



Translating PISA, translating the world

Camilla Addey^a and Radhika Gorur ^b

^aGEPS – Globalisation, Education and Social Policies Research Group, Department of Sociology, Autonomous University of Barcelona, Bellaterra, Spain; ^bLaboratory of International Assessment Studies, Deakin University, Melbourne, VIC, Australia

ABSTRACT

The OECD is extending the participation of low- and middle-income nations in its Programme for International Student Assessment (PISA). To explore how PISA can be made more relevant to these contexts, a pilot study, PISA for Development (PISA-D), was launched. Translating PISA into PISA-D required the development of instruments that had relevance to the new contexts while maintaining comparability across all PISA participants. Drawing on Science and Technology Studies and Callon et al's theory of the three stages of translation of research, and based on detailed empirical data, this paper describes how the technical and the political, and the material and the semiotic, work together to make PISA 'fit' new contexts, while at the same time making the new contexts 'fit' PISA. This paper demonstrates how international comparisons demand profound changes in the ways countries come to know, represent, and act upon their education systems.

转译PISA, 转译世界

摘要

世界经合组织正在拓展低等和中等收入国家在其“国际学生评估项目”(PISA)中的参与。为了探究如何使PISA项目更符合这些国家的情境,经合组织推出了“PISA促进发展”(即PISA-D)这一试点项目。将PISA转译成PISA-D需要开发与新情境相关的测试工具,同时还要维持所有PISA参与国(或地区)之间的可比较性。本文借鉴科学技术研究以及Callon等人提出的转译研究的三阶段理论,结合详细的实证数据,描述了PISA如何在技术和政治以及物质和符号这几个因素共同作用下“适应”了新情境,同时也使新情境“适应”了PISA。本文证明了国际比较如何要求这些国家在对其教育系统的认识、展示和采取相关行动的方式上发生深刻变化。

KEYWORDS

OECD; PISA for development; science and technology studies; socio-technical global metrics; international large-scale assessments; cross-national comparisons

关键词

经合组织; PISA促进发展; 科学技术研究; 社会技术的全球测量; 国际大规模评估测试; 跨国比较

Introduction

The OECD states that participants in its Programme for International Student Assessments (PISA)¹ make up nearly 90% of the world's economy (OECD 2009), and that PISA has become the global yardstick for education (OECD 2013). However, a look at the participation patterns in PISA shows that, while certainly international, PISA is far from 'global'. Until very recently, large parts of the 'global south'² were missing from its participants' list. Many middle- and low-income countries that had participated in PISA, had done so

sporadically or as a one-off exercise. Moreover, in those middle- and low-income countries that *did* participate, the influence of PISA on policies and practices has been minimal (Lockheed, Prokic-Breuer, and Shadrova 2015).

With Sustainable Development Goal 4 (SDG4) drawing global attention to the unacceptably low learning outcomes in much of the global south, PISA is now seeking to almost double its participants to reach 150 countries by 2030 by expanding the number of low- and middle-income nations in its fold. Since PISA was originally designed for high-income OECD member nations, this ambitious expansion plan required the OECD to consider how to encourage more nations to participate – and participate regularly – in PISA, and how to ensure that education policies and practices in these nations could be influenced by PISA.

In response to these twin challenges, the OECD developed a pilot study called PISA for Development (PISA-D) – a version of PISA that was specifically adapted to the contexts of middle- and low-income nations. Over the course of 2017 and 2018, PISA-D was piloted in eight low-income nations – Cambodia, Ecuador, Guatemala, Honduras, Bhutan, Paraguay, Senegal, and Zambia. The PISA-D results were announced in December 2018. Of the countries that participated in PISA-D, Cambodia, Guatemala, Panama and Paraguay have now signed up to participate in PISA in 2021, whilst Ecuador and Zambia have confirmed participation in 2024 and Bhutan, Honduras and Senegal have signalled their intention to participate.³ PISA-D will become part of the PISA 2021 suite of offerings, and will be known just as PISA (rather than PISA-D), and offered as a paper-and-pencil test. What we currently known as PISA will be an electronically administered, adaptive test. We can thus deduce that the OECD and the nations participating in the PISA-D pilot believe that the PISA-D pilot was successful in achieving its aims.

This paper focuses on the development of PISA-D and the ways in which the OECD went about the challenging task it had set itself – of developing a test that was deemed more suited to low-income nations and better able to influence policy in these nations. Our empirical data were gathered across three research projects: [PISA for Development for Policy, funded by Fritz Thyssen Foundation] (Camilla Addey), and [PISA for Development: Between Standardisation and Recontextualisation (Heidelberg Academy for Sciences and Humanities, Science, Numbers, and Politics project – 2016–2017) and Accountability and global education policy networks in the Indo-Pacific (Australian Research Council, 2017–2019)] (Radhika Gorur). The research projects span the period between 2014 and 2019, during which PISA-D was being developed and implemented, and its data analysed and disseminated. All three projects gathered data through: (a) document analysis (*Capacity Building Plans, Needs Analysis Plan, Communication Strategies*, but also working documents – mainly power points – made available on the PISA-D website at the end of each of the main meetings; (b) interviews with OECD officials, PISA-D contractors, and high level government actors and education stakeholders in Ecuador, Paraguay, and Cambodia; and (c) ethnographic observations of PISA-D meetings and social events in Paraguay, and of PISA-D data dissemination activities in Cambodia.

For our analysis, we draw on key concepts from Science and Technology Studies (STS), and in particular on the description of contemporary scientific practice in terms of the *three stages of translation* proposed by Callon, Lascoumes and Barthe (2011). Observing

these efforts to translate PISA into PISA-D provides an opportunity to understand the making of science, or describing *science in action* (cf. Latour 1987). But at the same time, it is also an opportunity to study the *politics* involved in developing large-scale innovations that aim to influence practices that may affect millions of learners across a diverse range of nations. Choosing to treat the development of PISA-D as a ‘scientific’ practice is a deliberate, political decision on our part. PISA’s normativity and its efforts to promote an economistic approach to education and further the course of particular types of datafication and accountability have been extensively studied (i.e. Fischman et al. 2018; Grey and Morris 2018; Baird et al. 2016; Auld and Morris 2016; Martens, Niemann, and Teltemann 2016; Sellar and Lingard 2013; Gorur 2011, Gorur 2016; Addey 2018, Addey 2019; Addey et al. 2017; Grek 2009). OECD’s influence on education policy has also been analysed in terms of globalisation and the travels of policy ideas and practices (for example, Verger, Novelli, and Altinyelken 2018; Steiner-Khamsi 2017, 2010; Waldow 2009; Rizvi and Lingard 2010). These and other scholars have shown that global policy initiatives such as PISA-D are driven by the wider political and power context – such as the strategic visions of organisations like the OECD, World Bank, and bilateral donors; the OECD’s new paradigm for development (Auld and Morris 2016); the dominant cognitive-economic model of education (Auld, Rappleye, and Morris 2018; Komatsu and Rappleye 2017), and the growing pressure to participate in large-scale comparative assessments bypassing government and multilateral donor concerns (Auld, Rappleye, and Morris 2018).

We contribute to this critical literature, building on the work of Latour and other STS scholars, by drawing attention to the less frequently described *politics of science*. We argue that the making of PISA-D was not just a *technical* exercise of translating survey instruments of PISA and making them ‘fit’ for low-income nations; the new low-income contexts were also translated and made to fit PISA – and this *political* exercise had to go hand in hand with the technical one – without the political part, the technical could not succeed. Whilst studies have shown how governments are affected by the *results* of PISA, we demonstrate that profound changes are required even for *participation* in PISA, even before the assessment is implemented and the results announced.

In keeping with the ethnomethodological flavour of our empirical engagements, we tell the story from the perspective of the actors, without presuming to judge them or to analyse what lies *behind* their statements, since we have no ‘god’s-eye-view’ that is superior to that of our informants about their own theories and intentions (Latour 2005). Rather than point out the politics of the actors, STS tries to allow actors to speak for themselves. This can lead to our own expressions, as authors, becoming somewhat deadpan – the better to allow the politics of science and of the actors involved to be perceived by the reader.

In the next section, we provide an elaboration of our conceptual framing. Then, we provide an overview of PISA-D, and empirically illustrate the three stages of translation in action. We conclude with a reflection of the usefulness of a socio-technical approach to illuminate how technologies and politics work together in ambitious contemporary exercises of developing global metrics.

Assessment in an uncertain world

Exclusion, expertise and the regime of exactness

At the heart of our paper is the question: what does it take for sociotechnical innovations to succeed? How do new practices or artefacts come to be developed by scientists and innovators, and how do they come to be accepted by others? To answer these questions, it is helpful to begin with a very brief history of the production of scientific truths presented by Callon, Lascoumes, and Barthe (2009) in *Acting in an Uncertain World*. Following Licoppe (1996), Callon and his colleagues distinguish three great periods of scientific truth production, each marking a specific type or location of 'laboratory' (where science was made); a particular relationship between the laboratory and the world; and differences in the extent and manner in which 'experts' and 'laypersons' came to engage in the making of science.

Breaking from Aristotelian science that relied on shared, empirical experience for validation (i.e. the expert had no more access to the truth than the layperson), in the seventeenth century Regime of Curiosity, scientific truth was sought through incredible, elaborately set up, one-off experiments legitimised by the validation of a restricted number of highly ranked and learned individuals who acted as witnesses (see also Shapin and Schaffer 1985). Science was not part of everyday experience, but was produced through specific, spectacular events in confined spaces to which only the privileged had access.

In the eighteenth century regime of utility, the production of science was validated by the utility of facts and depended on standardised instruments that allowed for the reproduction of phenomena across contexts. Replicability, made possible by the range of mobile, standardised instruments that were invented during this time, was a hallmark of good science, and the goal was generalisability, enabled through standardisation.

Towards the end of the eighteenth century, these regimes were displaced by the Regime of Exactness in which we now operate, according to Callon et al. In this era, science has increasingly become an isolated exercise, where the division between experts and lay persons is starkly defined and maintained. The sites of scientific endeavour, physically or metaphorically, are remote from the rest of the world. In these secluded spaces, science, characterised by purity, precision, and exactness, is made. Only an inner circle has access to the laboratory and the instruments to make fresh knowledge. Access to funds and the specialised infrastructure for validating their knowledge – instruments and equipment, conferences, peer review, etc. – sets scientists apart from the lay public, which can only access 'frozen knowledge'.

The three stages of translation

In this Regime of Exactness, Callon et al. delineate three 'stages of translation' in the doing of science, which we present briefly here, and elaborate in the next section. In the first translation, scientists translate a complex, abundant, colourful, noisy world – the 'wild' – into objects of manageable size and diversity which they can transport to their secluded laboratories, where they can work in peace, surrounded by their instruments and chemicals and expert colleagues. The second translation involves the development of new knowledge or objects – an effect of being able to study the reduced world. Soil

samples can be analysed, lab rats injected with various substances, particles made to reveal themselves. At the end of this second translation, scientists are able to state a scientific truth or fact, or produce an object. But the seclusion of science is only a virtue during the second translation. Unless scientists are able to influence the world through their science, their labours would be in vain. Unless they impact the world and induce others to take up their knowledge, they will be hard pressed to get funded for further research, gain citations, enhance their reputation or push an agenda. So in the third translation, scientists must re-enter the world and interest a series of actors to become their allies in engaging with and promoting their invention or discovery. This requires that the world is, in some way, 'laboratorised' – turned into a site or environment that supports the entity or facts produced in the carefully controlled, secluded laboratory.

The analytical trope of the three stages of translation lends itself most usefully to study the making of PISA-D. The OECD can be conceptualised as a secluded laboratory – the PISA laboratory – a space that can only be accessed by a restricted group of experts with specialised knowledge (psychometricians, statisticians, survey design specialists, and other experts involved in producing PISA-D). They are distributed across the OECD secretariat, PISA contractor agencies (like Educational Testing Services, Pearson, Westat); the PISA Governing Board; expert groups and the national project teams. They also include non-human actors like Item Response Theory, literacy frameworks and education standards. Education stakeholders, including teachers and students, do not enter this space, except as data samples. The OECD's offices, even while being in the heart of Paris, are as secluded and exclusive as any specialised laboratory. They are as difficult to enter – without a sponsor, security clearance, an identity card and pass, you cannot cross the electronic barrier. This is, of course, a conceit on our part, because it is not only a matter of buildings, and PISA-D is not produced exclusively in the OECD building in Paris. Its exclusivity is in the expertise, its networks, in the funds and the political capital it has garnered, and in its ability to exclude or displace other opinions and actors.

In the next section, we study the production of PISA-D in terms of the three stages of translation and explore how science and politics are folded into each other, and at the same time how divisions come to be drawn between science and politics, preserving the objectivity, purity and exactness of the statistical work whilst relegating the messy aspects to the domain of 'politics' (Latour 1987; Gorur 2015). As we argue in this paper, without the 'political', 'Science' (the capital S is used here to distinguish this specialised knowledge from the process of science – see Latour 1999) would not stand a chance of working in the new contexts in which it tries to establish itself.

Translating PISA, translating the world

When the OECD became interested in extending PISA to more low-income nations (OECD 2013), it commissioned Lockheed, Prokic-Breuer, and Shadrova (2015) to examine the barriers and challenges to the optimal use of PISA in lower-middle and middle-income nations that had previously participated in PISA (Ward and Zoido 2015).⁴

Lockheed et al identified three types of barriers. The first was cost. Even where the fee charged by the OECD is paid by donors, there are a number of in-country costs associated with field-testing, translation, implementation, marking, the training of staff, transport, etc. for which donor funding is not guaranteed. The second set of challenges they classify as

‘technical’ – challenges of instrument development, analysis of data, sampling, and working to a common global timeline and standard. The third challenge they diplomatically describe as ‘context challenges due to a country’s political economy, regulatory environment and national culture’ (Lockheed, Prokic-Breuer, and Shadrova 2015, 43). This is shorthand, we conjecture, for the lack of political will, convoluted and inefficient bureaucracies, an irreverent approach to systems and the maintenance of deadlines; and, not to put too fine a point on it, possibly corruption. They conclude that for PISA to be effective in low-income settings, the survey would require significant changes and adaptation, and also that the ‘capacity’ of PISA-D nations had to be enhanced so that PISA could actually ‘work’.

From the outset, then, it was acknowledged that the challenges involved were both ‘technical’ and ‘political’. The technical challenges included such questions as: how will the instruments be adapted? How will samples be drawn and standardisation ensured? How will comparison with regular PISA be maintained? Political questions were around how the ‘contextual’ or political situation would be managed, how the financial challenges would be resolved, and how it could be ensured that policy makers confronted the results that PISA would reveal and acted upon them rather than brush them under the carpet (Gorur and Addey, [forthcoming](#)). But these issues were not divorced from each other. The technical task of adapting PISA to fit the world of low-income nations, is, at the same time, a political process of adapting the world of low-income nations to fit PISA. We demonstrate this dual movement – of translating PISA to PISA-D to fit low-income nations, and translating low-income contexts to fit PISA-D – using the concept of the three translations described above.

The first translation: from the wild to the laboratory

The first [translation] is that of the reduction of the big world (the macrocosm) to the small world (the microcosm) of the laboratory. (Callon, Lascoumes and Barthe 2011, 48)

In the first translation, a research collective ‘grasps hold of the macrocosm’ (Callon, Lascoumes and Barthe 2011, 49) – described as big, wild, complex, enigmatic, scattered, unrelated, heterogeneous, abundant, colourful and varied – in order to simplify it, reduce it to a manageable scale, and make it amenable to being transported to the laboratory. A vast rainforest may become a few samples of soil and leaves and twigs which can be easily carried to a comfortable laboratory, far away from the hot and uncomfortable rainforest (Latour 1999). They could be further reduced – into photographs and drawings, chemical smears or colour coded inscriptions that can be stored easily or even emailed or otherwise transmitted to distant colleagues without deteriorating, unlike the actual twigs and leaves.

Translation 1 was quite challenging in PISA-D nations, despite the years of experience with PISA. There were significant differences between PISA nations, but the ‘wilds’ of the PISA-D nations presented a whole different level of challenge. To grapple with this challenge, in 2013, even as the Lockheed Report was in progress, the OECD began to assemble a group of contractors, advisors and consultants around the PISA-D initiative. This team was closely connected to the main PISA research collective, and the assemblage included former and new contractors (ETS, Pearson, Westat, Microsoft, The Learning Bar, Positivo

BGH). It drew in aid and development agencies such as global south NGO Pratham, the UNESCO Institute of Statistics (UIS), technical partners, experts in lower and middle income contexts, county representatives from the participating countries, and also non-human actors: the PISA tests, Item Response Theory, sampling procedures and so on. Together, the research collective set out to grasp hold of the complex macrocosm of the eight PISA-D nations. The question was, which parts should be grasped and how could they be reduced?

To begin with, as Lockheed et al. put it succinctly, 'The student being assessed in middle and low income nations is different to the student assessed in OECD nations' (Lockheed, Prokic-Breuer, and Shadrova 2015, 88). Unlike PISA 15-year-olds who were conveniently mostly in school, distributed between a couple of grade levels, many PISA-D 15-year-olds were out of school and were distributed across the landscape – on farms, in factories etc. Tracking them down would be no easy matter:

In Africa I suspect these kids are still in villages, they're doing agricultural work, some of them – they're not, they're not necessarily concentrated [in one place]. They're still with their families, I think, that's my guess – it's not clear which of these [survey methods] is going to work out ...
(Interview transcript, PISA sampling expert)

But going house to house looking for a 15-year-old, defined in PISA as being between 15 years and 3 months and 16 years and 2 months, can be inefficient. So the definition of 'student' in PISA-D had to be redefined to include a larger age range – 14–16-year-olds – to make the household surveys more efficient.

Those who were in school were scattered across several grade levels – some as low as Grade 5 – in contrast to their high income counterparts. Many were located in communities where a number of languages were spoken, with different languages introduced at different grade levels. Deciding on which languages would be most appropriate for testing, especially for testing the out of school children, and the funding of multiple language translations, also posed a challenge. While PISA students score more or less evenly across the spread of difficulty level in the PISA test items, which range from the easiest items (Level 1) to the hardest (Level 6), most children in low-income nations tend to score at Level 1 or below. Using the same cognitive survey instruments as in PISA would mean that the few items in Level 1 could not sufficiently describe students' performance. The cognitive surveys needed to be adapted. Similarly, the background survey for students was based on socio-economic conditions in OECD nations. It needed to be adapted – and to do this, researchers from The Learning Bar, the contractor for developing the background survey, would need to visit these countries or otherwise learn about them to determine how to survey them.

So, despite two decades of experience with PISA, and although many of the same contractors were used in both PISA and PISA-D, the research collective had to start more or less from scratch – beginning with asking, as one PISA sampling expert explained in an interview: 'What is a school?' and 'What is a student?'. So the research collective was faced with many decisions: defining student populations (who is in and who is out), choosing languages of assessment (which languages count and which do not), selecting schools (what can be defined as a school and what cannot), and finding out of school children (drawing a threshold between in and out of school – for example, children who were in school but in Grade 6 or below at age 15 were considered 'out of school'). They were

faced with the tasks of constructing literacy, numeracy and science skills (identifying what levels of difficulty were appropriate), defining educational backgrounds (i.e. school and community violence levels, home and school possessions), and so on. These decisions are finalised at International Advisory Group and Technical Advisory Group meetings where the research collective gathers (often in Paris) and are based on: commissioned studies by experts, study trips of contractors to the PISA-D countries, the many trips of the country representatives and experts to the offices of the PISA-D contractors, studies of country capacity to implement PISA-D, consultations between the contractors and National Project Managers on the instruments by email, software programmes keeping all actors on track with their tasks, instant communication software allowing weekly online meetings amongst all actors, and hundreds of power point slides from each meeting.

Through this process of many small and several major decisions, the research collective transports the wild world back to the laboratory. The complex world of children and schools is reduced through sampling to a manageable 'data set', different literacies are reduced to a few test items, and varied contexts and learning possibilities are pruned down to a restricted set of standardised options (cf. Gorur 2011). Of course, there are strict protocols that ensure that the samples that enter the laboratory represent the world they leave behind. It is this connection between the sample and the world – what Latour calls 'circulating reference' (Latour 1999) – that gives the laboratory its strange power (Callon, Lascoumes and Barthe 2011, 49). The process of selecting, reducing and transporting is the first transformation of the wild world.

The second translation: the mysterious manipulations in the laboratory

The second stage is that of the formation and setting to work of a restricted research group that, relying on a strong concentration of instruments and abilities, devises and explores simplified objects. (Callon, Lascoumes and Barthe 2011, 48)

It is important to note the overlap between the first and second translation. The research collective is already gathered to make the first translation feasible. But the first translation may require a number of actors with less specialisation, such as the in-country fieldworkers and survey administrators whose role is that of implementer rather than decision-maker, and who do not have access to the laboratory.

In the second translation stage, a restricted research collective is formed *inside* the laboratory. Here we find actors like Westat with their school and student lists, cApStAn with experts in new languages, The Learning Bar with descriptions of diverse backgrounds and policy priorities, and also selected experts, such as the Questionnaire Expert Group, national project managers with expert 'local' knowledge, OECD staff with their main PISA instruments, and technical partners who help make expert decisions on issues such as 'out of school children'. The actors are joined by a set of non-human actors such as the main PISA instruments and procedures, software and psychometric theories.

The way the questionnaires are adapted and made ready to travel to the wild again are illustrative of these manipulations inside the laboratory. The cognitive questionnaires need to be expanded at the lower end of performance to provide more usable information in

low-income nations. But there are not enough funds to develop new items. So items from PISA for Schools, the Literacy Assessment and Monitoring Programme, the Southern and Eastern Africa Consortium for Monitoring Educational Quality and other tests are summoned to the PISA lab. So are experts who can pronounce whether the items are suitable or not:

The expert group who are assigned to review the items – they look at the items and they look at the framework and see if they think that the items fit well in terms of the definition and in terms of the levels as they are already defined for PISA. In the process of doing that, we have expert group meetings, and we also involve the countries in reviewing items and there we question whether or not these are items that seem to fit in terms of their curriculum, and their students, but first we look at the framework. (Interview transcript, PISA questionnaire expert)

Secluded in their laboratory, just as Callon, Lascoumes and Barthe (2011) explain, they manipulate, stabilise and interpret the world ‘often through heated debates and discussions’ (56). Challenges are identified and discussed: if the ISCED (International Standardised Classification of Education) levels used in the main PISA do not fit the context of Paraguay, should they be changed? If the socio-economic index in PISA does not fit the socio-economic index in Ecuador, should it be revised? OECD officials describe the test items being ‘forged in a microcosm or a mini process’ (Interview extract, OECD staff) in which the restricted research collective is constantly challenged by the PISA-D Technical Advisory Group to defend each decision and that ‘through that kind of technical discourse we find, I think, the right way’ (Interview transcript, OECD staff).

In this confined space, they rely on instruments and on formalised procedures and protocols to devise and manipulate the simplified worlds they have brought back. The inscriptions they fabricate are perfected, stabilised and interpreted. At the same time that they are adapting the instruments to work in low-income contexts, they have to worry about maintaining comparability with the main PISA:

The main reason for keeping so many [60%] PISA items is the advantage of being able to link to PISA and having these items all being on the same scale. So that was the driving force – to have at least half – and we did choose 60%. What was the most difficult domain was science – because aside from PISA for Schools, there were no other existing items we could use. (Interview transcript, PISA questionnaire expert)

The expansion of Level 1, required to suit the characteristics of PISA-D students, and the policy and practice needs anticipated in PISA-D nations is worked upon:

We need to have enough items to tease out those key knowledge levels – not just Level 1, but 1a, 1b and 1c – they each have very distinct descriptions – we’re hoping to have enough items to really tease those differences out and be able to say ‘here we have a certain percentage, not just at Level 1, but in this particular level with these types of characteristics’. (Interview transcript, PISA questionnaire expert)

Our interviewees also provided other excellent examples of the collective at work. The translation team collaborates with the country level teams to ensure the test items are rigorously translated and ‘to find the right word and making sure it is the same register’ (interview transcript, PISA questionnaire expert). The country experts provide valuable input in this instrument development process:

You know there are a few items that we saw that seemed very specific to knowing something about cars or roads or something ... there's a certain type of cultural knowledge that a couple of countries said 'I'm not sure that this will be as appropriate to our assessment' and if we could find an alternative, we didn't use it. (Interview transcript, PISA questionnaire expert)

As the instruments are perfected in the laboratory, the restricted research collective is aware that 'if they have done their work well, it is a world that can be connected with the big world from which they have taken care to keep their distance' (Callon, Lascoumes and Barthe 2011, 50; cf Latour 1999). Ensuring this connectedness with the big world requires a constant back and forth between the restricted research collective and the big world:

They [the private contractors] have gone to the countries, they have reviewed the frameworks, they have finalised the framework for the questionnaires, they reviewed the items, inspected the items, helped to choose the items from the other assessments and basically checked it with the countries at every stage: 'Are we in agreement? Can we go forward? Are we in agreement? Can we go forward?' (Interview transcript, OECD staff)

Callon et al. describe the restricted research collective as occupying a strategic position 'that controls access to the world and to the discourses about the world' (Callon, Lascoumes and Barthe 2011, 52). Undisturbed, they apply their specialised knowledge and perfect the instruments so they will work universally and without error, eliminate all human bias and external intervention, settle relations of domination and exclusion (Callon, Lascoumes and Barthe 2011, 4) and produce enriched ways to better reveal education systems, understand them better, and act on them better. These decisions and actions have enormous ontological consequences – but their politics remain hidden behind the technical veneer.

But this seclusion is also an illusion. The experts have to make a range of concessions inside the laboratory that have everything to do with the outside world. To start with, the funding limitations require them to make do with questions borrowed from other tests, rather than create a bank of new items. They cannot afford to stick to the age grouping because it is too expensive to test the out of school children precisely. So the whole idea of making highly sophisticated assessments available to low-income nations takes a bit of a hit as compromises begin to be made. Perhaps the biggest compromise is that some of the PISA-D nations opted out of Strand C – the survey of out of school children (Guatemala, Honduras, Panama, Paraguay and Senegal participated but not Zambia, Cambodia, Bhutan and Ecuador). Since children being out of school is one of the biggest issues confronting these nations, leaving them out of the survey would likely mean that they would escape policy attention further.

And of course, the whole world of the *original* PISA – from which PISA-D wanted to distinguish itself, and which, it argued, was so unsuitable to low-income contexts that a whole new assessment needed to be developed at great cost and involving hundreds of people distributed around the world – this world is never far away. It is a looming presence in the 'secluded' PISA-D laboratory, because the requirement of maintaining comparability with PISA nations is ever present during the phase of instrument development.

Quite literally hundreds of experts from all the countries coming together every three years. [...] Those hundreds of experts from those countries, reflecting every three years and formalizing frameworks, that has built the consensus, so PISA-D basically inherits that consensus, it

inherits that because we are using the existing frameworks but we are trying to extend them down. (Interview transcript, OECD staff)

The assessments needs of low-income nations are thus already framed within the needs of the OECD nations.

Towards the end of this second translation phase, the new knowledge gained in the laboratory is prepared for release into the big, wild, world. Reports are prepared keeping in mind the sensitivities of the receiving publics, composed of non-experts and political actors who do not have the benefits of the specialised knowledge of the scientists in the laboratory. This carefully monitored and greatly standardised writing of the data interpretation was carried out step by step, word by word inside the laboratory – ensuring a greater hold on the instruments before they are let out into the wild world. In the regular PISA, the data are released to countries and in-country experts write the country report. But given the sensitivity of PISA-D and the OECD's interest in its success (which would impact its ability to realise the ambition of expanding to 150 nations and pushing forward PISA as the SDG4 benchmarking metric), national experts were invited to Paris to write the data analysis under close supervision of the OECD policy analysts.

The third translation: laboratorising the world, or making the world fit PISA

The third stage is that of the always perilous return to the big world. Will the knowledge and the machines produced in the confined space of the laboratory be able to survive and live in this world? (Callon, Lascoumes and Barthe 2011, 48)

Once the PISA-D instruments have been perfected and their interpretation stabilised in the laboratory, their return to the world needs to be organised. This process requires many resources and many allies. In this return, fraught with dangers, the research collective carefully protects the new knowledge as it is bound to meet adversaries who will resist it. The sequence of translation stages may give the false impression that these stages are linear but the stages overlap. In particular, the third translation is clearly on-going during the first and second translation stages:

We put a lot of resources into preparation of the country, every country had a needs analysis, a plan, we got everybody up to speed in terms of what to expect, we gave them a lot of support, the contractors had additional resources to hold the hands of the countries, more international meetings than we usually have to do. (Interview transcript, OECD staff)

The laboratorisation of the low-income contexts occurs throughout the PISA-D development phase. When signing up for PISA-D, the first step is to complete a *Capacity Needs Analysis* (CNA) which was described as a necessary step to establish 'the ability of the individuals and institutions responsible for the project in each country to perform the necessary functions' (*PISA-D: Capacity Needs Analysis Ecuador 2015*, 7) to implement PISA-D. A key aspect of this phase is that national actors, often including the minister of education and officials of allied ministries, such as finance and labour, as well as development partners are brought together for several meetings and workshops. By involving a range of actors, including universities, other ministries and so on, the PISA-D network becomes thicker and more entangled. Getting the minister to chair these meetings

ensures that high-level commitments are clearly and publicly made, thus reducing the chances of their being ignored if the results, when released, are not appealing.

The CNA and the capacity building plans were perhaps the most obvious attempts to domesticate PISA-D nations. It included evaluations of whether individuals are resistant to acquiring skills, the general level of acceptance of assessment data by the general public, the quality of the network of the main implementing body (key to enrolling allies and displacing resistance), the willingness to learn and develop PISA methodology, the on-going efforts to sway public opinion on the significance and acceptance of the data (particularly through the press and social media), the extent to which 'technical expertise' is challenged, and how to minimise local variances (e.g. absenteeism during harvesting seasons). The analyses can be blatantly political. For example, the Ecuadorian CNA states that 'the risk of having the questions and the methodology publicly exposed and criticized by the media is very real' (*PISA-D: Capacity Needs Analysis Ecuador*, 2015, 22) but also that there are no groups (e.g. unions) that are formally opposed to assessments.

Training processes are an important part of domesticating PISA-D nations and making them fit for PISA. PISA's (and PISA-D's) requirements are extremely specific. Their manuals are very detailed. This specificity and detail are helpful to peer partners who support the nations taking PISA-D. Based on this manual, and their own experiences in their country, an official from Korea, the peer partner for Cambodia,⁵ explained how they prepared a training plan, working out who needed to be trained and what the content of the training should be for each stage in the process. For implementers, for example:

[T]hose who are actually piloting the implementation, they have to know what they are doing. The first [thing] I think is *why* they are doing [something], and [then] *what* they are doing ... Why it is important for you to keep to all the regulations, all the guidelines. You cannot skip this and that. Because those will affect the final results. So because all the guidelines and the in-country activities which are clearly specified by all the manuals of the OECD, I can easily set up the step-by-step capacity building to different targets. (Interview transcript, senior official, Korean PISA team)

In Translations 1 and 2, the experts in the laboratory have a greater control over how various actors deal with the instruments and the data and so on. But in the third translation, once the data and the facts are let loose in the world, it becomes much more difficult to control the interpretations. What will a certain percentage of students at level 1a on the PISA scale mean in countries as diverse as Cambodia and Ecuador? At this point, the restricted research collective is preparing for the return to the big world. The pedagogising, political agenda of PISA-D is nowhere more starkly evident than in the processes of ensuring that the nation pays attention to the results and responds through policy changes. There is a clear process of training, teaching, deciding what information should be released or disclosed, anticipating and neutralising opposition, and managing the response of different groups.

In order to ensure the PISA-D data is 'correctly' interpreted and disseminated, the restricted research collective pre-formatted a standardised national report to be filled in by each national policy analyst:

So to get up with the analysis quicker, we prepared first an outline of the national report, in terms of the structure, the headings, and that of course came from the assessment and analytical framework, basically put in place to guide the whole survey, and then we moved from the

outline of the report to actually producing templates, sections of chapters, template paragraphs where we did generic analysis of data that we were expecting to be using based on PISA data to produce the template, and then we translated the template into French, Spanish, Khmer, because everyone worked in their own languages. (Interview transcript, OECD staff)

The research collective prepares the ground politically through strategic communication, high-level meetings and high-visibility launches. They train the national project team to re-imagine their nation as a PISA context.

The laboratorising reaches a critical point when it is time to announce the results and drive forward the policy recommendations. This can end in disaster. One interviewee tells of one such disaster, when India had participated in the out-of-cycle PISA 2009 test. It did extremely badly, and so the nation was not prepared to own the results at all:

[India] see that they are barely better than Kyrgyzstan and nobody wants to own this, nobody wants to be part of this, so [that is] the opposite of all of that is what we are trying to do, and if we don't do all of that, we know we could be faced with seven Indian situations [with PISA-D in 2018] because we are under no illusions, we don't expect that Zambia, Senegal or Cambodia are going to perform better than Tamil Nadu and Himachal Pradesh. (Interview transcript, OECD staff)

Particularly when the results are unappealing, the media and the politicians have to be managed and the laboratisation reinforced:

[...] how you prepare a country to come into PISA, how you then do more than just enable the country to take part, you have to be with them all the way through, in terms of presenting to their population why they are in PISA, helping them to understand the results, and to present those results, and managing the fall out. (Interview transcript, OECD staff)

The research collective is open about the dangers ('the fall out') of a return to the world that has not been sufficiently laboratorised – where the instruments do not work because the world does not sufficiently resemble the laboratory:

What we don't want is a situation where the report is presented and then people start publicly disagreeing with findings, with implications. (Interview transcript, OECD staff)

So how can the world be made to resemble the laboratory? A key strategy is the displacement of obsolete laboratories that support contestation and resistance of the newly made world. For example, a fall out may occur if previously administered assessments (i.e. national assessments) hold greater legitimacy over PISA-D. This helps understand why throughout the entire PISA-D process, efforts were made to analyse national assessments with the aim of linking them to PISA-D, but also to repeatedly delegitimise all other assessments, such as regional assessments, as being below the PISA gold standard (Addey 2017).

At the same time and in order to recreate the conditions of the laboratory in the wild world, extensive capacity building was carried out in each PISA-D country to institutionalise the methodology of PISA-D and promote a culture of independent evaluations of educational systems. This resonates with what Callon et al. describe as implanting new laboratories.

Another key strategy the research collective relies on is mobilising resources and support and assembling as many allies as it can – a process requiring mutual realignments. The example below illustrates this process:

Having gone through each of the three or four key policy messages and identified who it is that is going to own it, we then worked with them about how they can develop a plan to go to that person and have some structured discussion about how to take this forward. [...] So what is elaborated there, as a suggested way forward is something the owners of this, whether it is curriculum, whether it is teacher education, administration, resources, whatever it is, that those who are responsible own it, they own it. [...] When the national report is launched [...] those bits of the Ministry or of the government which are responsible for each of those areas, are standing up and saying yes, absolutely, this is the way forward, this is what we must do, and that leads to a broad consensus in terms of how to move forward in terms of an evidence-based policy dialogue in education. (Interview transcript, OECD staff)

Having become, hopefully, indispensable to those who support the research collective's needs, PISA-D can return to the laboratorised world where there will be no adversaries or obsolete laboratories, but only allies standing up saying 'Yes, absolutely, this is the way forward'.

Conclusions

The concept of the three translations provides another theory to account for the success or failure of policy transfer, and how new policy ideas may be temporarily stabilised even in unlikely environments. The theory highlights how material-semiotic rearrangements are required for the world to be made to accommodate new ideas and practices. It makes visible how the technical and the political, the material and the semiotic work together to enact new worlds, often with far-reaching consequences. Moreover, our analysis demonstrates how politics can travel, like a stowaway, through templates and formats, software, protocols, standards, meetings, funding – these are as important in terms of the politics of such exercises as the more overtly political activities such as managing political expectations, the messaging to the press and the carefully coordinated spectacle of the release of the results.

STS commentaries aim to describe – to show rather than to tell. They often adopt an agnostic position about the desirability of technology or innovation, so that attention can be focused on the politics of such technologies and on the many political decisions and manipulations that are embedded in technological innovations, and the profound ways in which they impact societies. It takes focus away from the 'usual politics' that focuses on ideologies and interests, not because ideologies and interests do not count, but to elaborate how they come to count – through material-semiotic arrangements and the forging of relations between heterogeneous entities.

However, our politics is evident in our choice of topic. Like others (i.e. Auld, Rappleye, and Morris 2018), we hold that the entry of the OECD and PISA into the low-income space could prove to be a costly distraction, since the most pressing education policy issues in these nations are not at the secondary school level. Surveys in India show, for example, that 50% of fifth graders in India cannot read at Grade 2 level (Pratham 2015). Unless that is remedied, there is not much that can be expected from 15-year-olds. Engaging in a costly and sophisticated test of 15-year-olds, a large proportion of whom are not even in school – is

wasteful, and the scarce resources could be put to better use. Most of all, rankings based on aggregated data can often mislead policy, as has been shown even in the case of a statistically sophisticated nation such as Australia (Gorur and Wu 2015).

A key argument we are able to make as a result of such description is that the ‘influence’ of PISA is not something intrinsic to PISA – i.e. it is not that PISA has some intrinsic quality that makes its appeal so widespread and its message universally influential. For example, low- and middle-income nations that had previously participated in PISA were not particularly ‘influenced’ (Lockheed, Prokic-Breuer, and Shadrova 2015). Neither was, for example, India when it participated in 2009, as our OECD interviewee pointed out. Our account of PISA-D through the three translations demonstrates that for PISA to be influential, it has to be carefully assembled and orchestrated. Contributing to the sociology of numbers, this paper also demonstrates that for PISA to remain a ‘standard global metric’ that can be used for policy anywhere, in any context, the OECD has to laboratorise the world to fit its standards.

The thick descriptions provide an understanding of how intentions and ideas get translated, how compromises are made, how some objections are quelled, how reluctant actors are cajoled and mollified. These processes of cajoling and mollification, of compromise and alternatives, involves the translation of interests – and herein lies the politics. For PISA-D to ‘fit’ the world, it was not enough for PISA to transform itself by, for example, testing at levels below the lowest level 1. PISA-D had to simultaneously seek to translate the low-income world in ways that resembled the laboratory, carried in the PISA instruments and the policy ideas that inspired them. Descriptions of the capacity needs analysis and capacity building, and the careful scaffolding that involved gradually enrolling a range of actors – ministers, bureaucrats and development partners – to prevent the rejection of PISA results provide strong evidence of the ways in which PISA-D nations underwent consequential transformations even as PISA itself was adapted to these contexts.

Moreover, by sticking to this particular type of description – based on the information provided by actors who are most closely associated with the innovation – we are able to engage in a form of respectful critique. The respect extends both to the actors who so generously met with us and provided details of their activities, and to the readers, who can engage with the politics from their unique standpoints, without excessive interpretation from us. Education policy research has come to be extremely polarised – an issue which we have sought to address through various activities – developing forums for cross-disciplinary and multi-stakeholder conversations, developing a critical voice that does not deride or denounce, and using theories that seeks to elaborate and to open, rather than close down, discussions. In this endeavour, we find what eminent scholar Fazal Rizvi has referred to in informal conversations, as ‘a politics of derision’ to be unproductive – it only serves to preach to the choir, and forecloses meaningful engagement with development practitioners, assessment experts, governments, global policy influencers and intergovernmental actors.

This paper is particularly significant at a time when the OECD is turning its attention to low- and middle-income nations because it highlights the kinds of changes that adopting PISA-D imposes on newly participating nations. The OECD’s recent (2019) appointment of the Australian Council for Educational Research (ACER) to provide support to countries new to PISA provides a perfect illustration of our findings. ACER’s ‘PISA 2021 Core E’

'aims to prepare countries for the successful preparation and implementation of PISA during all stages of the assessment, from planning and contextualisation to implementation, analysis and reporting' (our emphasis).⁶ Although these requisites present themselves as technical requirements for implementing PISA-D, they are also political because they fundamentally seek to transform the ways in which the countries come to know, represent, and act upon their education systems.

Notes

1. PISA is an international, comparative assessment of the performance of 15-year-olds in a range of countries. The triennial survey produces international rankings of nations on student performance, as well as a range of other analyses, such as the effects of social and cultural advantage, school structures, and administrative arrangements on student performance. In addition to a cognitive survey (a test of reading, mathematical and scientific literacy – and optional additions like financial literacy), PISA administers a background survey completed by students to assess their socio-economic and cultural advantage, and a school survey completed by principals to provide information about school-based factors that may influence student performance. Some countries also opt for a parent survey.
2. We use this disputed term somewhat hesitantly. While a discussion on this terminology is important, it is not central to the argument in this paper. For that reason, we have used scare quotes here, at first usage to indicate our discomfort, but have dropped these in subsequent usage to make for easier reading.
3. Governments cannot actually sign up to participate in PISA 2024 until 2021.
4. https://www.waxmann.com/index.php?elD=download&id_artikel=ART101865&uid=frei.
5. PISA-D nations were paired up with 'partner nations' who supported the PISA-D nations in various ways, including through providing technical assistance and implementation advice.
6. <https://rd.acer.org/article/supporting-countries-new-to-large-scale-assessment>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr Camilla Addey is a Marie Curie Research Fellow at GEPS - the Globalisation, Education and Social Policies - research centre at the Autonomous University of Barcelona, Spain. Formerly, Camilla was lecturer in Comparative and International Education at Teachers College, Columbia University (USA), and a researcher at Humboldt University in Berlin (Germany). She researches international large-scale assessments, global educational policy, and education privatizations. Currently she is carrying out a research project 'ILSAINC' on the private sector's involvement in international large-scale assessments.

Radhika Gorur is an Associate Professor at Deakin University, and a Director of the Laboratory of International Assessment Studies. Engaging in the fields of sociology of quantification and critical data studies, she explores the social and political lives of data and metrics and the practices of accountability in education. Her research interests include education policy and reform; global aid and development; data infrastructures and data cultures; classroom research; and the sociology of knowledge. Currently, with a grant from the Australian Research Council, she is researching the role of global policy networks and the new accountability practices in the Indo-Pacific, with an empirical focus on Cambodia and India to develop principles for sustainable, participatory accountability practices.

ORCID

Radhika Gorur  <http://orcid.org/0000-0002-4528-0793>

References

- Addey, C. 2017. "Golden Relics & Historical Standards: How the OECD is Expanding Global Education Governance Through PISA for Development." *Critical Studies in Education* 58 (3): 311–325. doi:10.1080/17508487.2017.1352006.
- Addey, C. 2018. "The assessment culture of international organizations: from philosophical doubt to statistical certainty through the appearance and growth of international education assessments." In *Assessment Cultures – Historical Perspectives*, edited by C. Alarcón and M. Lawn, 379–408. Frankfurt am Main: Peter Lang.
- Addey, C. 2019. "The appeal of PISA for Development in Ecuador and Paraguay: theorising and applying the global ritual of belonging." *Compare: A Journal of Comparative and International Education*. doi:10.1080/03057925.2019.1623653.
- Addey, C., S. Sellar, G. Steiner-Khamsi, B. Lingard, and A. Verger. 2017. "The Rise of International Large-Scale Assessments and Rationales for Participation." *Compare: A Journal of Comparative and International Education* 47 (3): 434–452. doi:10.1080/03057925.2017.1301399.
- Auld, Euan, and Paul Morris. 2016. "PISA, Policy and Persuasion: Translating Complex Conditions Into Education 'Best Practice'." *Comparative Education* 52 (2): 202–229. doi:10.1080/03050068.2016.1143278.
- Auld, Euan, Jeremy Rapple, and Paul Morris. 2018. "PISA for Development: How the OECD and World Bank Shaped Education Governance Post-2015." *Comparative Education*. doi:10.1080/03050068.2018.1538635.
- Baird, Jo-Anne, Sandra Johnson, Therese N. Hopfenbeck, Talia Isaacs, Terra Sprague, Gordon Stobart, and Guoxing Yu. 2016. "On the Supranational Spell of PISA in Policy." *Educational Research* 58 (2): 121–138. doi:10.1080/00131881.2016.1165410.
- Callon, Michel, Pierre Lascoumes, and Yannick Barthe. 2009. *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge: MIT Press.
- Callon, M., P. Lascoumes, and Y. Barthe. 2011. *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, MA: MIT Press.
- Fischman, G., A. Marcetti Topper, I. Silova, J. Goebel, and J. L. Holloway. 2018. "Examining the Influence of International Large-Scale Assessments on National Education Policies." *Journal of Education Policy*. doi:10.1080/02680939.2018.1460493.
- Gorur, R. 2011. "ANT on the PISA Trail: Following the Statistical Pursuit of Certainty." *Educational Philosophy and Theory* 43 (S1): 76–93.
- Gorur, R. 2015. "Producing Calculable Worlds: Education at a Glance." *Discourse: Studies in Cultural Politics of Education* 36 (4): 578–595.
- Gorur, R. 2016. "Seeing Like PISA: A Cautionary Tale About the Performativity of International Assessments." *European Educational Research Journal* 15 (5): 598–616.
- Gorur, R., and C. Addey. *Forthcoming*. "Capacity Building as the 'Third Translation': The Story of PISA-D in Cambodia." In *World Yearbook of Education 2021: Accountability and Datafication in the Governance of Education*, edited by S. Grek, C. Maroy, and A. Verger. Oxon: Routledge.
- Gorur, R., and M. Wu. 2015. "Leaning Too Far? PISA, Policy and Australia's 'Top Five' Ambitions." *Discourse: Studies in Cultural Politics of Education* 36 (5): 647–664.
- Grek, Sotiria. 2009. "Governing by Numbers: The PISA 'Effect' in Europe." *Journal of Education Policy* 24 (1): 23–37.
- Grey, Sue, and Paul Morris. 2018. "PISA: Multiple 'Truths' and Mediatized Global Governance." *Comparative Education* 54 (2): 109–131. doi:10.1080/03050068.2018.1425243.
- Komatsu, Hikaru, and Jeremy Rapple. 2017. "A new Global Policy Regime Founded on Invalid Statistics? Hanushek, Woessmann, PISA, and Economic Growth." *Comparative Education* 53 (2): 166–191. doi:10.1080/03050068.2017.1300008.
- Latour, B. 1987. *Science in Action*. Milton Keynes: Open University Press.
- Latour, B. 1999. "On Recalling ANT. In Actor Network Theory and After." *The Sociological Review* 15–25.

- Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Licoppe, C. 1996. "La formation de la Pratique scientifique." *Recherche* 27 (290): 100–105.
- Lockheed, Marlaine, Tijana Prokic-Breuer, and Anna Shadrova. 2015. *The Experience of Middle-Income Countries Participating in PISA 2000–2015*. Paris: OECD Publishing.
- Martens, Kerstin, Dennis Niemann, and Janna Teltemann. 2016. "Effects of International Assessments in Education – a Multidisciplinary Review." *European Educational Research Journal* 15 (5): 516–522. doi:10.1177/1474904116668886.
- OECD. 2009. *PISA Data Analysis Manual: SPSS Second Edition*. Paris: OECD.
- OECD. 2013. *The OECD's Contribution on Education to the Post-2015 Framework: PISA for Development* (OECD and Post-2015 Reflection Series). Paris: OECD Publishing.
- OECD. 2015. *PISA-D: Capacity Needs Analysis Ecuador*. Paris: OECD Publishing.
- Pratham. 2015. *Annual Status of Education Report*. New Delhi: ASER Centre.
- Rizvi, Fazal, and Bob Lingard. 2010. *Globalizing Education Policy*. London: Routledge.
- Sellar, Sam, and Bob Lingard. 2013. "The OECD and the Expansion of PISA: New Global Modes of Governance in Education." *British Educational Research Journal* 40: 6.
- Shapin, Steven, and Simon Schaffer. 1985. *Leviathan and the Air Pump: Hobbes, Boyle and the Experimental Life*. Princeton, NJ: Princeton University Press.
- Steiner-Khamsi, Gita. 2010. "The Politics and Economics of Comparison." *Comparative Education Review* 54 (3): 323–342. doi:10.1086/653047.
- Steiner-Khamsi, G. 2017. "The Rise of International Large-Scale Assessments and Rationales for Participation." *Compare: A Journal of Comparative and International Education* 47 (3): 1–20. doi:10.1080/03057925.2017.1301399.
- Verger, Antoni, Mario Novelli, and Hülya Kosar Altinyelken. 2018. "A Revisited Introduction." In *Global Education Policy and International Development*. 2nd ed., edited by A. Verger, M. Novelli, and H. Altinyelken, 1–34. London: Bloomsbury.
- Waldow, F. 2009. "Undeclared Imports: Silent Borrowing in Educational Policy-Making and Research in Sweden." *Comparative Education* 45 (4): 477–494.
- Ward, M., and P. Zoido. 2015. "PISA for Development." *ZEP: Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik* 38 (4): 21–25.