify the science
- Most of all, it makes it possible for such measurements to persist and to continue to be seen as ‘objective’ and more or less accurate representations of an already existing world, and to withstand the accusations of imprecision levelled at them

Jasanoff (2011, p. 313) argues that the fact that ‘policy controversies so often play out in a scientific idiom draws attention away from human decision makers, while modelling and quantification hide professional judgement behind the seeming objectivity of formal simulations and mathematical representations’. By focusing on the mundane and tacit practices and the range of selections and decisions that are folded into international comparisons, I have attempted here to reintroduce the human decision makers into the picture in this analysis.

While the discussion above highlights the instrumentalism needed to cobble together such measurements and explains their persistence, I want to press on further to argue, following Woolgar (1991), that such comparisons are not merely technically flawed *representations*, but are *performative*. The very attribution of the character of calculability (Mitchell, 2002) and the production of commensurability changes what it is we understand as ‘equity’ in education. The OECD average becomes a mean tolerance threshold for the amount of inequity that is deemed acceptable or ‘normal’. The extent to which equity gets defined as a ‘policy problem’ gets determined by this measure. This consequence is well illustrated in the case of Australia, which moved from a ‘low equity’ to a ‘high equity’ nation based on changes to the measurement of equity. Nations in the high-equity quadrant have become role models for those with more iniquitous systems as per PISA. The OECD uses these measures to identify practices that are believed to lead to high or low equity and promotes these practices through its policy advice, with consequences for nations and individuals.

### **Critiquing the Technology of International Comparisons**

The detailed exploration I have provided of how ‘equity’ is measured in international comparisons should by now have convinced the reader that such comparisons are constructions rather than descriptions (Knorr Cetina, 1981). But nothing of what I have described here is new – the issues I have raised are in fact those that the measurement community itself pointed out to me. Yet the practice of regarding these international comparisons as more or less accurate representations continues apace, both within and outside the measurement community. I am faced with the same question that confronted Woolgar (1991, pp. 320-321) in relation to the persistence of citation as a measure of quality: ‘under what circumstances does an institutionalised system of measurement engage with its critics?’

Woolgar finds that at the crux of the issue is the supposition, in such measurement, of a transcendental object to be measured. Such a supposition is the basis of such institutionalised practices of science as ‘triangulation’ which employs the following logic:

[I]f indicator A is suspected of inaccuracy, we can always check it against indicator B (and with indicators C, D etc.). Attention is devoted to the relationship between the two indicators rather

than to the question of whether or not there is something to measure out there in the first place. It is as if the more indicators one invokes, the more one can rely upon the base assumption of a transcendental object. (Woolgar, 1991, p. 322)

Challenging this assumption of a transcendental object is hard to do. Critiquing the accuracy of measurements only serves as 'a way of reinforcing the transcendental nature of the measured "it"', since such criticism derives from the same 'mundane idiom' of representation as the measurement itself (pp. 322-323). To critique the measurement of equity in PISA, we would need to problematise the very assumption of a transcendental object such as 'equity' and the possibility of its representation through measurement. Describing the ways in which the institutional practices of international comparisons are shaped by a multitude of practical decisions presents the detail of this performativity.

Understanding the performativity of PISA can pave the way for researchers to go beyond critique in the 'mundane idiom'. As Stengers (2011) argues, by pretending that processes of comparison are disinterested and objective, we give up the possibility of debating them and collectively involving ourselves in their construction. She advocates, instead, an 'active and interested comparison ... made by a collective sharing of the same matter of concern, privileging what can be associated with new questions and experimentally challenging consequences' (Stengers, 2011, p. 5). In order to open up possibilities for such active and interested inquiry, it is important to demonstrate, empirically, how politics are folded into methodologies and technologies, and how the separation of science and politics is achieved. As Woolgar (1991) reminds us, the instrumentalism of such measurement technologies need to be highlighted. The enormous influence of such measurements depends on the assumption of their 'objectivity' and representational capacity, so the more such descriptive studies which draw attention to their performativity, the better we might be able to break the old habits of misunderstanding the nature of science.

Adopting an ontologically relativist position, in this article I have asserted that the very act of measurement changes the entity being measured. Approaching critique from a stance that is neither 'technical' nor 'ideological', but is grounded in the notion of ontological variability, it becomes possible to see the consequences for this phenomenon we call 'equity' when its measurement is performed through the particular practices of international comparisons in PISA. By focusing on the practical accomplishment of such comparisons, the performative nature of these measurements is highlighted.

In doing this type of critique, my hope is to develop a sociology that speaks to the 'scientist' and the 'politician', the technician and the ideologist. In presenting the processes by which PISA equity measurements are put together, my purpose is not to discredit these data by pointing out that they are not objective or precise. I am not concerned with 'exposing' the limitations of comparison. Rather, I am attempting to create spaces for interdisciplinary engagement to examine how comparison works, to respect its performativity and its limits, and to develop other, and possibly more useful, ways of engaging with comparison – and indeed with the objects of comparison, such as equity in education. While this case study is of international comparisons in the field of education, such global categorisations, standards and indicators are deeply implicated in the production of calculable worlds in a variety of fields. They are, to use Knorr Cetina's (2007) term, the 'machinery' through which global societies are being produced. As part of the new sociologies that are evolving to describe emerging phenomena in these times of mobility, this article demonstrates the critical possibilities offered by a 'sociology of measurement' in understanding global machineries of knowledge production.

## Notes

[1] <http://www.harryganzeboom.nl/isco08/qa-isei-08.htm>

[2] The group became 'restricted' when a more inclusive group was unable to reach consensus despite two years of work, as Sauvageot documents.

[3] PISA is taken by students who are between 15 years and three months and 16 years and two months at the time of the test.

## References

- Appadurai, A. (1996) *Modernity at Large: cultural dimensions of globalization*. Minneapolis, MN: University of Minnesota Press.
- Appadurai, A. (2006) *Fear of Small Numbers: an essay on the geography of anger*. Durham, NC: Duke University Press. <http://dx.doi.org/10.1215/9780822387541>
- Ball, S.J. (2009) Privatising Education, Privatising Education Policy, Privatising Educational Research: network governance and the 'competition state', *Journal of Education Policy*, 24(1), 83-99. <http://dx.doi.org/10.1080/02680930802419474>
- Beck, U. (1992) *Risk Society: towards a new modernity*, trans. M. Ritter. London: Sage.
- Busch, L. (2011) *Standards – recipes for reality*. Cambridge, MA: MIT Press.
- Callon, M., Lascoumes, P. & Barthe, Y. (2009) *Acting in an Uncertain World: an essay on technical democracy*. Cambridge, MA: MIT Press.
- Derksen, L. (2000) Towards a Sociology of Measurement: the meaning of measurement error in the case of DNA profiling, *Social Studies of Science*, 30(6), 803-845. <http://dx.doi.org/10.1177/030631200030006001>
- Ganzeboom, H.B.G. (2010) A New International Socio-Economic Index [ISEI] of Occupational Status for the International Standards Classification of Occupation 2008 [ISCO-08] Constructed with Data from the ISSP 2002-2007; with an analysis of occupational measurement in ISSP. Paper presented at the annual conference of International Social Survey Programme, May, in Lisbon. <http://www.harryganzeboom.nl/pdf/2010-ganzeboom-isei08-issp-lisbon-%28paper%29.pdf>
- Ganzeboom, H.B.G. & Treiman, D.J. (1996) Internationally Comparable Measures of Occupational Status for the 1998 International Standard Classification of Occupations, *Social Science Research*, 25, 201-239. <http://dx.doi.org/10.1006/ssre.1996.0010>
- Gorur, R. (2008) Explaining Global Policy Phenomena using the Small and the Mundane: a network analysis of PISA. Paper presented at the AARE 2008 International Education Research Conference, 30 November-4 December, in Brisbane.
- Gorur, R. (2011a) ANT on the PISA Trail: following the statistical pursuit of certainty, *Educational Philosophy & Theory*, 43(5-6), 76-93.
- Gorur, R. (2011b) Policy as Assemblage, *European Educational Research Journal*, 10(4), 611-622. <http://dx.doi.org/10.2304/eej.2011.10.4.611>
- Gorur, R. (2012) The Invisible Infrastructure of Standards, *Critical Studies in Education*, 54(2), 132-142. <http://dx.doi.org/10.1080/17508487.2012.736871>
- Gorur, R. (2013) My School, My Market. Special Issue, *Discourse*, 34(2), 214-230. <http://dx.doi.org/10.1080/01596306.2013.770248>
- Gorur, R. & Koyama, J.P. (2013) The Struggle to Technicise in Education Policy, *Australian Educational Researcher*, 40(5), 633-648. <http://dx.doi.org/10.1007/s13384-013-0125-9>
- Hannam, K., Sheller, M. & Urry, J. (2006) Editorial: Mobilities, Immobilities and Moorings, *Mobilities*, 1, 1. <http://dx.doi.org/10.1080/17450100500489189>
- Heyneman, S.P. (1999) The Sad Story of UNESCO's Education Statistics, *International Journal of Educational Development*, 19, 65-74. [http://dx.doi.org/10.1016/S0738-0593\(98\)00068-6](http://dx.doi.org/10.1016/S0738-0593(98)00068-6)
- Hopman, S.T. & Brinek, G. (2007) Introduction: PISA According to PISA – does PISA keep what it promises?, in S. Hopman, G. Brinek & M. Retzel (Eds) *PISA According to PISA – does PISA keep what it promises?* Berlin: LIT Verlag Dr. W. Hopf.
- Jasanoff, S. (2011) The Practices of Objectivity in Regulatory Science, in C. Camic, N. Gross & M. Lamont (Eds) *Social Knowledge in the Making*, pp. 307-339. Chicago: University of Chicago Press.
- Knorr Cetina, K. (1981) *The Manufacture of Knowledge: an essay on the constructivist and contextual nature of science*. Oxford: Pergamon.
- Knorr Cetina, K. (2007) Culture in Global Knowledge Societies: knowledge cultures and epistemic cultures, *Interdisciplinary Science Reviews*, 32(4), 361-375. <http://dx.doi.org/10.1179/030801807X163571>
- Lampland, M. & Leigh Star, S. (Eds) (2009) *Standards and their Stories: how quantifying, classifying, and formalizing practices shape everyday life*. Ithaca: Cornell University Press.
- Latour, B. (1999) *Pandora's Hope: essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- Latour, B. (2004) Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern, *Critical Inquiry*, 30, 227. <http://dx.doi.org/10.1086/421123>

- Latour, B. (2005) *Reassembling the Social: an introduction to actor-network-theory*, Oxford: Oxford University Press.
- Lawn, M. & Grek, S. (2012) *Europeanizing Education: governing a new policy space*. Oxford: Symposium Books.
- Marks, G.N., Cresswell, J. & Ainley, J. (2006) Explaining Socioeconomic Inequalities in Student Achievement: the role of home and school factors, *Educational Research and Evaluation*, 12(2), 105-128.  
<http://dx.doi.org/10.1080/13803610600587040>
- Mitchell, T. (2002) *Rule of Experts: Egypt, techno-politics, modernity*. Berkeley, CA: University of California Press.
- Niewöhner, J. & Scheffer, T. (2010) Foreword, in J. Niewöhner & T. Scheffer (Eds) *Thick Comparison*, pp. vii-viii. Leiden: Brill.
- Organisation for Economic Cooperation and Development (2002) *PISA 2000 Technical Report*. Paris: OECD.
- Organisation for Economic Cooperation and Development (2007) *No More Failures: ten steps to equity in education*, p. 143. Paris: OECD.
- Organisation for Economic Cooperation and Development (2009) *PISA 2009 Assessment Framework – Key Competencies in Reading, Mathematics and Science*, p. 292. Paris: OECD.
- Organisation for Economic Cooperation and Development (2012) *PISA 2009 Technical Report*. Paris: OECD.  
<http://dx.doi.org/10.1787/9789264167872-en>
- Pickering, A. (1995) *The Mangle of Practice: time, agency, and science*. Chicago: University of Chicago Press.  
<http://dx.doi.org/10.7208/chicago/9780226668253.001.0001>
- Rizvi, F. & Lingard, B. (2010) *Globalizing Education Policy*. London: Routledge.
- Rutkowski, D. & Rutkowski, L. (2013) Measuring Socioeconomic Background in PISA: one size might not fit all, *Research in Comparative and International Education*, 8(3), 259-278.  
<http://dx.doi.org/10.2304/rcie.2013.8.3.259>
- Sauvageot, C. (2008) A Tool for International Comparison: the International Standard Classification of Education (ISCED), in A. Charras & C. Sauvageot (Eds) *Education and Formations: international comparisons*. Paris: Ministry of National Education, Ministry of Higher Education and Research Evaluation, Forecasting and Performance Directorate.
- Schulz, W. (2005) *Measuring the Socio-economic Background of Students and its Effect on Achievement in PISA 2000 and PISA 2003*. Paper presented at the Annual Meeting of the American Educational Research Association, 7-11 April in San Francisco,.
- Sharma, Y. (2011) How China is Winning the School Race, *BBC News – Business Section*, 11 October.  
<http://www.bbc.co.uk/news/business-14812822> (accessed 5 March 2012).
- Sørensen, C.W. (1994) Success and Education in South Korea, *Comparative Education Review*, 38(1), 10-35.  
<http://dx.doi.org/10.1086/447223>
- Stengers, I. (2011) Comparison as a Matter of Concern, *Common Knowledge*, 17(1), 48-63.  
<http://dx.doi.org/10.1215/0961754x-2010-035>
- Stronach, I. (2010) *Globalizing Education, Educating the Local*. Oxford: Routledge.
- Thévenot, L. (2009) Postscript to the Special Issue, *Social Studies of Science*, 39(5), 793-813.  
<http://dx.doi.org/10.1177/0306312709338767>
- Timmermans, S. & Berg, M. (2003) *The Gold Standard: the challenge of evidence-based medicine and standardization in health care*. Philadelphia: Temple University Press.
- Woolgar, S. (1991) Beyond the Citation Debate: towards a sociology of measurement technologies and their use in science policy, *Science and Public Policy*, 18(5), 319-326.

---

**RADHIKA GORUR** is a Post-doctoral Research Fellow at the Victoria Institute, Victoria University, Australia. Her current research focuses on contemporary education policy, PISA, evidence-based policy, collaborative knowledge production and youth policies. She is interested in how policy ideas evolve, circulate and stabilise. Her sociologies of measurement and numbers are guided by an impulse to engage in productive critique that provides platforms for interdisciplinary conversations rather than critique that seeks to debunk or deconstruct. *Correspondence:* [radhika.gorur@vu.edu.au](mailto:radhika.gorur@vu.edu.au)