1 ASSEMBLING A SOCIOLOGY OF NUMBERS

Radhika Gorur (The Victoria Institute, Victoria University, Australia)

INTRODUCTION

That numbers are now ubiquitous in education is beyond dispute. Large-scale surveys of literacies now routinely produce rankings, benchmarks and standards that participate in significant ways in national policies. Not only are they routinely used and, many would argue, misused, they appear to have colonised our collective imaginations, reducing the scope of policy issues by rendering them into variations of positions on league tables and comparative accounts. Emanating from a few 'centres of calculation' such as UNESCO and the OECD, numbers have become the international language of literacy, with such programmes as Education for All and Programme for International Student Assessment (PISA) which are almost global in scope.

The alarming ubiquity and influence of numbers as measures of literacy have been accompanied by strong critique of the use of numbers in policy. These arguments can be summarised as follows: (a) quantification cannot capture the complexity of education and is inherently reductive; (b) numbers are becoming hegemonic and marginalising other ways of knowing; (c) numbers are a technology of governmentality and should be resisted; and (d) numbers are being misused in policy and should be viewed with suspicion.

In this chapter, I shall elaborate some of these criticisms, and argue that whilst all of these criticisms are legitimate and useful, they are not sufficient. Taking a cue from scholars in Science and Technology Studies, I suggest that measurements of literacy are *performative* – i.e. world-making – processes (Gorur 2014a; Gorur 2014b; Scott 1998; Woolgar 1991). I explore the consequences of such an understanding for critique, and propose that a 'sociology

1

of measurement' (Gorur 2014a; Gorur 2014b; Espeland and Stevens 2008; Woolgar 1991) is required to adequately and effectively critique the ever-expanding field of literacy measurements.

To craft my argument, I will focus on the Programme for International Student Assessment (PISA), the flagship international literacy comparison of the OECD. I will also draw upon a range of interviews with PISA and OECD officials and measurement experts that I have conducted across several related projects over the last six years. Whilst PISA makes an ideal 'case study', I suggest that the argument I develop here holds for critique of numbers in general, and applies not only to other literacy assessments, but more widely to the current translations of the world into numbers across a range of social policy terrains.

PISA presents itself as a worthy focus of attention because of its extraordinary influence. With its high-profile media coverage and its international 'league tables', it epitomises the growth of 'literacy as numbers'. Initially developed for the rich countries' club of the Organization for Economic Cooperation and Development (OECD), PISA now has more non-member participants than member nations in its surveys. There are now extensions of PISA in the form of PISA for Schools to assess literacies comparatively at the school level, as well as plans for PISA for Development to assess literacies in developing nations (see Guadalupe, Maddox and Addey, this volume). So powerful has PISA become, and so threatening is its growing influence, that in April 2014, a group of high-powered academics (including Stephen Ball, Noam Chomsky, Robin Alexander and Henry Giroux) and parents, principals and others wrote an open letter[1] to OECD's Deputy Director of Education, Andreas Schleicher, to bring to his attention the negative consequences of PISA, and suggesting that the OECD immediately make some changes to mitigate against the worst of its effects.

That so many thoughtful people are becoming so alarmed as to write such a letter is a testament to the influence of PISA. It is hard to believe that the first PISA survey was done as recently as 2000, not only because it is now so widely used to inform and justify policy, but also because it has so easily displaced other forms of assessing, evaluating and understanding education systems. Before PISA came along to tell us that Germany had a shocking system and Finland had a great one, many countries, including Canada and Finland, looked to Germany to learn from them. As one well-informed policy expert said to me in an interview:

> Canadians used to be constantly going to Germany to study them so we could copy their system! I went to this meeting in Berlin in 2001, where their Federal

> Minister got up and said, 'Well, we need to learn from Canada, because they're doing so much better', and I wanted to yell out, 'Give us our money back for all the trips we've made!' Finland has got new hotels to accommodate all the PISA visitors. And they were looking to Germany before the PISA results came out. Everyone was going to the US all the time – no-one goes to the US any more to see how they do schooling … no-one thinks that the US is the international model for how to do schooling. (Interview transcript: senior policy official)

Whatever criteria Canada and Finland had used over the years to conclude that the Germans had something to teach them about education were instantly trumped when the PISA literacy numbers were announced. Other evaluations become mere intuitions or vague feelings – easily dislodged with 'That's not what the *numbers* say!' Indeed, I would say that currently numbers appear to be speaking louder than those who are critical of them.

PISA was developed explicitly for the purpose of informing policy. So PISA results are presented in specific ways with the purpose of speaking to policy-makers – framing policy issues, describing the parameters of relevance, and offering solutions. The logic of its comparisons is that systems that are not doing as well can learn from those that are. Analyses that accompany these PISA reports contain interpretations and suggestions that particular practices in policy, governance and schooling would contribute to higher rankings.

With PISA results being tied to economic benefit to nations (Hanushek and Woessmann 2012), the release of PISA rankings is eagerly awaited, the anticipation mounting as the results are released at the same instant around the world. Each release is accompanied by a media blitz. This has meant that the results have to be further simplified into bytes that are suitable for newspapers and the television. As one Australian policy and measurement expert said in an interview with me in 2009:

> When the results come out, there is quite a flurry of excitement. I was on Channel 7, 9, 10 and ABC and Sky News – all on the same day! But the media likes to hone in on a story and this time they just focused on 'reading results having gone down'. Politicians too want to do this – they want to think in the short term – only for so long as they remain in that portfolio. (Interview transcript, senior measurement expert)

Moreover, officials from the OECD also are often advisors to government and have the ear of the highly placed politicians and policy-makers in various countries and are able to spread their policy message quite effectively. The fact that Andreas Schleicher is a German did a lot to make PISA important in Germany; 'he was all over the press in Germany because he could speak

German,' as one interviewee explained. More recently, explaining the rise of the importance of PISA in the US, a senior policy expert said that not only do the highest officials from the OECD, including the Secretary-General, Angel Gurria, come to the US to release PISA results and make public presentations with a large media presence, but Schleicher also meets with members of the Congress and with people in state governments and other influential policy-making bodies, spreading PISA's policy messages. For these reasons, PISA makes a wonderful exemplar of the spread and influence of 'literacy as numbers' in contemporary education policy.

CRITIQUING 'LITERACY AS NUMBERS'

The growing takeover of the education policy space by numbers has not gone unnoticed by critics in education, policy and sociology. Not only numbers, but the current ideological climate of neoliberalism and the accompanying market orientation within which numericisation has been flourishing has also come under the critical gaze of education and policy commentators. Most of the critique can be clubbed under the umbrella of 'debunking' numbers in one way or another, using a variety of arguments. The reductionism argument is that literacy assessments and the rankings that emerge from such surveys fail to capture the complexity of classrooms, schools and other learning contexts. Therefore, it is argued, literacy assessments, which are single-event snapshots, make poor proxies for 'student performance' or, worse still, 'outcomes' of educational systems. Not only do they reduce learning within a particular 'literacy' to the ability to respond to a small range of questions, but also they only include a small number of subjects. This is in part because many valued aspects of education cannot easily be assessed internationally in a comparative way. As one PISA official explained:

> Reading, science and maths are there largely because we can do it. We can build a common set of things that are valued across the countries and we have the technology for assessing them. So there are other things like problem-solving or civics and citizenship – that kind of thing where there would just be so much more difficulty in developing agreement about what should be assessed. And then there are other things like teamwork and things like that. I just don't know how you'd assess them in any kind of standardised way … So you are reduced to things that *can* be assessed. They've tried writing – but … the cross-cultural language effect seems too big to be comparable. So the things we assess are a combination of the things we value and the things we can do – I think it sends an odd message about

science, perhaps, but I don't think anyone would argue about literacy and numeracy. (Senior PISA official, cited in Gorur 2011)

Moreover, large-scale, sample-based surveys are expensive and so necessarily brief. They are limited in the breadth of their coverage. Not all aspects of even the limited range of literacies, let alone the schooling experience of students, can be captured in such surveys, however well they are designed and conducted, and however carefully they are interpreted. The inclusion of diverse cultures necessarily ignores differences to produce commensurate entities. Indeed, it is only by giving up the ambition to know things in depth that such assessments can become large-scale (Gorur 2011; Latour 1999).

While the reductionism argument is important to make because policymakers appear not to realise the limits of the warrant of numbers, I would argue that reductionism, in itself, is not necessarily problematic. Nor is reductionism merely a product of ignorance or poor science. In my exploration of the history of indicators that underpin PISA (Gorur 2014b), and in the construction of PISA itself (Gorur 2011; Gorur 2014a), I have found that the laborious processes that culminate in the PISA surveys are not trivial exercises but developed through lengthy negotiations, with a complex understanding of the challenges to adequacy and fitness for purpose of these assessments. Moreover, it is only through having reductionist inscriptions that certain patterns become visible and certain actions are imaginable and made possible. Scott explains how a certain 'tunnel vision' can be useful in producing certain types of knowledge:

> Certain forms of knowledge and control require a narrowing of vision. The great advantage of such tunnel vision is that it brings into sharp focus certain limited aspects of an otherwise far more complex and unwieldy reality. This very simplification, in turn, makes the phenomenon at the center of the field of vision more legible and hence more susceptible to careful measurement and calculation. Combined with similar observations, an overall, aggregate, synoptic view of a selective reality is achieved, making possible a high degree of schematic knowledge, control, and manipulation. (Scott 1998, 11)

Similarly, Latour (1987) has also talked of the value of such reductions. Numeric inscriptions, like the maps of the early explorers, can be transported back and forth without losing fidelity, and can provide the information required to develop strategies and focus energies and other resources in ways that are not possible without such maps to guide action. Maps are an excellent illustration of the usefulness of reductionist renderings. Maps are useful *because* they are reductionist, *because* they inscribe a less complex

version of the world. Moreover, maps make it possible to plan and strategise from a distance, without requiring physical presence on site to understand the terrain. Whilst local, detailed and complex knowledge is important for certain decisions, others require only the broad-brush picture. Of course, maps developed for one purpose may be useless for another, and maps that are ridden with errors can lead its users down the wrong path. But reductionism, per se, is not a reason to dismiss numbers as useless for the purposes of research or governance.

A second reason for the discomfort with numbers (and quantification in general) is that they are becoming hegemonic, displacing other forms of knowledge. Porter explains that even the knowledge that numbers have serious limitations does not limit preference for them over other ways of knowing:

> Critics of quantification in the natural sciences as well as in the social and humanistic fields have often felt that reliance on numbers simply evades the deep and important issues. Even where this is so, an objective method may be esteemed more highly than a profound one. (Porter 1995, 5)

Numbers – measurements and calculations – have a 'cold, hard facts' feel to them that is difficult to argue against; they close off spaces of debate (Barry and Slater 2002; Hammersley 2001). Other evaluations become mere intuitions or vague feelings. As one policy expert said:

> Governments have become ravenous for information and evidence. A few may still rely on *gut instincts, astrological charts or yesterday's focus groups.* But most recognise that their success – in the sense of achieving objectives *and retaining the confidence of the public* – now depends on much more systematic use of knowledge than it did in the past. (Mulgan 2005, 215, my emphases)

Non-numeric evidence and qualitative research based on such methods as 'yesterday's focus groups' now occupy the same status as 'gut instincts and astrological charts'. Part of the reason for this privileging of numbers is the persistence of the division between 'quantitative' and 'qualitative' methodologies, and a serious lack of collaborative engagement of scholars across these research traditions. This division is reinforced by institutional requirements, where researchers have to declare themselves as belonging to one camp or the other, and strengthened with journals often dedicated to one or other type of research. As a result, neither side trusts or understands the other. Even the category of 'mixed method' reinforces that the two methods are fundamentally different from each other. The privileging of 'quantitative research'

in education is supported by several 'clearing houses' that filter out research that does not conform to the requirements of quantitative methodologies.

That such a division is unjustified has been successfully argued by several theorists (see, for example, Gorard and Taylor 2004). There is no 'quantitative' work that does not involve qualitative decisions in terms of models, assumptions, choices of what to count, or what value or weight to place on different factors – 'quantitative research' is not just a matter of counting and measuring in some 'objective' way, but a science that is infused with decisions of all manner based on common sense, intuition and professional judgement. These processes, and the historical stabilisation of methodologies, have been described in detail (in, for example, Gorur 2014a; Desrosières 1998; Porter 1995). And 'qualitative research' could never report any findings without resorting to some sense of how much, how many, when, for how long, how frequently and so on – all quantitative judgements. Despite this, mutual suspicion with regard to methodologies and understandings persists, leading to a lack of engagement with – and even dismissal of – each other's research. 'Quantitative' work is often dismissed as 'positivist', and numbers are debunked on the basis that they are not objective but political. Indeed, I have myself worked in this arena of 'revealing' that numbers are political and not a pure 'view from nowhere' (Haraway 1988) (see, for example, Gorur 2011). But the point is this: no knowledge is 'objective', and no knowledge is detached from the methodologies, assumptions and world views that underpin it, or the vast networks of institutional and social structures that support it; neither the research seen as 'quantitative' nor that which is classified as 'qualitative' is immune to this limitation. So a dismissal of numbers on the basis that they are not 'objective' or 'apolitical' requires a corresponding dismissal of 'qualitative' knowledge on the same basis.[2]

A third objection to the widespread use of numbers in education has been on the grounds of their contribution to the surveillance mechanisms in the heightened governmentality that is the signature of our times. Numbers of all manner seek to shine a stark and unforgiving light on the effectiveness of schools and teachers and the progress of students (or lack thereof). However, some of the critique argues that audits and accountability are instruments of governmentality, and *therefore* to be resisted. I suggest that accountability in itself is desirable – even *essential* – for good government, as the effects of lack of accountability in corrupt nations demonstrate adequately. Democratic societies that expect governments to take responsibility for organising institutions such that they are fair and run ethically, efficiently and in a well-organised manner will rightly expect that adequate accountability measures

are in place. So protesting accountability itself is not only futile but would be dangerous if successful. This kind of advocacy would throw the baby out with the bath water and is, moreover, unlikely to be taken seriously by policy-makers.

While the 'governmentality' argument was attractive and even useful perhaps a decade or so back, particularly in the face of growing neoliberalism, it has also been a distraction, serving to marginalise 'qualitative researchers' even further. It has achieved little through the challenges it has posed; indeed over the last couple of decades, the education world has become yet *more* obsessed with the very things the critics of governmentality sought to resist. Such critique, perhaps, has run out of steam, to borrow Latour's (2004) expression.

The misuse of numbers – deliberate or otherwise – both by transnational organisations such as the OECD and by national and other governments and institutions – has been another line of critique. There is much merit in this – though perhaps it would be more effective to wage this battle in the popular press than in scholarly journals. That politicians cherry-pick numbers that suit, that they choose numbers according to convenience, that they suppress unfavourable numbers, and that they seek numbers to support already-made decisions is now well known. However, every instance of it needs to be challenged because decisions on literacy have far-reaching consequences for societies.

To summarise so far: some of the critique of numbers has focused on the inability of numbers to capture and represent the 'true picture' of the state of literacy. Other critiques have focused on the politics of *interpretation* and *use*, often claiming that policy-makers are either ignorant of the limitations of numbers, or are deliberately obtuse, in order to serve their own agendas. These criticisms are to a large extent irrelevant, to my mind. Merely showing that numbers are reductive is not enough – it is because they reduce, and because they can express a large amount of information in a parsimonious way that they are useful in governance. To argue that they are political rather than neutral is to deny that *all* forms of knowledge are political. Similarly, if the argument is that numbers are misused, the same can be said of non-numeric forms of knowledge. To merely object on the grounds that they are an instrument of governance is not enough; governance is an inherent part of all formal societies, and statistics is an invaluable tool of good governance as well. Espeland and Stevens point out that:

> Measurement can help us see complicated things in ways that make it possible to intervene in them productively (consider measures of global warming); but

> measurement also can narrow our appraisal of value and relevance to what can be measured easily, at the expense of other ways of knowing (consider how education became years of schooling in American sociology). (Espeland and Stevens 2008, 432)

To critique literacy as numbers more rigorously and usefully, I suggest, it is important to understand their *performativity*: that is, their productive role in knowledge creation and governance.

ASSEMBLING A SOCIOLOGY OF NUMBERS

Scholars in STS have demonstrated that measurement is not merely a descriptive exercise, but a *productive* one (Knorr Cetina 1999). Thinking in this *performative* idiom (Pickering 1995) requires that we understand literacy measurements as 'world-making' practices. The performative or productive aspects of measurement and its nuances need to be carefully understood if we are to engage with literacy measurements seriously and to develop useful critique.

The most readily understood 'world-making' quality of measurement is that as soon as measurements are instituted, they begin to act on the world (Gorur 2014b). PISA, for example, has set in motion large-scale changes in Germany, Poland and other countries. In the US, the Race to the Top scheme is seen as influenced by PISA (Alexander 2014). In Australia, the desire to rank in the 'top five' in PISA is now inscribed in the Education Act of 2013 (Gorur & Wu 2014). It has been demonstrated that in some countries, students are spending an inordinate amount of time training for high-stakes accountability measurements of literacy, at the expense of other types of learning (Berliner 2011). There are reports of schools and teachers gaming the system or downright cheating – suggesting that certain children stay home on days when literacy assessments are being conducted, so as not to adversely affect the school's average performance, and even changing students' responses on the test (Polesel, Rice and Dulfer 2014). Literacy comparisons are changing policies and practices on a very consequential scale. So whilst the intent of measurement might be to just represent an existing situation or 'reality', the act of measurement sets in motion a series of changes to that situation, thus changing the very reality it purports to measure.

There are more subtle and deeper impacts as well. Surveys and indicators present new patterns, produce new understandings of the world and highlight new problems to battle. They redirect attention and resources. Scaled up at

national and global levels, they create a collective imaginary that requires, and is reified through, specific sets of routines and practices (cf. Appadurai 1996; Taylor 2004; Hamilton 2012). Using a variety of frameworks, tables, graphs and diagrams, technologies of quantification order and clarify the chaotic jumble of the world, acting as technologies of visibility, accountability and control (Miller 2005). For example, during World War II, UNESCO began the task of statistically mapping literacy globally, measuring levels of literacy around the world; new patterns began to emerge, marking out regions as objects of inquiry, defining nations in terms of deficit, causing alarm and creating funding priorities. Advancing 'the common welfare of mankind', as per UNESCO's mission, demanded that the welfare of mankind be understood in commonly held terms. Statistical 'norms', as Desrosières (1998) pointed out, soon begin to suggest desired conditions, create new aspirations and even provoke new moral norms.

Numbers translate realities into their abstract versions, and these abstractions can then be imposed on realities themselves. Scott (1998) has presented an excellent example of this process through his historical tracing of scientific forestry practices in Prussia. Efficient and scientific forest managers recognised the fiscal value of timber, and their measurements ignored the rest of the forest – the shrubs with their medicinal offerings, the biodiversity which made forests resilient, the insects and other life forms which were necessary for enriching the soil and so on. As Scott tellingly observed, this translation of 'nature' into 'natural resource' began to create pictures in government tables and charts of an abstract forest that was devoid of biodiversity and that ignored the multiple uses of the forest to the communities around them, as well as the complexity of the forest ecology. Eventually, the Germans began to impose the abstract forests of their audit books onto reality itself: land was cleared and forests began to be planted with single species of the same age in neat rows so that they could be inspected, monitored and harvested with great ease. Neat and uniform forests became the new aesthetic of forestry.[3] Verran (2010) provides a fine-grained analysis of the performativity of numbers in her study of the enumeration of Australian waters. She finds that numbers are used in indexical and symbolic ways to represent 'the ecological health of Australia's creeks and rivers, lakes and billabongs', and that a shift to an iconic use of numbers makes possible the constitution of a 'water market'.

But it is not just that measurement changes the world once it is performed; importantly, *the world has first to be changed in order that the measurement becomes possible.* As Ted Porter put it: