

6 Governance by data circulation? The production, availability, and use of national large-scale assessment data

*Nelli Piattoeva, Vera Gorodski Centeno,
Olli Suominen, and Risto Rinne*

Introduction

Education quality represents a key discursive justification for diverse ongoing education reforms. Using data as a tool of governance (Rose & Miller 1992), quality assurance and evaluation (QAE) is a tool for attempting to reinforce central control at a distance while allocating more autonomy to local actors and simultaneously creating a need for and reliance on new experts and data infrastructures (Lawn & Segerholm 2011). Governance at a distance relies on the provision and translation of information about subjects, objects, and processes to the centres of calculation and power (Hansen & Flyverbom 2014). In this form, the new architecture of governance relies on data production and circulation (Ball 2016).

This chapter starts from the premise that contemporary education governance increasingly depends on and utilises quantitative data. However, paraphrasing Beer (2016), our assumption is that it is in the circulation of the outputs of measurement systems that we may see the unevenness and adaptability of governance at a distance through data, revealing something about the variations and messiness of the realities of the broader political orders to which data may belong. As Beer (2016) writes, frictionless data circulation is an ideal which often remains unfulfilled because of context-specific blockages and frustrations. In this chapter, we investigate what we term “the intended political project of data”, that is, how education policy documents and statements by officials depict where data is supposed to circulate, for what purposes, and how. We then look at whether data circulation meets its claimed goals and aspiration by analysing data production, data availability, and data use as distinct but very intertwined phases of overall data circulation. In following the scholarly appeal to study actual rather than assumed processes of governance by data (e.g., Selwyn 2016), this orientation advances one of the main issues raised in Chapter 1 and elaborated throughout the book, namely, the difficulty of policy-planning and the unrealistic promises made in the name of QAE instruments. In doing this, we follow the book’s abductive methodological premises to assist in teasing out the complex web of possibilities and limitations which exist in conducting (new forms of) governance. We thus treat governance attempts as potentially

incomplete and riddled with tensions and unfulfilled expectations and seek to understand existing governance at a distance better and in greater detail.

Brazil, China, and Russia have entered the era of education governance by data differently and at different moments, bearing somewhat different legacies of the link between knowledge and decision-making (see Chapter 3). We focus our analysis on data produced through one type of National Large-Scale Assessment (NLSA) in each country: the SAEB (the Evaluation System of Basic Education) in Brazil, the assessments conducted by the NAEQ (the National Assessment of Education Quality) in China, and the USE (the Unified State Exam) in Russia. Despite differences in the formats of these three assessments (see Chapter 7), their comparative analysis is compelling because they share three important criteria. First, they represent crucial sources of information for national policymaking; second, they are produced and used as quality evaluation tools; and third, they aim to make data comparable across space (among self-governing sub-national units in Brazil, semi-autonomous sub-national units in Russia, and regions with limited autonomy in China) and time (to determine changes in learning outcomes). The chapter opens with a theoretical note to problematise data circulation as a tool of governance. We utilise the term “data circulation” to argue that circulation unfolds through distinct but interlinked phases of data production, availability, and use. We see all these phases playing a role in the prospects of governance at a distance by data. The chapter then moves to our research’s empirical findings, which first address the intended political project of data and then the production, availability, and use of data explored under the respective sub-titles. The section on the intended political project of data uses official policy documents and interviews as empirical evidence, while the three subsequent sections rely mostly on individual interviews with officials and assessment experts.

Governance at a distance through data circulation

Scholars use the concept of governance at a distance to refer to modes of governance in which formal prescription is complexly absent, indirect, or enmeshed with more or less voluntary commitment to accountability, that is, to the submission of organisations and individuals to external performance measurement, which is often combined with (self)-evaluation (Rinne & Ozga 2011: p. 66). It relies on those at the centre having information about distant persons and events (Miller & Rose 1990). The concept of governance at a distance thus emphasises how the behaviour of governed actors is directed by the processes of collection and use of information by the authorities, who seek to direct the actions and behaviour of those targeted by these data. However, this form of governance does not replace but complements what might be seen as traditional governance, which has primarily relied on direct prescription (legislation and hierarchical principles absorbed in rules and procedures; see Maroy 2009). Various governance styles can therefore be integrated within the governance continuum. Researchers thus see governance at a distance as

reconciling “decentralized action (subsidiarity, self-responsibility) with centralized assessment (standardization) to facilitate exchange and valuation in vast spaces and to make long distance control something the actors aim to achieve by pursuing their interests” (Rottenburg & Engle Merry 2015: p. 22).

Effective governance at a distance necessitates calculations in one place to be linked to actions in another not through direct imposition but through assembling and connecting different actors and agencies in a functioning network. These actors, however, may often evince diverse, if not opposite, perceptions of and interests in data, reflecting the heterogeneous background networks which constitute them. These heterogeneous actors can represent both “producers” and “users” of data. To name a few, they comprise the commercial industry which produces software and hardware; the highly complex scientific expertise which enables the construction of test items and data analysis; policymakers who shape the demand for data and its application and set the necessary legislation which frames data collection, analysis, and use; and the multiple data audiences expected to use but which often also need to produce data (for example, teachers or officials who first collect data and then use its interpretations for various purposes). Government agencies, consultants, auditors, NGOs, and ICT personnel, among others, thus need to both supply and apply data to generate and consolidate networks as the means and channels of education governance (see also Hansen & Mühlen-Schulte 2012; Piattoeva 2015, 2016; Hartong 2016).

We use the notion of circulation to describe how data such as the NLSA needs to become integrated in practices, decisions, and processes (Beer 2016). Otherwise, data cannot govern (Grek et al. 2011). In other words, the power of measurements is located “not just in what they record or calculate, but in what then happens to those numbers, how they are used and by whom” (Beer 2016: p. 78). Three elements are thus of major importance in data circulation analysis: data production, data availability, and data use. Those who order and produce data contribute to defining policy problems and solutions; they frame the complex production chains and the horizon of policy interpretation when data are transformed into knowledge. However, researchers emphasise the role of the actual availability of data (Souto-Otero & Beneito-Montagut 2016). Data should be made accessible to and used by policymakers for evidence - based policy and as “a political strategy to make the policy process seem less political and more rational” (Steiner-Khamsi 2016: p. 585). Data also need to circulate widely across governmental-administrative levels and among different state and non-state actors to direct action (Piattoeva 2015). Specifically, they need to be employed by local governments and in schools to influence curricula and pedagogy through teachers’ and the local administration’s data usage (for this aspect see Chapter 8).

The circulation of numbers can enhance their legitimacy and power if they move and accumulate new networks of constituents, technologies, and things (Espeland & Sauder 2012: p. 86). The importance of data circulation has led researchers to call for a new sociology of numbers and quantification of

education which considers how numbers are “mobilized, circulated, consumed and contested” (Gorur 2015: p. 13). The flexibility, stability, and combinability of numbers, in contrast to the written or spoken word, are said to enable them to transcend contexts and find governmental roles in new institutions and often for other than their original purposes (see Rose & Miller 1992; Lascoumes & Le Galès 2007; Hansen & Mühlen-Schulte 2012). By examining the professed political project of data circulation and then looking at the actual production, availability, and use of NLSA data across the three national contexts, this chapter contributes to such a sociology of numbers in education, especially the debates about actual governance at a distance through data and the realities of evidence-based policy and QAE tools.

The political project of data circulation in Brazil, China, and Russia

Brazilian politics has historically favoured political decentralisation in education, while constantly projecting the construction of a coherent nationwide education system (see Chapter 3). The 1990s reforms deepened administrative, financial, and pedagogical decentralisation and motivated and justified numerical data as the main instrument for obtaining a national perspective of a disparate education system (see Kauko et al. 2016). According to the main national data expert body, the National Institute for Educational Studies and Research (INEP), NLSA data are intended to support policymaking and enable the monitoring of public policies at the three levels of government (municipal/state/federal), with the main goal of “contributing to the improvement of education quality, equity and efficiency. Besides, [NLSAs] seek to offer data and indicators about [the] factors [which] influence pupils’ achievements” (INEP, n.d.c).

NLSA data are assumed to circulate freely and widely in Brazil. They are considered public data: information which should be disclosed by the national provider and which may be freely used and reused, redistributed and re-appropriated without legal restriction. As an official from INEP put it, “[I]t is not reasonable here in Brazil, with our tradition, to say that you are not going to disclose that data” (BR-N-03). NLSA data are designed to “introduce information to the society” (BR-N-01) and “to produce clear and trustworthy information to managers, researchers, educators and the general public” (INEP, n.d.b). Besides this broad public aim, NLSA data are intended to circulate among policymakers working in both the executive and legislative federal branches, with the aim of informing education policies (INEP, n.d.b). NLSA data should also travel to sub-national units such as state and municipal governments, local councils, and schools (INEP, n.d.b). Interviewed policymakers and officials stated that NLSA data can and should trigger a vertical chain of reaction, resulting in improved teaching practices and education quality. As an official said,

[The INEP has] the task of collecting the information, and to handle that information and to render that information clear to the society. The

ministry has the task of taking that information and orienting policies and handling policies . . . The results that are produced here are often handled by teachers on television [the open broadcast TV School] . . . and that helps teacher training, helps to work in the classroom . . . There is an attempt to ensure that the teachers and directors should be able to use the results that come from these evaluations, that they are able to use [them] to change the teaching practice and the practice within the schools.

(BR-N-02)

Having closely followed international trends since the late 1970s, Chinese scholars started to raise concerns about the shortage of coherent and trustworthy data on the education system's performance. Against this backdrop, the Chinese NLSA, the assessments conducted by the National Assessment of Education Quality (NAEQ), was born. An interviewee explained,

[The NAEQ assessments were established] because there is no data. Because the data in Chinese education is largely a mess and there are many discrepancies. So, we need data from a place that we can trust. We need to get first-hand data.

(CN-E-02)

Although the system is still under development, the initial testing phase, which officially began in 2007, was phased out in 2015. Documents such as the Outline of China's National Plan for Medium and Long-term Education Reform and Development 2010–2020 (State Council of the People's Republic of China 2010) have elevated assessments to serve as the core element of the education system. The purpose of assessments is to understand the system better to implement evidence-based policy, which leads to qualitative improvements in education (Zhou et al. forthcoming). The previously mentioned reform plan for 2010–2020 (State Council 2010) also suggests that it is necessary to establish a national information database to develop the quality of education on the local level (see also Fan & Liao 2013; Rasmussen & Zou 2014; Zhang & Wan 2017). Assessment reforms take place amidst broader attempts to diversify evaluation approaches and consequently reduce the number of exam and score-oriented practices. Although in the initial experimental phase, NLSA results have been distributed only within the government, the plan is to release all the results (but not the original data obtained through assessments, a difference which will be elaborated on later) to the general public to increase accountability (as of 2017 no fixed time frame has been set). As one of our interviewees asserted, "I think it [releasing the results to the public] serves a purpose. You shouldn't keep data a secret, but it should be published. Only in this manner can we have true accountability. Only then can it produce a marked effect" (CN-E-02).

The Russian USE (Unified State Examination) started as an obligatory graduation exam for upper secondary schools in 2009, but with growing political concern about the unsatisfactory quality of education, a variety of other uses

were attached to it, including informing pedagogy, ensuring accountability, and monitoring education quality through the production of quantitative performance measurement data based on exam scores (see Luk'yanova 2012; Tyumeneva 2013; Piattoeva 2015). Because these developments transformed the USE into a very high-stakes examination for many stakeholders, Russian ministerial authorities have recently started to diversify national assessments to relieve pressure on USE results. However, at the time of the study, the USE remained the most important, as well as the longest existing, tool for education quality assessment, and it is entrenched in the Russian Law on Education.

The Russian state's education project is targeted at the unification and standardisation of the content and results of national school education. The discourse of "putting the state back in", meaning that the central authorities should have more say in the content of school education, is strong and widely supported by the authorities and public. National authorities also seek to establish an "effective state" system to increase the efficiency of federal regulation and local self-governance and bridge the gap between national policy and its implementation. However, as the implementation of national school policy is the responsibility of the municipalities and depends on their funding, disposition, and expertise, assessment data are expected to incentivise local policymakers, school administrations, and teachers to understand and target problems in learning achievements and introduce changes in line with national priorities. In this respect, the initiation of standardised testing runs in tandem with another central policy, the introduction of a more standardised, output-based curriculum model in Russia (Piattoeva 2015).

Government programmes for various public sectors, including education, embrace detailed, quantitatively articulated objectives and a system of numerical performance indicators to monitor the accomplishment of results annually. Following this, the results of the USE can be employed as objective performance and quality indicators, for example, in monitoring the quality of higher education or the efficiency of local authorities (Piattoeva 2015). Objectivity is defined as using standardised, expert-developed evaluation tools as sources of impartial information on the functioning of education. Alongside objectivity, the ideal of comparability is an underlying value and political goal of the initiatives linked with the USE. Comparability is expected to function on different levels: teachers compare their students' results and simultaneously examine the results of other teachers and students in their city, region, or country, and the regional comparability sought at the national level (by *Rosobrnadzor* – the National Agency for Control and Supervision in Education, for example). Teachers are held accountable for learning outcomes measured by standardised assessments, and the recent introduction of bonuses and "effective contracts" has intensified this trend. Teachers' informed professional judgement is highly mistrusted because, it is argued, it is either arbitrary or biased and incommensurable beyond its context, and "those who produce the results should not be put in a position to evaluate them" (Civic Chamber of the Russian Federation 2014).

In this section, we have explored the justifications and objectives for the NLSA in the three contexts, which illuminate, in CADEP's terminology, the discursive context (political possibilities) within which NLSA data emerge and circulate as a tool of governance at a distance. Our analysis makes clear that the countries share a discursive space, and all aspire for NLSA data to play a significant role in achieving: 1) the production of evidence for better policy and better understanding of the state of education in the country; 2) the closing of the policy implementation gap – a goal mainly linked to the vastness and internal heterogeneity of the countries and, in the cases of Brazil and Russia, to their nature as federal states with substantial autonomies delegated to the sub-national level; and 3) the management of education quality (with quality bearing a slightly different connotation in each case: see Minina et al. 2018). This analysis affords an important background to our exploration by characterising the discursive context and the national policies to which NLSA data are expected to contribute in the three country contexts. In what follows, we explore the actual NLSA data circulation.

The uncertainties of data production

In Russia, the production of the USE is centralised, being tightly controlled by *Rosobrnadzor* and the technical and analytical units functioning under it (for example, the Federal Institute of Pedagogical Measurement in charge of developing test items and analysing test scores, and the Federal Testing Centre responsible for infrastructure – the distribution and collection of test packages). However, local actors like regional information centres are responsible for the enactment of exams in strict accordance with national legislation, and non-multiple-choice tests are marked by groups of subject specialists usually composed of teachers. The national authorities seek to standardise and control data collection procedures to the greatest possible extent (through strict and detailed legislation and observation, training, and harsh penalties for rule violation), while still failing to ensure maximum compliance (see Piattoeva 2016). Thus, the quality of the NLSA is a constant focus of political and academic discussion, largely because data are inevitably produced by and rely on multiple actors with different stakes, and data are therefore expected to play too many essentially incompatible roles.

Problems are caused by the conflicting demands for the USE to become a nationwide standardised examination, a tool of accountability, and a source of impartial policy – relevant knowledge. In the general climate of increased control of the functioning and results of school education in general and assessments in particular (Piattoeva 2016; Gurova & Piattoeva forthcoming), even attempts to produce information on the exigencies of the education system are reinterpreted and diverted to the discourse of control:

There are no groups, except for external experts, who would be interested in objective results. For instance, I would be interested in looking at them.

But for that we need to refrain from using them for evaluation because now the USE carries too many burdens, that is, it is the graduation certificate, entrance examination to higher education, and in addition it assesses everything possible starting from the teacher.

(RU-E-03)

In China, the NAEQ assessments are only starting to take root, and the people involved still have little experience of their implementation. Our evidence on the topic is therefore limited, especially concerning the actual use of assessment data. Nevertheless, our limited research material communicates that there is a strong perception of assessment data's fragility. Assessments are expected to counter the practice of using school examination results as the primary indicator of quality in education (Rasmussen & Zou 2014), stressing the importance of providing timely feedback to different stakeholders based on assessment data. However, there is a substantial fear among scholars and politicians that new assessments may become another high-stakes examination because of China's century-long history of defining education quality in terms of examination excellence. The fear is that results will no longer capture the reality, that is, the measured and assessed variables. An interviewee explained,

We don't want like the schools or students [to] take it as another examination. . . . [I]f they think . . . they will rank us, then they will do some reactions, yeah. You know that . . . right now . . . we get the true answers from students. Like I don't have one hour for my outdoor activities . . . but later if you take it as high-stake, I am not quite sure they will answer no. Maybe the teachers will teach them only to answer yes.

(CN-E-01)

The Brazilian academic community frequently contests the methodological and political choices underlying the design of the Evaluation System of Basic Education (SAEB), maintaining that the SAEB does not enable the identification of schools' impact on pupils' learning achievements (for a review of the literature see Candido et al. submitted). However, our interview material reveals that the INEP's technical capacity to produce data is rarely contested. It is rather the capacity of the SAEB to account for the reality of Brazilian education and influence education planning at different levels which is challenged. Whereas policymakers, civil servants, evaluation experts, and technicians at the three levels of government generally stress the necessity of a national standardised evaluation to achieve education quality, many scholars and practitioners question the suitability of quantitative national standardised approaches to assess or make sense of educational realities. They consider the Brazilian education system too complex to be measured by quantitative indicators, which they characterise as superficial. For example, some interviewees raised questions about the significance and method of evaluation by contending its irrelevance either for daily school life or Brazilian education as a whole:

[T]he pertinence between what is required in these assessments or evaluation systems, as is the case with the SAEB, and the relationship they have with what in fact takes place in the daily life of Brazilian education classrooms, that's a question that still needs to be better explored in the country.
(BR-E-02)

While views diverge concerning the content the SAEB should measure and the SAEB itself as an evaluation tool, there is agreement on the inherent logistical challenges of data production in such a large and diverse country. The production of SAEB data is centralised. The INEP is responsible for the design of the SAEB, although scholars, experts, and teachers are engaged to work on content. The INEP hires private companies to collect data and provide certain analyses, the results of which are crosschecked with the INEP's to confirm data validity. The INEP works directly with one key person in each State Secretariat for Education, whose perceptions about the SAEB and degree of involvement in the process seem to affect the smoothness of data collection and the subsequent application of results. However, data production remains challenging in a country where the main NLSA covers 58,000 public schools and 4,000,000 pupils (INEP 2016). The time lapse between data collection, analysis, and dissemination does not allow for an in-depth treatment of data because, as an INEP official puts it, "[W]hen you [have] to publish the result, many times you are already engaged in another process on another [assessment] to start again" (BR-N-07).

Quantification relies on a series of interpretative decisions about what to quantify and how to categorise and label it (see e.g., Desrosières 1998). Previous research argues that antithetical attitudes towards data usually emanate from high degrees of negotiation of validity and reliability and the tensions and dynamics which may be created by methods of data collection (Landahl & Lundahl 2013: pp. 57–58). In our cases, these issues also easily conflict with data production and explain why actors do not always treat the NLSA as entirely objective, precise, or accurate, even when the technical side of producing quantitative assessments is little contested. In Brazil, criticisms of validity are especially tangible. In Russia and China, criticism of produced data focuses mainly on its reliability and emanates from a mistrust of data production processes. This mistrust is linked to political contexts in which data carry high stakes and there are historically specific assumptions about the meaning and role of external evaluation.

The ambiguities of data availability

In China, data generated directly by assessments are only available to the relevant NAEQ staff and to the highest levels of the Ministry of Education. Outside this circle of trust, the lower government units, the media, the public, and even external researchers – as our research team discovered as a result of a small misunderstanding in the project's early stages – are not allowed access to data. There is an evident fear that if such unfiltered data spread too far, control

of them may be lost and the media will focus on naming and shaming, again reviving the high-stakes image of assessments which the government is trying to counteract. An interviewee explained,

In China there is this problem. That is, there is this peculiarity that if you publish [assessment] data, they [whoever has access to data] will start making mutual comparisons . . . It might be that the consequence is that there will be rewards and punishments. But this will cause the objective of the evaluation to stray out of line.

(CN-E-06)

However, there are some minor exceptions to this rule. As mentioned, the assessment system has been developed by relying on foreign expertise. Thus, scholars and experts working at the NAEQ have sometimes consulted external experts to collaborate on assessment tools. In these cases, national experts can choose to provide outsiders with data. During our meeting, an interviewee reflected, “For example, I could use our data as an example for you today, but it’s absolutely not fine for others to see it” (CN-E-02).

Despite restrictions of the availability of assessment data, the NAEQ provides ready feedback in the form of information or results to sub-national units. An interviewee working for the NAEQ stated, “[A]nother division [in the NAEQ] is our reporting . . . division. They use our data to write reports for the governments” (CN-E-01). In practice, this means that the scores of a single school or the questionnaire answers of a single teacher cannot be extracted from results distributed to the sub-national governments. Instead, a county-level government, for example, will only receive information about average performances and results in their county and about the national average. They can therefore only make comparisons between their own performance and the national average or on a time axis by comparing their current and previous performances. However, there is some ambiguity concerning whether local governments will be allowed to distribute these results to whomever they wish. Moreover, in the currently available source material (publicly accessible documents, our interviews, and existing research), there is conflicting information regarding whether the school level receives any feedback. What is known is that individual students or teachers participating in assessments do not receive it. Again, the fear is that assessments will be treated as a high-stakes test. The Chinese authorities thus face a situation in which they seek a delicate balance concerning which assessment data and results are distributed and how. They fear that if they are spread too extensively, assessments will become a high-stakes test. An interviewee explained,

So, it [the feedback] is not for the individual. And we don’t want them [the students] to get nervous about their individual score. . . . [W]e don’t want [the assessment to be] taken as a high-stakes test . . . If you report

individuals, maybe . . . the students will get nervous thinking how they are going to perform.

(CN-E-01)

In Russia, individual test scores of the results and participants of the USE are stored in the federal database. The users of the database are said to be all the interested federal and local authorities, educational institutions, and other organisations involved in the USE. The establishment of the databank, therefore, constitutes an unprecedented achievement – never has the federal state possessed comparative data of such geographical scope and detail (see Bakker 1999). However, evaluation experts argue that the data accessible for research are insufficient; the average scores available for public attention, for example, are far from being the research data they want or can work with analytically. Thus, actors who can perform a scientific analysis of assessment data as a precondition for informed political decisions are largely unable to complete the task because their access to data is limited:

The ministry and *Rosobrnadzor* are keen to accumulate all the information for themselves, which they receive from the [regional] centres of quality evaluation, from the regional departments and ministries, and it is very difficult for society to receive this information. For example, Russia took part in TALIS [the OECD's Teaching and Learning International Survey], you know . . . So now in order to productively analyse TALIS data, we need the USE results of the schools that took part in TALIS. These are only two hundred from fourteen or fifteen regions. We cannot obtain this information. Can you imagine? We are the national coordinator, our institute is the operator of the project, we don't need this information for the purpose of publicising it in the newspapers . . . We need to implement a comprehensive analysis of an international study. Big money was invested. Who is interested in the analysis? Who ordered the study? The ministry.

(RU-E-04)

Uncertainties concerning data availability surface as a struggle between the necessity to collect data for governance and the management of education quality and simultaneous realisation that production of performance data can easily weaken decision-makers' immunity and lead to calls for reverse accountability. As one Russian respondent eloquently commented, "No one is interested in receiving objective information." He continued that the ministry did not wish to see poor results because if poor performance data surfaced, the next question would be "What do we need the Ministry of Education for?" (RU-E-03). Yet assessment data are widely available in the form of crude average scores (of schools, regions, in a particular subject, and so on) as performance indicators employed by officials at different levels of education administration and in league tables supported by private and public actors. While

authorities frequently state that schools or teachers should not be compared or ranked based on testing scores – at least not crudely, with no consideration of diverse contextual factors – no strong efforts are undertaken to deal with the practice. The national ministry’s general message concerning this topic remains controversial, as the ministry itself has ordered and refers to league tables.

In Brazil, the situation is the reverse of that in China and Russia. Data dissemination is encouraged by the fact that the INEP produces a large amount of varied data, which scholars use extensively. The INEP provides a research service and space for consultants, teachers, and researchers to access INEP databases, the only criteria being that data are used for research or study and ethical principles are met. The INEP has also supported several initiatives to meet with other education actors to discuss the NLSA and other monitoring tools in the context of the National Educational Plan (e.g., INEP 2015). Furthermore, it recently created an online platform for teachers to provide them with direct and almost individualised pedagogical feedback (data are presented by schools and localities, pupils’ proficiency levels, comparisons with similar schools in the region, and NLSA items and their pedagogical explanation; see INEP n.d.a).

Besides initiatives targeting the educational community and relevant political actors, the INEP makes data available to the wider public. It publicly disseminates the results of the SAEB, which consist of the mean mathematics and Portuguese literacy scores combined with several other variables. The INEP’s website design and performance has recently been improved, and data are extensively available in several formats. Despite these recent developments, all the INEP interviewees acknowledged that data dissemination was the INEP’s main challenge. Interviewees described INEP reports and databases as providing “arid” information which was difficult to understand and recognised that civil society organisations were the most important actors in helping with data availability and visibility. Even if these organisations frequently draw on data to push their own agendas (sometimes their policy proposals are aligned with the government’s, sometimes not), their contribution to data dissemination is crucial. They are seen to have better communication strategies and display results in easily accessible, user-friendly ways. Indeed, local-level interviewees frequently referred to the QEdU platform, which is administered by a privately funded civil society organisation and extensively and interactively displays NLSA data.

In Brazil, data dissemination is vulnerable to pressure from every sector of society, including policymakers, the media, civil society organisations, and even school principals, wanting to advertise their place in the rankings (e.g., BR-N-02). The media diffuses them extensively. Whenever NLSAs take place, assessment processes are scrutinised, technical or procedural mistakes identified, and results used to support arguments, usually according to the speaker’s political and education stances. Freely drawing on these results, various parallel polemical rankings are constructed. Politicians clearly try to channel dissemination according

to the political agenda, and professional associations' and unions' engagement with data greatly contributes to their circulation.

Despite the shared discourse on rational decision-making and making data a prime tool of education governance at various levels, our cases demonstrate considerable differences in data availability, that is, who has access to assessment data, of what kind, and for what purpose. Here, as in the previous section, we see a pattern in which the realisation of shared policy is embedded in the case-specific institutional and discursive arrangements and particularities of socio-historical dependencies, which can both enable and constrain actual data access.

Too much or too little data, and who knows how to use them?

All Brazilian interviewees acting nationally, and therefore involved – though differently – in national education politics, expressed the view that NLSA data have fuelled the debate on education quality. As a representative of a civil society organisation stated,

[B]y putting the discussion of the metrics based on a large-scale assessment you visualise within the education system the inequality [provided by the lack of quality] and therefore you make more tangible the talks and the formulations and actions about equality.

(BR-NNGO-05)

Despite this, the effects of NLSA data on national education policy are disputed. Several interviewees suggested, as is aptly systematised by a senior civil servant working at the Federal Senate, “the consolidation of the evaluation system [i.e., of the SAEB] has been an important driving force of public policies in education” (BR-N-11). Other interviewees suggested that the impact is still far from what it could and should be. NLSAs collect information which is not comprehensively used to formulate concrete education policies which might target the improvement of the school as a whole. For many, there is enough information to understand which aspects need to be improved in each school, but there is no consequential national political use of collected data.

Several INEP interviewees mentioned that at the state level many governments use data carefully in education planning. However, other national, state, and municipal interviewees mentioned that most state and municipal administrators had little capacity to use data: they were either unqualified to do this or they lacked resources. Even in the richer states data use seems to be in its early stages. An interviewee from the state level, for example, related that they were still trying to “create a culture of data usage”, and that in most states NLSA data is still “less used than [...] it should be because there are no established competencies among the secretaries [of education] to know what to do with these results” (BR-S-01).

Nevertheless, some richer states and municipalities have chosen to create their own assessments in the guise of preparation for NLSAs and local monitoring and only rhetorically react to national results. It is argued that NLSAs do not fully meet the needs or particularities of the states. However, many states and municipalities acknowledge severe financial implications and given the existence of the national system, mention only the need to complement it. This partial disengagement with NLSAs seems connected to an attempt to ensure that education responsibilities are not appropriated by the national government, which has used central control in steering certain education aspects, such as the establishment of a common curriculum. Although these initiatives are generally welcome, it is obvious that states and municipalities see their own assessments as a strategy to retain political authority over education issues.

While NLSA data met their target in informing education debates and, partly, at least, policymaking, they seemed to fail in influencing school practice as the political project foresaw. Historically, governmental actions, regardless of level, are the object of the education community's public mistrust. More specifically, mistrust of NLSAs emanates from accountability policies. Although neither the INEP nor the Ministry of Education supports data use for accountability purposes (interviewees were unanimous here), some states and municipalities use NLSA data to make schools accountable. Furthermore, teachers do not understand results and do not know how to engage with them. This has triggered the construction of the online platforms mentioned in the previous section, which aim to inform teachers by explaining NLSAs and translating data into classroom reality. However, because of the great number of teachers in Brazil and their diverse educational backgrounds, the link between data and classroom activities remains a major challenge. NLSA results are still far from being used to improve teaching and learning in Brazilian schools.

In China, utilisation of gathered data is considered a major problem, as practitioners do not always know how to interpret NLSA data and how to employ them in practical solutions. In this opaque context, policymakers sometimes choose the feedback that best supports their political agenda, and policy surfing means feedback is used to justify a predetermined policy choice rather than to identify problems for which policy solutions might then be sought. Once the NAEQ has analysed scores and the final assessment results are released, the NAEQ staff holds meetings with education officers, inspectors, and school principals from different provinces and municipalities. NAEQ experts suggest how to tackle problems identified by assessments. These meetings' emphasis is on cooperative action. An interviewee explained, "The key is in finding the problems and then helping schools by solving the problems together" (CN-E-07). Evaluation experts thus gain in power and become indispensable in helping to translate data into potential policy solutions and pedagogical tools. However, this also has its limits: feedback does not oblige the recipient to make any changes.

As previous research (Zhang & Wan 2017) and our interviewees point out, the assessment system is a somewhat recent innovation. Data use is considered

a major problem, the feedback system remains incomplete, and follow-ups are rare. Moreover, as our interviews frequently indicated, these new assessment practices and ideas are still largely emphasised only in official documents, whereas in practice, examinations still reign supreme as the primary quality indicator. Local practitioners do not always know how to produce solutions based on results of a new assessment, even when provided with suggestions. A scholar reflected on the general situation:

I think the challenge is . . . how we can use the data well . . . the results. So, I mean, we can do the national assessment. We can easily do all these things . . . But most important [is] how we can use it . . . We want to find a way to use results.

(CN-E-05)

Sometimes, there is almost an atmosphere of apathy and disinterest concerning assessments, because they are not always seen as helpful in solving local problems:

For example, tens of counties in Shandong and Henan provinces and in Jiangsu province each year receive national-level basic education quality monitoring [i.e., they take part in assessments], but they are ignorant of the results, and just take care of organising them.

(CN-E-13)

Interestingly, the fact that assessments are not used for ranking is partly to blame. An interviewee sarcastically noted,

Some people take it [the feedback] very seriously, but most people don't. Because in our education system they care about things that have to do with selection, and they are not too concerned about these things that have to do with quality monitoring. When would they care? If you used these results to compare, they would take notice!"

(CN-E-06)

A further challenge is that assessments reveal a significant amount of detailed information (teachers' work burden in one school was an example given by an interviewee) beyond levels to which feedback extends; feedback does not target the levels at which data were collected in the first place. It appears a compromise of sorts has been struck, whereby school principals are informed of a problem at a general level so they can attempt to address it themselves. An interviewee elaborated,

So they [the school principals] need to find the causes. But we usually we cannot go that deep for them. But we . . . will . . . say: 'Okay, this is the phenomenon we've discovered. Okay, this information we've got.' . . . We kind

of give them a kind of a direction: ‘Okay, this is the way you should go to try to find out the problem’.

(CN-E-01)

In Russia, interviewees often mentioned a conflict between the demand for evidence-based decisions on the one hand and the fast pace of performance evaluation on the other. National-level civil servants criticised local administrations for their lack of intellectual investment in the decisions that would bear fruit, and their preference for quick fixes (such as dismissal of personnel).

There are two negative aspects in managerial competences. First, non-intellectual reaction, when a head sees a ranking position and says “punish the bad ones and reward the good ones” . . . And second, making decisions without analysing the causes of problems, or without the adequate resources.

(RU-O-04)

At the same time, the support provided at the national level to improve managerial resources was often limited to simply channelling more data, especially comparisons between different regions’ achievements. Analytical support was thus supplanted by an increase in the volumes of data distributed during internal meetings between the national ministry and regional authorities, for example.

National experts questioned officials’ capacity to work with data and whether data were used for rational decision-making instead of heavy control or one-sided reporting.

The civil servants have no time to pause and think. They work at a very fast pace. And if [s/he] didn’t have the required competences to start with, it is pointless to hope they will develop.

(RU-E-03)

Russian interviewees often complained about the expansion of reporting at different governance levels, meaning that, in addition to NLSAs, schools and regional authorities spent progressively more time on collecting and channelling different types of statistical and other information requested and processed by higher levels. Interviewees expressed a somewhat contradictory position about the simultaneous abundance and shortage of data. Many mentioned that data were of the wrong kind. As we discussed earlier, high-stakes data cannot be trusted fully even when they are available because data lack the contextual information needed for a deeper academic analysis. The difficulty and high cost of data collection were also mentioned. In this respect, experts compared Russian NLSA data to international large-scale assessments and praised the latter for their richness, openness, and depth.

Concrete understanding of management and data use is still in the making. Some policymakers admit that data cannot be used on their own and that

common sense is still needed in decision-making processes, acknowledging the limitations of data such as the problems caused by high stakes and the complexity of learning outcomes on which the NLSA is intended to reflect (e.g., RU-O-05). However, some individual and still rare examples of concrete data use in the decision-making process were reported, for example, targeted training in Russian language teaching for regions where Russian is the second language, as well as the decision to divide the USE mathematics test into basic and advanced levels, reflecting the status of the USE as a graduation exam and its recognition of the different needs of students in further education (e.g., RU-O-05). Surprisingly, despite the frequently cited lack of resources, time, or possibility to analyse available data and use it meaningfully for evidence-based policy, the motivation to produce more data has remained strong:

When we receive a set of information on the state of affairs, then in a good sense we always have requests for further development. For example, when we get a snapshot, we want to see dynamics. When we receive dynamics, we want to see the added growth, individual progress. When we know the individual progress in maths, for example, we want to have an idea about the individual progress in thinking . . . And second, [data] should be used for decision-making. I want to emphasise again that between data and decisions there is a space in the form of the manager's head, which is regularly missing.

(RU-O-04)

The idea behind evidence-based policy is that the more information policymakers collect, and the faster they collect it, the more informed decisions they will be able to make and, thus, the more effective the management of social problems will be (Tsoukas 1997). However, this section problematises this image of evidence-based policy and points to the complexities of putting quantified education data to work in education realities, whether at the policy-making level or in regions, schools, and classrooms.

Conclusion

Our comparative analysis of the three national cases has explored the diverse ways by which data are intended, but often fail or only partly succeed, to be integrated in education governance practices at a distance. The political project of QAE in general and the NLSA in particular in Brazil, China, and Russia envisions NLSA data providing a better general view of the education systems of these enormous and heterogeneous countries and a more nuanced view of learning achievement at the student, teacher, and school levels, thus optimistically and at least rhetorically incorporating very diverse stakeholders in the data project. In reality, data are not easily integrated in these actors' practices, limiting their capacity to govern at a distance. Echoing our other chapters, the variation and friction the chapter describes reflect the specific political situations

and possibilities embedded in the socio-political and historical contexts of these countries, such as the various country-specific reforms in education and beyond. In Brazil, for example, the strongly democratic open and multi-vocal policymaking agenda provides a platform to make performance data fully available – a discursive feature absent in China and Russia. Restrictions on data availability in China and Russia reflect different contextual factors. They emanate from synoptic data possibilities, because in one case governance at a distance afforded by data comes at the potential cost of making people in political power visible (see Landahl & Lundahl 2013). In another case, there is a fear that publicly available data may cease to reflect educational reality ‘objectively’. Thus, in contrast to Brazil, NLSA data in China and Russia are guarded and used exclusively by national officials. Our interviews with civil servants and evaluation experts often signalled ambiguity and scepticism regarding data analysis and use even within official circles. Moreover, restricting data to official usage has negative consequences for the overall project of evidence-based policymaking, as academic experts may also be excluded from the data sources necessary for a scientifically sound analysis.

In all cases, we have detected concern about data objectivity, especially the reliability of data collection, which is inevitably performed by actors who are more or less external to the evaluation agencies. Data production is problematic either because it is conducted by locals with varying degrees of professional interest in and commitment to the process or because it is generated by actors who are then evaluated based on these data. Fragility of trust in the processes of data collection is compounded by mistrust of the officials and experts involved in administering and interpreting assessments and the tensions between experts and officials at different levels of governance and between representatives of national and sub-national education administration. These complications weaken data, paving the way for further mistrust and criticism. Quantification thus highlights the important role of trust between civil servants and assessment experts because of the civil servants’ reliance on the technical and analytical capacity of the experts (see the case of Russia especially). Our analysis equally points to the importance of a working relationship between the national and sub-national administration levels, especially because of the national level’s reliance on the willingness and possibilities of local actors to engage with the NLSA. This relationship seems deeply distrustful, rooted in controversies concerning NLSA use as data for both evidence-based policy and accountability, as well as attempts to partly recentralise certain features of the education system by means of NLSA data (especially in Brazil).

While smooth governance at a distance through data is envisaged because of the assumed easy circulation of numerical data, circulation is contingent on the successful abstraction and de-contextualisation which allow new meanings to be attached to data when they are produced for a delimited purpose (Porter 1994). Complete or partial failure to achieve an accepted abstraction and de-contextualisation impairs or alters the proliferation of networked actors

and constellations of new relationships, undermining successful governance by data. The simplification of complexity presents an argument against NLSA data which is difficult to refute, and various actor segments may make accusations of reductionism at any time. As the case of Brazil shows, this kind of criticism is made more easily the more data are made publicly available to diverse groups of actors. Thus, paradoxically, while data should be made accessible if they are to be integrated in diverse practices to influence others' conduct, wider availability increases the possibility of questioning data and the diversion of original governance intention. In summary, whereas NLSA data are frequently argued to offer solid evidence, our study shows that NLSA data can be easily dismissed even for mutually controversial reasons. It is either described as too rigid to capture complex realities or as too subjective and unreliable to be trusted as an accurate and detached description.

These frictions occur during different data circulation stages and give rise to a situation which is characterised not by more data use and reflexivity but by the intensification of data production, thus ultimately accentuating a particular phase of data circulation. Contrary to David Beer (2016), quoted in this chapter's introduction, Anna Tsing (2005) describes friction as essential for motion: friction is productive and enabling. Inspired by this alternative perspective, we conclude that while data circulation and thus governance by data do not function as a well-oiled machine, friction does not slow processes down but instead bears a productive power to retune them in new ways. However, as friction emanates from historical trajectories and the particularities of political and discursive contexts, unexpected directions are neither random nor open to any possibility. In our cases, we especially emphasise the politically and historically charged dynamics of trust and mistrust in the possibilities of NLSA data, linked to the complicated relationships of trust and mistrust between and among different actors. This situation occurs both nationally and locally and between levels and leads to the intensification of data production. For example, in Brazil, local actors produce new data to both prepare for the NLSA and generate data more suited to local needs. In Russia and China, data production is stimulated by problems of data reliability and objectivity.

Bibliography

- Bakker, S. A. (1999). Educational assessment in the Russian federation. *Assessment in Education: Principles, Policy & Practice*, 6(2), 291–303.
- Ball, S. (2016). Following policy: Networks, network ethnography and education policy mobilities. *Journal of Education Policy*, 31(5), 549–566.
- Beer, D. (2016). *Metric Power*. London: Palgrave Macmillan.
- Candido, D. H. H., Kauko, J., & Centeno, G. V. (submitted). Portraying quality assurance and evaluation in Brazil: An analysis of how quality is problematized in Brazilian basic education.
- Centeno, G. V., Kauko, J., & Candido, D. H. H. (2017). Quality assurance and evaluation through Brazilian lenses: An exploration into the validity of umbrella concepts. *Comparative Education* 54(2), 132–158.

- Civic Chamber of the Russian Federation. (2014). *USE Shifts to Independent Rails*. Retrieved May 6, 2015, from www.oprf.ru/ru/print_datas/26091 [in Russian].
- Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.
- Espeland, W. N., & Sauder, M. (2012). The dynamism of indicators. In K. E. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by Indicators* (pp. 86–109). Oxford: Oxford University Press.
- Fan, Y., & Liao, Q. (2013). 基础教育质量监测:国际经验与本土路向 [Basic education quality monitoring: International experiences and the local path], *教育导刊*, 4, 37–40.
- Gorur, R. (2015). Assembling a sociology of numbers. In M. Hamilton, B. Maddox, & C. Addey (Eds.), *Literacy as Numbers – Researching the Politics and Practices of International Literacy Assessment* (pp. 1–16). London: Cambridge University Press.
- Grek, S., Lawn, M., Lingard, B., Ozga, J., Rinne, R., Segerholm, C., & Simola, H. (2011). National policy brokering and the construction of the European education space in England, Sweden, Finland and Scotland. In J. Ozga, P. Dahler-Larsen, C. Segerholm, & H. Simola (Eds.), *Fabricating Quality in Education: Data and Governance in Europe* (pp. 47–65). London & New York, NY: Routledge.
- Gurova, G., & Piattoeva, N. (2018). A post-Soviet audit culture: Changing practices and subjectivities of schoolteachers in a Russian region. In L. M. Carvalho, L. Levasseur, L. Min, R. Normand, D. Oliveira, & D. Andrade (Eds.), *Education Policies and the Restructuring of Educational Professions*. Dordrecht: Springer.
- Hansen, H. K., & Flyverbom, M. (2014). The politics of transparency and the calibration of knowledge in the digital age. *Organization*, 22(6), 872–889.
- Hansen, H. K., & Mühlen-Schulte, A. (2012). The power of numbers in global governance. *Journal of International Relations and Development*, 15(4), 455–465.
- Hartong, S. (2016). Between assessments, digital technologies, and big data: The growing influence of “hidden” data mediators in education. *European Educational Research Journal*, 15(5), 523–536.
- INEP. (2015, August 12). *Notícias: Inep promove encontro de trabalho para debater instrumentos de monitoramento do PNE*. Retrieved from http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-promove-encontro-de-trabalho-para-debater-instrumentos-de-monitoramento-do-pne/21206
- INEP. (2016, September). *Sistema De Avaliação Da Educação Básica, Edição 2015, Resultados*. Retrieved from download.inep.gov.br/.../saeb_2015_resumo_dos_resultados.pptx
- INEP. (n.d.a). *Devolutivas Pedagógicas*. Retrieved July 16, 2017, from http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-promove-encontro-de-trabalho-para-debater-instrumentos-de-monitoramento-do-pne/21206
- INEP. (n.d.b.). *Institucional*. Retrieved March 24, 2017, from <http://portal.inep.gov.br/web/guest/conheca-o-inep>
- INEP. (n.d.c). *Saeb*. Retrieved March 24, 2017, from portal.inep.gov.br/web/guest/educacao-basica/saeb
- Kauko, J., Centeno, G. V., Candido, D. H. H., Shiroma, E., & Klutas, A. (2016). The emergence of quality assessment in Brazilian basic education. *European Educational Research Journal*, 15(5), 558–579.
- Landahl, C., & Lundahl, J. (2013). (Mis)trust in numbers: Shape shifting and directions in the modern history of data in Swedish educational reform. In M. Lawn (Ed.), *The Rise of Data in Education Systems: Collection, Visualisation and Use* (pp. 57–78). Oxford: Symposium Books.
- Lascoumes, P., & Le Galès, P. (2007). Introduction: Understanding public policy through its instruments – from the nature of instruments to the sociology of public policy instrumentation. *Governance*, 20(1), 1–21.

- Lawn, M., & Segerholm, C. (2011). Europe through experts and technologies. In J. Ozga, P. Dahler-Larsen, C. Segerholm, & H. Simola (Eds.), *Fabricating Quality in Education: Data and Governance in Europe* (pp. 32–46). London & New York, NY: Routledge.
- Luk'yanova, E. (2012). Russian educational reform and the introduction of the unified state exam. A view from the provinces. *Europe-Asia Studies*, 64(10), 1893–1910.
- Maroy, C. (2009). Convergences and hybridization of educational policies around “post - bureaucratic” models of regulation. *Compare*, 39(1), 71–84.
- Miller, P., & Rose, N. (1990). Governing economic life. *Economy and Society*, 19(1), 1–31.
- Minina, E., Piattoeva, N., Centeno, V. G., Zhou, X., & Candido, D. H. H. (2018). Transnational policy borrowing and national interpretations of educational quality in Russia, China, and Brazil. In I. Silova & M. Chankseliani (Eds.), *Comparing Post-Socialist Transformations: Education in Eastern Europe and Former Soviet Union*. Oxford: Symposium Books.
- Piattoeva, N. (2015). Elastic numbers: National examinations data as a technology of government. *Journal of Education Policy*, 30(3), 316–334.
- Piattoeva, N. (2016). The imperative to protect data and the rise of surveillance cameras in administering national testing in Russia. *European Educational Research Journal*, 15(1), 82–98.
- Porter, T. (1994). Objectivity as standardization: The rhetoric of impersonality in measurement, statistics, and cost-benefit analysis. In A. Megill (Ed.), *Rethinking Objectivity* (pp. 197–238). Durham: Duke University Press.
- Rasmussen, P., & Zou, Y. (2014). The development of educational accountability in China and Denmark. *Education Policy Analysis Archives*, 22(121).
- Rinne, R., & Ozga, J. (2011). Europe and the global: The role of the OECD in education politics. In J. Ozga, P. Dahler-Larsen, C. Segerholm, & H. Simola (Eds.), *Fabricating Quality in Education: Data and Governance in Europe* (pp. 66–75). London: Routledge.
- Rose, N., & Miller, P. (1992). Political power beyond the State: Problematics of government. *British Journal of Sociology*, 43(2), 173–205.
- Rottenburg, R., & Engle Merry, S. (2015). A world of indicators: The making of governmental knowledge through quantification. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The World of Indicators: The Making of Governmental Knowledge Through Quantification* (pp. 1–33). Cambridge: Cambridge University Press.
- Selwyn, N. (2016). “There’s so much data”: Exploring the realities of data-based school governance. *European Educational Research Journal*, 15(1), 54–68.
- Souto-Otero, M., & Beneito-Montagut, R. (2016). From governing through data to governmentality through data: Artefacts, strategies and the digital turn. *European Educational Research Journal*, 15(1), 14–33.
- State Council of the People’s Republic of China. (2010). 国家中长期教育改革和发展规划纲要 2010–2020年 [Outline of China’s national plan for medium and long-term education reform and development 2010–2020]. Retrieved from www.gov.cn/jrzq/2010-07/29/content_1667143.htm
- Steiner-Khamsi, G. (2016). Global indicators and local problem recognition: An exploration into the statistical eradication of teacher shortage in the post-socialist region. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The Handbook of Global Education Policy* (pp. 573–589). West Sussex: John Wiley & Sons.
- Tsing, A. (2005). *Friction: An Ethnography of Global Connection*. Princeton, NJ: Princeton University Press.
- Tsoukas, H. (1997). The tyranny of light: The temptations and the paradoxes of the information society. *Futures*, 29(9), 827–843.
- Tyumeneva, Y. (2013). *Disseminating and Using Student Assessment Information in Russia*. Washington, DC: World Bank.

- Xinhua. (2010, 29 November). 我国加强基础教育质量监测 监测学生全面素质 [*China Strengthens Basic Education Quality Monitoring – Monitoring the Overall Quality of Students*]. Retrieved from www.gov.cn/jrzq/2010-11/29/content_1755670.htm
- Zhang, H., & Wan, D. (2017). Status of Chinese science education reforms: Policies and development framework. In L. L. Liang, X. Liu, & G. W. Fulmer (Eds.), *Chinese Science Education in the 21st Century: Policy, Practice, and Research* (pp. 5–30). Dordrecht: Springer.
- Zhou, X., Suominen, O., Fan, Y., Kallo, J., & Rinne, R. (forthcoming). *Trajectory of Reforms to Education Supervision in Contemporary China*.

