



Numbers and their contexts: how quantified actors narrate numbers and decontextualization

Nelli Piattoeva¹

Received: 2 May 2020 / Accepted: 2 June 2021 / Published online: 24 August 2021
© The Author(s) 2021

Abstract

The proliferation of standardized testing and administrative statistics in compulsory education is embedded in the rise to prominence of quantified accountability as a mechanism of education governance. Numbers work by stripping away the contexts of their production and the granular and ambiguous detail of the phenomena they claim to represent. The article re-examines qualitative interviews collected during a completed international project that studied policies and practices of accountability reforms and quality evaluation in Russian school education to understand how actors involved in quantified accountability—from producers to users of large-scale assessments of learning outcomes and administrative statistical data—articulate measuring or being measured in a decontextualized manner. The article offers two theoretical contributions on the enactment of accountability policies. First, it shows the importance of analytical attention to the materiality of accountability policies, that is, their specifically numerical nature. Second, it proposes that the study of accountability enactment should address the question of how actors deal with the decontextualizing propensities of quantified accountability. I conclude that enactment of accountability is fuelled by quantified decontextualization and the diverse ways in which actors experience, make sense of and act upon it.

Keywords Quantification · Russia · Accountability · School education · Education policy
Materiality of policy

1 Introduction

Governance has become increasingly and more overtly techno-rational (Le Galès, 2016), and governance by numbers, which reduces “complex processes to simple numerical indicators and rankings for purposes of management and control” (Shore & Wright, 2015, p. 22), is a prominent characteristic of our times. Accountability

✉ Nelli Piattoeva
nelli.piattoeva@tuni.fi

¹ Tampereen Yliopisto, Tampere, Finland

has broken loose from its historical association with the realm of democratic answerability that requires decision-makers to justify their actions to those subject to their power, and has become more narrowly associated with finance and the efficient use of material resources on the one hand and monitoring of measured performance on the other (Strathern, 2000, p. 2; Espeland & Vannebo, 2007; Piattoeva & Boden, 2020; Grek et al., 2011). Moreover, “some rather specific procedures have come to carry the cultural stamp of accountability, notably assessments which are likened to audit.” (Strathern, 2000, p. 2.)

This paper is anchored in the paradox of numbers used for accountability, namely, that their legitimacy as objective representations of reality or impartial tools of governance and decision-making relies on de-contextualization, opacity and being removed from the complex living texture of the world. I examine how actors involved in quantified accountability—from producers to users of large-scale assessments of learning outcomes and administrative statistical data in Russia—articulate measuring or being measured in a decontextualized manner. Empirically, I am interested in how these actors discuss “reality”, that is, their personal or professional selves and contexts represented through numbers, and deal particularly with the decontextualizing qualities of quantification. Even though this was not the question that we directly asked the interviewees during qualitative data collection for an international project described in more detail in the methodological section of the article, re-reading the data brought this question into sharp focus, calling for more detailed analysis.

The article makes two theoretical contributions to the ongoing discussions on the enactment of accountability policies. First, it shows the importance of analytical attention to the materiality of accountability policies, that is, their specific *numerical* nature. The seminal work by Braun et al. (2011) interpreted enactment as recontextualization of policy and alluded to the role of material context in the creative processes of policy translation and interpretation. Following in their footsteps, I approach the material context of accountability policies as the “materiality of policy” (Ball, Maguire and Braun, 2012, p. 41). This perspective, anchored in actor-network theorizing (cf. Fenwick & Edwards, 2011; Piattoeva, 2018), examines the heterogeneity and distributed nature of policy and the role of non-human actors in policy enactments. It foregrounds the performative qualities of the materiality that makes up policy and argues that they afford the “recognition of certain entities and attributes while excluding others from the reckoning” (Gorur, 2015, p. 92). The latter is a particularly important insertion, as accountability policies largely, though not exclusively, mobilize numbers as a means of holding to account, and their transformative effects stem from the specifics of quantification (Espeland & Vannebo, 2007).

Second, while theories of policy enactment have emphasized the diversity and multi-layered nature of educational contexts and their key role in policy enactment, sociology and anthropology of quantification have shown how numbers act as technologies of de-contextualization, producing abstractions rather than detailed descriptions of accountable entities (Rottenburg & Engle Merry, 2015). Accordingly, the second contribution of this article proposes that the study of accountability enactment should address the more specific question of how actors deal with the

decontextualizing propensities of quantified accountability. I propose that enactment of accountability is fuelled by quantified decontextualization and the diverse ways in which actors experience, make-sense of and act upon it.

The article is structured as follows. The first section offers a brief outline of the key dimensions of quantification in accountability policies, complemented by a discussion of the epistemological approach to numbers as actors. In the next section, I move on to describe my methodological approach of “number narratives” and present the research data. The following part sets the context by introducing education accountability reforms in Russia since the early 1990s. The empirical section examines the research findings according to three narratives that depicted the data/context problematique: (1) data objectivity, (2) data lacking in context and (3) data as misrepresentation, followed by a concluding section.

2 Examining numbers and quantified accountability: theoretical background

Quantification has become a prominent mode of knowledge production on the one hand and policy instrumentation (Le Galès, 2016) and technologies of government (Rose & Miller, 1992, Piattoeva, 2015) on the other. Porter has argued that the “bureaucratic use of numbers presupposes their objectivity but not their truth” (1994, p. 211), meaning that the general perception of numbers as objective does not stem from their being “true” to reality, but from claiming an impartiality of procedure through which this reality is rendered legible. Moreover, quantification came to be allied with both science and governing because numbers are transportable and combinable (Desrosières, 1998; Miller & Rose, 1990). Porter (1994) calls numbers a means of communication across (vast) distances because of their capacity to enable contact and co-ordination despite differences in disciplinary background and national or local tradition. Thus, numbers exercise power by becoming part of the world as they circulate, and their power is a product of circulation (Beer, 2016; Grek et al., 2011).

Numbers produce a synoptic view, a kind of abstracted tunnel vision (Scott, 1998), that brings into focus a limited aspect of a complex reality, focusing on that element of the reality that the state or the governing agent considers important or amenable to being rendered legible. A general synoptic view is preferred over granular detail. This is no oversight or inevitable cost of quantification, but rather its very premise and purpose. Most quantification can be understood as commensuration because quantification implies the evaluation of characteristics normally expressed by reference to their distinctive and often incompatible qualities through a common, standardized metric (Espeland & Stevens, 1998, p. 314). It creates relations between different entities through a common metric, transforming differences into magnitudes that can be understood and valued against each other. Commensuration impacts contextualization by both generating and inhibiting relations between attributes or dimensions. As “a set of connections between the object of inquiry and its surroundings” (Morita, 2014, p. 215) context moves from being intrinsic to the phenomenon under study to becoming extrinsic to it, that is, relevant contextual

factors are derived primarily through commensuration and comparison. Thus, commensuration changes the meaning of context and predefines and standardizes relations between explanandum and explanans (cf. Espeland & Stevens, 1998). In this sense, commensuration (re)constructs the world through both the socio-technical processes that it requires to become effective and through the representations of the world that it brings into being.

Audit, as one of the most prominent ways to perform accountability of public education, manifests how contemporary education governance has come to rely upon tools that function at a distance through numerical forms of knowledge production, assessment techniques and evaluation data. This development has, for instance, accelerated the proliferation of standardized testing of learning outcomes (Benavot & Tanner, 2007). Whether we describe this expansion of quantification as neoliberal governance (Fouger, 2008), New Public Management (Hood, 1991), audit culture (Shore & Wright, 2015; Strathern, 2000) or metric society (Mau, 2017), what all these have in common is that numbers are increasingly treated in a realist modality and are expected to enable and reconcile centralized accountability with decentralized action (Rottenburg & Engle Merry, 2015). The synoptic, decontextualized data is thus taxed with making judgment on a broad spectrum of intrinsically and contextually rich entities (Rottenburg & Engle Merry, 2015). Moreover, in this context, quantification introduces certain normative “orders or worth” that establish evaluation standards and a vision of how things are to be viewed and assessed, and which activities, achievements or characteristics have a “value” (Boltanski & Chiapello, 2005; Boltanski & Thévenot, 2006, cited in Mau, 2020, p. 23).

There is a certain determinism by which numbers are often described as powerful actors. I do not start from the ontological assumption that numbers are all-powerful means of accountability, policymaking or knowledge-production (cf. Piattoeva & Boden, 2020) but see numbers and numbering as inherently complex socio-cultural and socio-technical achievements. Although often portrayed as robust, objective and neutral, numbers are nevertheless inherently interpretive, fluid and amorphous (Piattoeva & Boden, 2020). This sensibility originates from science and technology studies and actor-network theory, which approach actors such as numbers as “a fragile assemblage performing itself as solid and immutable” (Fenwick & Edwards, 2011, p. 719). Numbers as socio-material, heterogeneous arrangements are not neutral, but rather performative devices emerging as a result of political work and temporal consensus. Because numbers shape and in turn are shaped by practices, and practices are uncertain, they remain open, contested and contestable. Numbers are also potentially internally inconsistent and incoherent, and these characteristics enable numbers to perform different tasks, but also to travel routes unforeseen by their strategists (Law & Ruppert, 2013).

Scholars of quantification have called for systematic investigations of quantification and governance by numbers in situ (Mennicken & Espeland, 2019) even though quantification itself is perceived as detrimental to a contextualized account. They take issue with quantification exerting uniform effects across distinct social domains (Gorur, 2018; Kipnis, 2008; Mugler, 2015) and study how numbers wield power by becoming part of the world as they circulate (Beer, 2016). Methodologically, understanding circulation is about examining the continuity or discontinuity between

descriptive measures and prescriptive outcomes (Beer, 2016). Circulation is a recursive game of subtraction and addition of contextual narratives of what measures are, what they are good for and what they should achieve (Beer, 2016; Espeland, 2015). The process of circulation creates possibilities that may strengthen or weaken numbers used for generating understanding, and enacting governance or accountability (Piattoeva & Boden, 2020), potentially because “politics and accounting practices are not questions of one will or of one desire, but rather of encounters and confrontations between competing projects and desires” (Asdal, 2011, p. 7).

The actors involved in quantification may express numeracy and reflexive competencies that help them to reflect on and even strategically manipulate the simplifications of numerical data and the policies of accountability based thereon (Mugler, 2015). Such narrative (re)counts of numbers were earlier examined by Espeland (2015, p. 61), who documented how they frequently appeared “causal and defensive, and often focus on unpacking some formalized aspect of identity that is a component of rankings. Some person or organization is telling a story and the story is a causal story that accounts for how events or entities are linked”. People affected by quantification may not take numbers at face value, but engage with their virtues and limitations, promises and failings (cf. Espeland, 2015). They may “reclaim information that has been stripped away, including reasons for numbers, their context and identity claims about their organizations” (Espeland, 2015, p. 65). However, the point is not that such stories necessarily disclose resistance to or escape from numbers (Gorur, 2018). Rather, they talk back to or with numbers depicting actors’ sense-making or emotional charge. Even descriptions that seem to explicitly oppose or question measures may still show how “organizational agents must define themselves in relation to measurement systems that dominate organizational and managerial discourses”, thus reiterating hegemonic discourses (Power, 2004, p. 778).

The diverse ways of understanding the experiences and sense-making of numbers among those who engage with and are affected by them have been studied by sociologists and anthropologists of quantification as number narratives (Lynch, 2019; Espeland, 2015), arguing that the effects of quantification are made possible through the interpretative meaning-making of the narrative (Lynch, 2019, 38). This sensibility enables researchers to focus on more obscure ways in which actors interact with numbers, often recognizing and engaging with their possibilities and limitations. Commensuration, as Lynch has argued, is more than just “making up people” (Hacking 1990, 6, cited in Lynch, 2019, 37). The processes may also be described as controversial and “dual-directional, so people, in turn, ‘make up’ and use numbers to label and order and categorize their a priori assessments” (Lynch, 2019, 37). Hence, my original point is to draw attention to actors’ productive engagement with the decontextualizing properties of quantified accountability and argue that it makes a fruitful contribution to the enactment approach to understanding accountability. In the following section, I explore the methodological nuances of studying number narratives.

3 Research materials and methodological approach

I adopt a qualitative approach to ascertain how people make sense of and deal with the simplifications and de/re/contextualizations produced by large-scale assessments and administrative data deployed for accountability (cf. Espeland, 2015). I present the findings of this analysis through three main themes: (1) data objectivity, (2) data lacking in context and (3) data as misrepresentation. Narrative is about “a distinct form of discourse...meaning making through the shaping or ordering of experience, a way of understanding one’s own or others’ actions, of organizing events and objects into a meaningful whole, of connecting and seeing the consequences of actions and events over time” (Chase, 2011, p. 421, quoted in Chase, 2018, p. 951). The general interest of narrative inquiry also encompasses the question of identity, that is, who I, we or they are. Personal or institutional narratives construct versions of self, others and the social world embedded in interpersonal, cultural, institutional and historical contexts (Chase, 2005; Chase, 2018). Thus, instead of conveying expressions of “true” identity, they construct and perform an identity by introducing an order and meaning to things and events. Relatedly, a narrative approach is interested in the question of power and, particularly, how narratives may function as modes of resistance—as counternarratives—to societal structures and power dynamics. Even though narrative inquiry starts from a single biographical subject, the narratives transcend this singularity (Squire, Andrews & Tamboukou, 2013; Phoenix, 2013).¹

The interrelated focal points of identity, order-making, counternarrative and contextuality are pertinent for the study of quantified accountability as a policy that judges and potentially redefines who one is (Ball, 2003; Mau, 2020) through both the reductionist nature of commensuration and the classification of one’s worth thereby, as explained earlier. At the same time, they help to avoid a strict binary between quantification and qualitative narration, making space to study how numbers and narratives co-exist and co-evolve (cf. Lynch, 2019). I read the data through the lens of these focal points with an interest in understanding how people articulate the relationship between numbers, contexts and selves, and how they might attempt to reconstruct the narrative that numbers recount about them or about the quantified entity that matters to them (their school, region, or country).² The three pervasive themes presented in the empirical section are the result of this analysis.

¹ There is a growing body of narrative inquiry in education policy research, e.g. Hansen, Wallenius, Juvonen and Varjo (2020), Bansel (2015), Crow and Jones (2018). In this article, I build on the general accounts of narrative research in social sciences and specifically the study of narratives in relation to quantification practices.

² It must also be noted that Espeland’s (2015) study focused on narratives derived from observations of events and media materials that publicly evaluated education institutions (US-based law schools in her case) produced to explain/explain away numbers. In contrast, the research materials analyzed in this article were produced by means of an immediate intervention by the researcher (through interviewing) based on an interview questionnaire adjusted for each individual interviewee. The interview guide was not developed with a narrative approach in mind, but still contained broadly and openly phrased questions about data and its role in the governance of education across levels, among other issues, allowing the respondent to engage with the question and to use as much time as needed to provide a detailed response.

I revisit published work and original qualitative research data collected for a large international study (see Kauko, Takala & Rinne, 2018). The study collected national policy documents, media sources and interview data with policymakers, education experts and school personnel to understand how quality of education has become one of the central tools of governing school education, and how the rise of quality and its measurement by means of quantitative data reforms education contexts in Brazil, China and Russia. I was personally involved in planning and conducting national-level interviews with Russian education policymakers and experts—people working for ministerial agencies, trade unions, social media involved in producing and publicizing school rankings and research institutions. In addition to the national-level data, I make use of materials collected at the sub-national and school levels. The study of the school context was conducted in the Republic of Chuvashia (population 1.3 m), located approximately 650 km from Moscow. The region's policies of auditing education quality received acknowledgment from national and international authorities. The education reforms implemented in the republic in focus were guided by a World Bank project (2001–2006). The specific case from Chuvashia examined in this research is the city of Cheboksary (0.5 m inhabitants), the capital of the republic. The sub-national and school level data were collected primarily through participant observation and interviews with teachers and administrators conducted by my colleague and were utilized earlier in e.g. Gurova (2019).

In addition to reanalyzing³ the data cited in our published work, I returned to the initial coded extracts from interviews where respondents discuss various matters related to accountability data. The passages in the interview materials tagged with the codes “data”, “monitoring” and “statistics” in Atlas.ti are conversations on data production, analysis and use and refer to either large-scale assessments of learning outcomes or statistical administrative data on education (for more details on the overall project data, data management and coding, see Kauko, Takala & Rinne, 2018 Ch. 2 and Appendices 1–3). When necessary, I returned to the original transcribed interviews to verify and contextualize the text passage entered into the code-specific compilations of quotes.

The group of respondents consisted of school teachers, school administrators and regional authorities from the locality studied in the project as well as people working for ministerial agencies, trade unions, social media involved in producing and publicizing school rankings and research institutions located in Moscow (see Kauko, Centeno, Piattoeva, Gurova, Suominen, Medvedeva, Santos & Xingguo, 2018) for a detailed description of data collection, categories of respondents and experiences of fieldwork in Russia). This article does not make a sharp distinction between those governed by numbers and those eliciting and producing numbers in the first place. This is because the latter, too, must create narratives

³ The re-purposing of research data according to the analytical logic outlined above does not contradict the content of the consent form based on which the respondents agreed to participate in the research project. We described the research project as focusing on the politics of quality and evaluation in school education in their respective countries. The findings and arguments presented in this paper build on and further develop analytical work started during the funded phase of the project.

of why quantifications are needed and useful (Espeland, 2015, p.70), and, at the same time, contemporary neoliberal orders equally impact on those making quantifications (Desrosières, 2015). Flexible appropriation of accountability data causes many categories of education actors to be scrutinized through these data, producing mutual accountability and shared discourse (Piattoeva, 2015).

However, I pay attention to the professional position of the respondent to better contextualize and analyse their stories. The Russian national-level interview material consisted mostly of interviews with researchers and education experts, as access to ministerial officials and policymakers was restricted. The researchers interviewed engaged actively with numerical data, and were invited to participate in our study due to their somewhat vague or dual roles in between research and policymaking: they engaged with international and national large-scale assessments, or other forms of numerical student assessment—thus, assessments used by the authorities to hold to account and adjust policy. Many of these individuals had worked as policymakers or policy advisers or could be described as public intellectuals, who regularly comment on educational affairs in the public domain. Some of the researchers worked for governmental agencies during data collection, but their work was primarily analytical rather than executive.

Before turning to the empirical findings, I examine the Russian education policy context in which large-scale learning assessment data and expanding collection of statistical administrative data have become a prominent feature in policymaking and practice on the ground.

4 The rise and effects of accountability in Russia

In the early post-Soviet period, that is, the first 5 to 8 years after the collapse of the USSR, when ideas of accountability in education began to emerge, self-assessment at the level of schools, teachers and students and immediate feedback and improvement were preferred over external audit. School administrations were encouraged to become accountable to students and their families rather than to the state, and thus to define their own desired educational outcomes on the basis of the needs and expectations of their immediate stakeholders (students, families, local community, society and the economy at large), and then to seek information on whether these outcomes were achieved. Education was conceptualized as individually tailored, but still amenable to portrayal in the format of indicators, albeit of a descriptive and locally produced rather than a numerical and standardized nature (Gurova, Piattoeva & Takala, 2015).

Since the late 2000s, Russia has experienced a rise in accountability and management by results in the spirit of New Public Management throughout the public policy sector, and this development was embedded in the advice offered to the Russian authorities by major international organizations, such as the World Bank and the OECD (Organisation for Economic Co-operation and Development) (see Gusarova & Ovchinnikova, 2014; Piattoeva, 2015). The policy documents of the time stated that their successful realization was contingent upon the availability of data to monitor results and that absence of objective information would put the policy agenda

at risk (e.g. Government of Russia, 2011). This way, the government attempted to bridge the gap between policy and its implementation. Government programmes for different public sectors, including education, started to embrace detailed, quantitatively articulated objectives and a system of numerical indicators to monitor the achievement of results on an annual basis that included data from national and international assessments of learning outcomes. Currently, Russia's approach, on the macro level, is characterized by an acceptance of policies that foster accountability as an external audit of performance along with the monopoly of the state over the definition of quality and quality assurance practices (Minina et al., 2018).

The central (federal) government is directly authorized to determine subnational education policymaking (Starodubtsev, 2018). The imaginaries of a uniform national education space manifest in the implementation of national assessments and institutions that mediate assessments between the federal and the subnational, as well as the national equalization of subnational units as data (Hartong & Piattoeva, 2021). Generally, in the last 15–20 years, Russia has moved decisively towards education standards and data-based modes of more centralized school governance, including the introduction of standardized assessments. The collection of administrative data has also expanded (Gurova, 2018; Gurova & Piattoeva, 2018). The requirement to develop and implement measures of accountability, including new assessments, and to mediate assessment activities between the federal, national and the subnational, has led to the establishment of Rosobrnadzor (Federal Service for Supervision in Education and Science) including several affiliated institutions (i.e. the Federal Testing Centre, the Federal Institute of Pedagogical Measurement, the Federal Institute for Education Quality Evaluation and the National Agency for Education Accreditation). Assessments have thus paved the way for new national/federal agencies that represent central nodes in the emerging infrastructures for accountability charged with tasks ranging from the development and supervision of assessment frameworks to the construction of test items and reporting instruments (see also Hartong, 2018; Kauko, Suominen, Centeno, Piattoeva & Takala, 2018).

The Unified State Exam, the standardized school-leaving examination (Yediniy gosudarstvenniy ekzamen, USE, also called GIA-11), was first piloted by the federal authorities in several regions in 2001 and in 2009 and became compulsory for all school leavers across the country. The examination, which complies with the federal education standards (see <http://www.edu.ru/abitur/act.31/index.php>), combines the functions of the school graduation test, the university matriculation test and serves as a source of “objective” information for evidence-based policymaking and quality assurance. Even though the examination has multiple functions and is only administered to those students who have continued their education after 9 years of general education, the Ministry of Education mentions it as a key measure of education quality (www.osoko.edu.ru); the standardized graduation exam for grade nine (the State Final Attestation SFA, or GIA-9, in the last grade of general education) was later also modelled on the USE.

Other types of educational assessment procedures have recently been developed by the federal agencies, which arguably carry less weight than the USE examinations. The sample-based National Study of Education Quality (Natsional'niye issledovaniia kachestva obrazovaniia, NIKO), initiated in 2014, examines proficiency in

annually rotating school subjects and grades on primary, secondary and high school levels. NIKO data are said to inform a range of stakeholders—from teachers and parents to regional and federal authorities—about the current quality of education, while rankings are discouraged, although not prohibited by law. Finally, in 2015, the federal ministry also initiated the All-Russia Examinations (*Vserossiiskie proverochnie raboti*, VPR), testing end-of-school-year learning results in different federally mandated school subjects annually to supply information on individual learning achievements for school, local and federal actors. Even though the involvement of and control over NIKO and VPR by the federal authorities varies, i.e. whether or not federal experts are present as observers of testing situations and whether or not tests are scored internally by school teachers or externally by specially appointed expert groups, all the results are duly reported back to the federal authorities through electronic databases. The list of performance indicators introduced in 2013 already contained some items based on the number of international assessments in which Russia participates and the ranks achieved in these. Most recently, the aim of education modernization for 2018–2024 is to ensure the international competitiveness of Russian education and a position for Russia among the top ten leading countries with the best quality of education according to international education rankings (Government of Russia, 2018).

The new measurements, comparisons, public league tables and incentives tied to high performance as manifestations of state-initiated accountability have all added to rather than replaced the traditional Soviet and early post-Soviet instruments, such as reporting or inspections (Gurova, 2018; Gurova, Candido & Zhou, 2018). In the case region of our study, 80 numerical performance indicators were introduced to report on the quality of education combining measurements introduced and mandated by the Federal Ministry of Education with those stipulated by other authoritative agencies and stakeholders across levels of governance. For instance, the “quality of educational results” of schools is measured through students’ grade point averages, average scores of students and numbers of failures in national examinations, numbers of prizes won in subject Olympiads and educational contests (Gurova & Piattoeva, 2018). The same indicators serve as the criteria for teachers’ and principals’ performance-based remuneration and for the promotion of teachers to higher professional categories. In other words, numerical indicators of students’ educational achievement are highly significant for all those involved in education: students, teachers, administrators and schools.

At the school level, accountability has resulted in the expansion of written and oral reporting. Teachers prepare self-assessment reports as a requirement for calculating the performance-related part of their salaries. These include personal portfolios for compulsory teacher attestation procedures and it is important to renew the document on time, as teachers’ qualifications count in external accountability and accreditation. Every school is required to present an annual self-evaluation report on its website. Additional evaluation activities are connected to preparing students for national examinations and schools organize mock exams for compulsory achievement testing. The results of these are discussed in staff meetings and serve as a basis for modifying pedagogy (Gurova & Piattoeva, 2018; Gurova, 2018).

Thus, in schools and in sub-national administration alike, the data are primarily discussed and dealt with in terms of managing the reporting and analyzing the data to improve immediate test results. Data are mostly needed to ensure accountability, and school administrators and deputy headteachers anticipate continuous requests for data, including those demonstrating their compliance with state requirements. As a school administrator explained:

If our institution somehow violates some norms, we may lose our accreditation, we may lose our license. Hence all these monitoring studies, self-evaluation reports, all these different reports [exist] – all this is just so that the institution works as it should work by law. Do you understand? It is very serious. (cited in Gurova, 2018, p. 408)⁴

5 Number narratives

5.1 Data objectivity

In the study reported here, the interviewees often mentioned a conflict between the demand for evidence-informed decisions on the one hand, and the fast pace of performance evaluation on the other hand (Piattoeva, Centeno, Suominen, Rinne, 2018). Nevertheless, researchers, experts and policymakers were attracted to numbers as a means of making education “reality” known, seeing the expansion of testing procedures as a means of gaining a better overview. The quote below shows that even a respondent overtly critical of measuring education through narrowly defined, countable proxies, proposes expanding and enriching numbers to make them “truer” to what they claim to represent rather than abandoning numerical representations. The interviewee makes use of the medical metaphor of symptoms and syndromes to describe the problem of distance and depth, and to call for complex measurements to get closer (in both distance and depth) to the reality of education.

Using a clinical language, which to me is closer in clinical psychology, many systems of quality evaluation [focus on] how many tables there are in the school, how many computers there are in the school. They identify symptoms, not syndromes. You may have a cold, but it doesn't tell your illness, it is on the surface. If we [act] superficially, without seeing the syndromes and the overall illness of the system, we will lose. Thus indices of [...] should be developed on

⁴ The interviews are tagged according to the interviewees' positions in the field of education at the time of the empirical work, following the system introduced in the final, landmark project publication (Kauko, Takala & Rinne, 2018). In the case of interviews or textual citations referring to earlier published work, the tagging system of the source is retained. As mentioned in the article, most interviewees held or had held various positions, simultaneously or not; therefore, the tag is inevitably a simplification of a more complex and nuanced professional background and identity. The code abbreviations denote: N = national level actor (federal or union level politician or civil servant); SN = sub-national level actor (region or municipality level politician or civil servant); E = Expert.

the basis of a vast amount of qualitative indicators, not one-sided parameters that would be extremely dangerous for education. (NE-01)

However, at the national level, policymakers were mainly concerned with accumulating data in a standardized, detached and commensurate manner—in safeguarding a mechanical, procedural objectivity (Porter, 1994) that removes contextual “noise” from data collection, analysis and representation. At the same time, they expected regions and schools to react to data contextually, recontextualizing data for targeted decision-making. National-level actors criticized sub-national authorities if they did not follow suit, that is, did not address the problems “revealed” by data contextually by probing further their underlying causes.

The [regional ministers] gather here several times a year to receive feedback on their successes, the successes of regions, their different positions. [...] The situation is monitored rigorously, and regional ministers receive objective feedback about where they stand, including their position compared to other regions. How good or poor their achievements are. (N-04)

So, the most important thing is, in fact, that everyone knows that they are all measured equally, say, by a machine, people do believe that a machine can measure and count objectively and not because it likes or dislikes me. As a manager I have to say that, if we forget about the children, which we shouldn't do, then, of course, these are supermanagerial mechanisms. (SNE-01)

The All-Russia Examinations are needed as a means of objective feedback for schools. And, overall, the general logic of the new wave of changes in the system, quality assessment developed by Rosobrnadzor, and us too, is to lower the fiscal burden on the schools, to lower the volume of final accountability (that is often pointless for schools) towards objective feedback [...] It is important to move to independent and objective forms, as if from the outside. [...] We stimulate education administration, the schools themselves, to administer these honestly with a view to improving the quality of teaching. This is the main goal; unfortunately, there are examples, two negative points, in the competences of administration. (N-04)

The notion of “objective feedback” is perplexing because in this context, feedback means increasing volumes of plain numbers circulated, first, from the sub-national level to the federal centres of calculation, to be verified and organized there into grids and rankings, and sent back. Policymakers rarely expressed concern with context in any other way than worrying about the unreliability of numbers if the processes of data collection were left to the discretion of regions or schools. Yet they expected numbers to travel to the local level, where subnational actors would make better sense of them. This situation is reflected in the experiences of schools and teachers who often described the punitive conditions in which data are collected:

Nevertheless, all teachers are worried. One month before, during every lesson, we discussed how to write an essay. We sent a very complicated form for accountability. Then I have a big sheet: every child [is inserted on the sheet] according to five criteria, “yes” or “no”. Then I sign, and hand it over to the

person responsible for the essay exam. These are all fed into a computer, and she [the person responsible] sends it on. Everything is somehow coded. And we were lucky that no inspector came to our school. They could've come, spent time, spent, by the way, money, because [...] the department would've had to pay. (NE-13)

While policymakers organize decontextualization of assessment and the actual means of collecting assessment data objectively, and teachers are responsible for ensuring procedural objectivity in practice, another way to see the objectification of context is through the verification that takes place when the actor producing the data analysis and displaying the data in public does not trust the origins of the data—when data production has not yet been satisfactorily standardized and made transparent (cf. above):

I will tell you, in fact, we have three or four pairs of eyes there. And I as the person first in line will, naturally, examine each product before it goes public, each database. [...] If it is a league table, we travel to see the leading organization, so, for instance, we see that a kindergarten in, say, Volgograd oblast or Vladimir, holds the first position, we will find a way to go there to observe it on site. It is not only information from open sources, but what is the reality, because we cannot accept the situation of having an organization ranked first, when in reality, it is of some different kind and does not verify the story. (NE-02)

For actors producing data, trust in numbers is paradoxically both strong and fragile. When numbers are used in the politics of accountability or high-stakes decision-making on the allocation of funds or could have damaging effects on one's reputation, they become incentives to manipulation and mistrust (Lim, 2020). Trust in numbers is thus contingent upon trust in the expertise and procedures behind those numbers. In the Russian context, especially at the time of data collection, scandals of data falsification frequently appeared in the media and had a harmful effect on the reputation of the USE data. This further fuelled prolonged debates on the overall problematic nature of the USE as a standardized and compulsory graduation exam and, simultaneously, a uniform measure of education quality. In response, the government introduced stringent surveillance measures to restore the reputation of the data (cf. Piattoeva, 2016). This explains the meticulous attention to the procedural standardization of data collection reported above.

5.2 Longing for context

The relation to context was very different for the researchers interviewed, as I explain next. The idea that numbers should represent context in richer terms is evident in the next citation, where a respondent discusses international large-scale assessments and their lack of depth to understand the heterogeneity of education inside Russia. However, as already described in the preceding section, the belief that diversifying numerical data is the answer still prevails:

*Unfortunately, the data is not representative on the regional level. Our country is very diverse [...] There is a saying “average temperature in the hospital” and this is what this is [the respondent means that ILSA data offer such an average account]. If I would have data on all regions, representative of the region, then we could talk about serious analysis, because for now, yes, the socio-economic background of the family [the respondent means that ILSAs have established a correlation between socio-economic factors and ILSA scores]. Yes, money doesn’t have a strong effect. Yes, teacher quality has a minor effect. But in order to make decisions [...] we need to look at a much lower level because decisions regarding Moscow and decisions regarding Tyva will be radically different, and factors that affect education there will be radically different.*⁵ (NE-03)

Nationally collected data were equally regularly characterized in terms of lacking in contextual detail and depth. USE data, for instance, were not only hard to obtain for academic use, but when available in an aggregate format, lacked contextual information to enable interpretation. For some experts, this situation was incomprehensible, that is, why collect such vast and expensive data if they are not available for scholarly purposes to support evidence-informed decision-making? Ironically, this situation caused some of the interviewees to be more attracted to the ILSA data, claiming that they offer unrestricted access to a large volume of background information to explore correlations and even causal relationships between achievement and variables like family support, school administration or resources available. All in all, then, the researchers’ overarching narrative combined the appeal of quantitative data with frustration about the lack of contextual detail to engage with performance data in a meaningful way.

The ministry and Rosobrnadzor are keen to keep to themselves all the information which they receive from the [regional] centres of quality evaluation, from the regional departments and ministries, and it is very difficult for society to obtain this information. For example, Russia took part in TALIS, you know..... So now in order to productively analyse TALIS data, we need the USE [Unified State Exam – a national standardized graduation examination after grade 11] results of the schools that took part in TALIS. These are only 200 from 14 or 15 regions. We cannot obtain this information. Can you imagine? We are the national coordinator, our institute is the operator of the project, we need this information not 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.” (NE-04) (cited in Piattoeva, Centeno, Suominen & Rinne, 2018, p. 125.)

⁵ The respondent refers to Tyva (also Tuva) in order to underline the diversity of Russian regions. The Tyva Republic is located 3700 km away from Moscow and lies at the geographical centre of Asia, in southern Siberia. Its population is slightly above 300,000 (2010 census), while the population of the city of Moscow exceeds 12 million (Wikipedia).

6 Numbers misrepresenting contexts

The paradox that I turn to last is how the amplification and complexification of data—and the belief that more data will mean more context—are in fact manifest in the narratives of concurrent misrepresentation in two intertwined meanings of the term: misrepresentation as a misleading account and misrepresentation as exclusion. The schools are becoming producers of large amounts of numerical data and reporting, and “data on data production” (Piattoeva, 2016). The main point is that the availability of data will enable actors, ranging from administration to parents, to know the characteristics and performance of schools, teachers and students to make decisions and launch interventions that exert a direct influence on the reputation, financial means or future operations of the accountable actors. The paradox of the situation lies in the fact that educators collect or produce the very data that are then used to “[m]onitor, measure, and, potentially, punish them” (Anagnostopoulos & Bautista-Guerra, 2013, p. 56), but these generate realities that are perceived as misrepresentations. Moreover, the measures invented at the federal and regional levels as synoptic overviews, become, in the school context, individualizing instruments evaluating individual employees. In this manner, they turn from detached and impersonal to embodied personal accounts.

A common phrase to describe authorities, national and local alike, is “they have not worked at school for a single day”. In the views of both teachers and administrators, policymakers are remote and uninformed, driven primarily by the desire to create an image of their own efficiency (Gurova & Piattoeva, 2018, p. 183). The measures of audit and accountability that are meant to convey knowledge of schools to the centres of calculation and decision-making end up widening the distance, at least as seen from the perspective of school practitioners. And this feeds into the sensation of misrepresentation. Even though our interviews and observations were not initially guided by interest in the question of misrepresentation, it is clear that the relationship of school professionals to accountability is construed through their interpretation of these as misrepresenting education in general and teachers’ work in particular. The latter concerns both teaching in an abstract sense of vocation and teaching as an everyday activity consisting of distinct, entangled tasks. Teachers take issue with both the idea that pedagogical “impact” is measurable and with the metrics deployed to measure performance. I provide examples of these next.

6.1 Misrepresenting education

Misrepresenting education is described, for instance, in terms of education quality being something very different from what is measured, that is, not being describable in terms of short-term learning outcomes. Nor was education seen as being in the hands of a single teacher—thus taking issue with the individualizing nature of the measurements and the fact that they ignore the complexity of school as an evolving community.

[Administrator 3]: Quality means something else. Quality of education materializes in one’s adjustment to life, how a person finds a place in life. Not in

academic achievements. Sometimes you see straight A students who can't find a place (...) and there are mediocre ones (...) but their lives turn out perfect. So the new education standards are correct in their practice orientation (...), but no one knows how to put them into practice. (cited in Gurova & Piattoeva, 2018, p. 179)

It is very, very difficult to evaluate the contribution of an individual teacher in the overall input of the school. We can't know what exactly influences a child's results. No one can. They will have to build utopian schools (...): if we isolate the children from their parents, cut off the internet, television, stop their interaction with anyone except for the wise specialists, then, in theory, perhaps, we'll be able to establish the influence of the school. So, it is impossible to tell that a student achieved results thanks to the work of a given teacher. (NE-13)

Many existed then, but now, it is principally nonalgorithmic issue. I cannot evaluate how I have improved patriotism in this concrete student, there are no criteria. Or I cannot say how ethnically tolerant a student is, that is, he was graded three and now became three plus, so I have done my work well. The standard is full of such indicators that should be evaluated. In results, I must report, and I write all sorts of nonsense. (NE-14)

6.2 Misrepresenting teaching as a vocation

Teaching as a vocation was described as being remote from the requirement of accomplishing certain pedagogical activities merely for the sake of their being identified as important by an external actor, motivated by reporting rather than professionalism.

One more important issue: who is controlling and what are the criteria of inspections? Today, these criteria are paper-based, absolutism of numbers, technically this looks like a scoring system. These performance-based components of the salary, the teacher should amass a certain number of points, his every sneeze is counted as a certain number of points to receive the performance component. The teacher is engaged in collecting points rather than work. But his vocation is different. A normal person would not talk to the parents and then run to tick a box that I've done it. (NE-14)

6.3 Misrepresenting everyday work

Performance-based payment is calculated individually for each teacher and does not take into account the collaborative nature of school work, let alone the fact that students' family backgrounds vary a lot. Teachers are randomly assigned to different classes in different years. They are assisted by other teachers and staff, but their performance is determined solely by the results of their students of that academic year. A large part of the work that teachers perceive as central elements of their everyday work, including lesson preparation, motivating students, liaison with parents

or working as the class teacher (“klassnyy rukovoditel”), is not included in the performance metrics.

[Teacher 23]: The teachers are ranked on the basis of the average grade, the quality percentage. (...) Of course, this is unpleasant. There are different children in the classes, in primary school - we sort them out - someone works in a “difficult” class, and will end up with the lowest ranking, even though he/she may be a very good teacher, highly qualified. Or I substitute in another class, and then the teacher of that class gets a better ranking. (cited in Gurova & Piattoeva, 2018, p.180)

[Teacher 26]: There is an internal ranking of teachers - whose students rank highest in GIA. (...) It is one thing if I have been with this class since grade seven, or five - then I can be held responsible. But what if I have only taught them in grades ten and eleven? (cited in Gurova & Piattoeva, 2018, p.180)

An administrator described the actual schoolwork as something only briefly reflected in what is controlled and measured by the authorities, or even paid for:

[Administrator 1]: Our work does not end with official hours. There are even more working hours that are not compensated at all. We never worked for the sake of salary. (...) What is requested from us: grades, results in the GIA, percentages of students entering higher education, crime rates. Our projects interest no one. We can work around the clock, or not at all, no-one will care. (cited in Gurova & Piattoeva, 2018, p.180)

Local education authorities demand “analytical reports” (i.e. reports containing numerical information) not only in connection with teaching, but also with many other activities, such as the organization of sports and patriotic education events, or school measures for drug abuse and crime prevention. The quantities of documentation requested are such that it is not feasible to organize all the activities. Neither is it necessary: the only controlled request is to provide a timely report (Gurova & Piattoeva, 2018).

The purpose of “objective” evaluation is to reveal problems in learning which without external control would remain hidden or go unnoticed, and to discipline teachers to assign grades more accurately (reflecting the learning results). However, the intention to reveal problems contradicts the official requirement to produce good results and compliance with state regulations. The indicator named “quality percentage” calculated from end-of-quarter and end-of-year grades assigned by teachers is particularly problematic. It creates incentives for teachers to give higher marks to push up the performance scores, but at the same time serves as a control measure to ensure that teachers do not fabricate high grades. It also treats grades as an absolute measure of achievement and disregards the fact that teachers may use grades to stimulate students.(Gurova & Piattoeva, 2018)

[Teacher 23]: On the one hand, we are reprimanded for having given a student the lowest grades. On the other hand, if we give satisfactory grades, but the child does not pass the GIA or receives the lowest grades the following year - we are reprimanded again: “Either you falsified the grade or the new

teacher cannot teach well”- they say. In any case the teacher always gets the blame, the entire responsibility is on him/her. (cited in Gurova & Piattoeva, 2018, 182)

7 Discussion

Numbers exercise power by circulating in the world (Beer, 2016; Piattoeva, 2015). In this paper, I examined circulation as subtraction and addition of contextual and context-referencing narratives of what measures are, what they are good for and what they should achieve (cf. Beer, 2016; Espeland, 2015). This interest was motivated by the observation that quantification is an act of de- and recontextualization. Numbers produced to hold actors to account balance between a demand for decontextualized centralized accountability and contextualized decentralized action. A further premise was that in an era of neoliberal governance of states, numbers enjoy the reputation of being true to reality and free from human bias. The practice of holding individual actors to account by calculating and representing their performance numerically manifests the blurring of the distinction between individual and collective issues (Jany-Catrice, 2016). That is, indicators that once measured the aggregate performance of collective actors are (simultaneously) used to measure the respective performances of individual actors.

In the analysis, I paid attention to the role and position of the respondent in the accountability system as I expected this to lead to different number narratives (cf. Asdal, 2011). The respondents discussed data interchangeably as data for accountability and data for knowledge production, and the distinction between the two usages was not always clear-cut. State-level actors used numbers to verify the allocation of public funds and the adherence of schools to the federal standards and expected learning outcomes. Media actors enable numbers to travel beyond the state and their purpose is to make “education quality” known to the wider public, thereby authorizing the public to monitor education institutions and supporting or at least expecting the public to make informed decisions based on data frequently presented in the form of rankings. These actors’ conventional critique of measuring education numerically focused on modifying and often expanding how education is measured. Their critique called for more “objective”, “valid” and “reliable” indicators, manifesting how critiques cannot bypass the language of quantification and accountability (cf. Taubman, 2009; Piattoeva & Saari, 2018). It is deemed paramount that the processes of producing data be rigorously standardized to prevent subjective distortions. The researchers criticized the lack of contextual data affording a richer and more meaningful understanding of the numerical data. The teachers and school administrations regularly described data in terms of being removed from the actual workings of the school. Paradoxically, the amplification and complexification of data that paralleled the evolution of accountability policies led to narratives of concurrent misrepresentation in two intertwined meanings of the term: misrepresentation as a misleading account and misrepresentation as exclusion.

The narratives of standardized assessments and administrative statistics examined here as data for accountability suggest that the expansion of quantification is

further powered by the experiences of a trade-off between commensuration and context. Quantification, then, engenders both numerical simplifications and qualitative accounts. Authorities across levels are concerned with the objective reputation of data, expanding the collection of data for accountability and evidence-informed decisions, although the nature of the latter remains ambiguous. People affected by this regime do not take numbers at face value but evoke narratives that explain their virtues and limitations, promises and failings etc. (cf. Espeland, 2015). These accounts point to different forms of numeracy, sense-making and emotional charge that encompass quantification. But all participants, in one way or another, contribute to the production, continuation or change in the project of quantification (Gorur et al., 2019).

Methodological sensitivity to number narratives is anchored in approaches to quantification as performative and generative (e.g. Porter, 1994, Kitchin, 2014, Espeland & Stevens, 1998, Merry & Wood, 2015, Beer, 2016, Rottenburg & Engle Merry, 2015, Mau, 2017), reminding us that instead of merely condensing reality, numbers also leave us with more, not less (Sellar, 2015, p. 132). Simplified abstractions as products of commensurating practices “are also added to the world”, leaving us with “the complex qualities subject to commensuration and the simplified representations produced through this process” (Sellar, 2015, p. 132). I suggest that narratives of numbers add a third, perhaps intermediary “layer”: we are left with complex qualities, commensurate abstractions and narratives that bridge the two. Thus, narratives are not the opposite of numbers but their very lifeblood. Understanding numbers in this way helps us to examine the equivocal practices of quantification and their role in accountability policies.

Based on these, I claim that enactment of accountability is fuelled by quantified de- and recontextualization in two ways. First, returning to the argument of Ted Porter that numbers constitute a means of communication that transcends contexts, quantification per se seems to make policies mobile and enables them to reach out to and assemble diverse actors to enact accountability in their local contexts. Second, the decontextualizing quality of quantification generates diverse ways in which actors experience, make-sense of and act upon quantified accountability. As a result, quantification and its decontextualizing attributes may operate as a red herring—they redirect actors’ attention to and engage them in the possibilities and pitfalls of quantification that enable the politics of accountability and its consequences to persist in the background. Related to this observation, further work might explore whether absence of substantial resistance to accountability policies (Shore & Wright, 2015) could be explained by the deceptive grip of quantification on actors’ imaginary, practices and sense-making of accountability reforms.

On an ethical note, attentiveness to and analysis of narratives are the researcher’s tool for contextualizing and taking issue with quantified accountability by reinstating the distinctiveness of the quantified entities—whether these are people or schools. Inviting people to recount numbers on their own terms offers a means of engaging in a reverse form of accountability where people can hold numbers to account.

Funding This research was supported by the Academy of Finland grant 273874 for the research project “Transnational Dynamics of Quality Assurance Policies and Evaluation in Brazil, China and Russia” 2014–17.

Availability of data and material (data transparency) Available on request from the author.

Declarations

Conflicts of interest/Competing interests Nothing to declare.

Numbers and their contexts How quantified actors narrate numbers and decontextualization.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anagnostopoulos, D., & Bautista-Guerra, J. (2013). Trust and numbers: Constructing and contesting statewide information systems. In D. Anagnostopoulos, S. A. Rutledge, & R. Jacobsen (Eds.), *The infrastructure of accountability: Data use and the transformation of american education* (pp. 41–56). Harvard Education Press.
- Asdal, K. (2011). The office: The weakness of numbers and the production of non-authority. *Accounting, Organizations and Society*, 36(1), 1–9.
- Ball, S. J. (2003). The teacher’s soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How Schools Do Policy. Policy Enactments in Secondary Schools*. London: Routledge.
- Bansel, P. (2015). A narrative approach to policy analysis. In K. Gulson, M. Clarke, & E. Bendix Petersen (Eds.), *Education policy and contemporary theory: Implications for research* (pp. 183–194). Routledge.
- Beer, D. (2016). *Metric power*. Palgrave Macmillan.
- Benavot, A., & Tanner, E. (2007). *The growth of national learning assessment in the world, 1995–2006*. UNESCO.
- Boltanski, L., & Chiapello, È. (2005). The new spirit of capitalism. *International Journal of Politics, Culture, and Society*, 18(3–4), 161–188.
- Boltanski, L., & Thévenot, L. (2006). *On justification: Economies of worth*. University Press.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 585–596. <https://doi.org/10.1080/01596306.2011.601555>
- Chase, S. (2011). Narrative inquiry: Still a field in the making. In N. Denzin & Y. Lincoln (Eds.), *Sage handbook of qualitative research* (4th ed., pp. 421–434). Thousand Oaks, CA: Sage.
- Chase, S. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In N. Denzin & Y. Lincoln (Eds.), *Sage handbook of qualitative research* (3rd ed., pp. 651–679). Thousand Oaks, CA: Sage.
- Chase, S. E. (2018). Narrative inquiry: Toward theoretical and methodological maturity. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (5th ed., pp. 546–560). SAGE.

- Crow, D., & Jones, M. (2018). Narratives as tools for influencing policy change. *Policy & Politics*, 46(2), 217–234.
- Desrosières, A. (1998). *The politics of large numbers. A history of statistical reasoning*. Harvard University Press.
- Desrosières, A. (2015). Retroaction: How indicators feed back onto quantified actors. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The world of indicators. The making of government knowledge through quantification*. Cambridge University Press.
- Espeland, W. (2015). Narrating numbers. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The world of indicators: The making of governmental knowledge through quantification* (pp. 56–75). Cambridge University Press.
- Espeland, W., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24, 313–343.
- Espeland, W., & Vannebo, B. (2007). Accountability, quantification, and law. *Annual Review of Law & Social Science*, 3, 21–43.
- Fenwick, T., & Edwards, R. (2011). Considering materiality in education policy: Messy objects and multiple reals. *Educational Theory*, 61(6), 709–726.
- Fougnier, T. (2008). Neoliberal governance of states: The role of competitiveness indexing and country benchmarking. *Millennium: Journal of International Studies*, 37(2), 303–326.
- Gorur, R. (2015). Situated, relational and practice-oriented: the actor-network theory approach. In K. N. Gulson, M. Clarke, & E. Bendix Petersen, (Eds.), *Education policy and contemporary theory : implications for research* (pp.123-143). London: Routledge
- Gorur, R. (2018). Escaping numbers? Intimate accounting, informed publics and the uncertain assemblages of authority and non-authority. *Science & Technology Studies*, 31(4), 89–108.
- Gorur, R., Maddox, B., & Sørensen, E. (2019). Standardizing the context and contextualizing the standard: Translating PISA into PISA-D. In M. J. Prutsch (Ed.), *Science, numbers and politics*. Palgrave.
- Government of Russia (2011). Conception of the Federal Programme of Educational Development for 2011–2015. <http://elementy.ru/Library9/r163r.htm> Accessed 13 November 2015 [in Russian].
- Government of Russia (2018). Passport of the national project 'Education'. <http://mo.mosreg.ru/download/document/1524233> Accessed 24 November 2018 [in Russian].
- 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). Routledge.
- Gurova, G. (2018). Soviet, post-Soviet and neo-liberal: Governing Russian schools through quality assurance and evaluation. *Policy Futures in Education*, 16(4), 398–415. <https://doi.org/10.1177/1478210317743648>.
- Gurova, G. (2019). *Quality Assurance and Evaluation as a Mode of Local Education Governance: The Case of Russian Schools. Doctoral dissertation*. Tampere: Tampere University Press.
- 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. A. Oliveira (Eds.), *Education policies and the restructuring of educational professions* (pp. 173–187). Springer
- Gurova, G., Candido, H., & Zhou, X. (2018). Effects of quality assurance and evaluation on schools' room for action. In: J. Kauko, R. Rinne, T., & Takala (Eds.), *Politics of quality in education: a comparative study of Brazil, China, and Russia*. London: Routledge
- Gurova, G., Piattoeva, N., & Takala, T. (2015). Quality of education and its evaluation: an analysis of the Russian academic discussion. *European Education*, 47(4), 346–364.
- Gusarova, M., Ovchinnikova, M. (2014). Introducing performance management in the system of public administration in Russia: Approaches and lessons learnt from the last decade of implementation. *Voprosy Gosudarstvennogo i Munitsipal'nogo Upravleniya* 1 [Public Administration Issues]. www.ecsocean.hse.ru/mags/en/vgu/2014-1.html Accessed 22 May 2015 [in Russian].
- Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press
- Hansen, P., Wallenius, T., Juvonen, S., & Varjo, J. (2020). Moving landscapes of Nordic basic education: Approaching shifting international influences through the narratives of educational experts. *Compare: A Journal of Comparative and International Education*, 50(6), 773–791. <https://doi.org/10.1080/03057925.2018.1557509>

- Hartong, S. (2018). Towards a topological re-assemblage of education policy? Observing the implementation of performance data infrastructures and 'centers of calculation' in Germany. *Globalisation, Societies and Education*, 16(1), 134–150. <https://doi.org/10.1080/14767724.2017.1390665>
- Hartong, S., & Piattoeva, N. (2021). Contextualizing the datafication of schooling – a comparative discussion of Germany and Russia. *Critical Studies in Education*, 62(2), 227–242.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Jany-Catrice, F. (2016). Evaluating public policies or measuring the performance of public services? In I. Bruno & F. Jany-Catrice (Eds.), *The social sciences of quantification: from politics of large numbers to target-driven policies - logic, argumentation & reasoning* (pp. 123–135). Springer.
- Kauko, J., Rinne, R., & Takala, T. (Eds.). (2018). *Politics of Quality in Education: A Comparative Study of Brazil, China, and Russia*. London: Routledge.
- Kauko, J., Centeno, V. G., Piattoeva, N., Gurova, G., Suominen, O., Medvedeva, A., Santos, I. & Xing-guo, Z. (2018). Layers of Reflectivity in Comparative Research. In J. Kauko, R. Rinne, & T. Takala (Eds.). *Politics of Quality in Education: A Comparative Study on Brazil, China, and Russia* (pp. 18–43). Routledge
- Kauko, J., Suominen, O., Centeno, V. G., Piattoeva, N., & Takala, T. (2018). Established and emerging actors in the national political arena. In J. Kauko, R. Rinne, & T. Takala (Eds.), *Politics of quality in education: a comparative study on Brazil, China, and Russia* (pp. 71–90). Routledge
- Kipnis, A. B. (2008). Audit cultures: Neoliberal governmentality, socialist legacy, or technologies of governing? *American Ethnologist*, 35(2), 275–289.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Law, J., & Ruppert, E. (2013). The social life of methods: Devices. *Journal of Cultural Economy*, 6(3), 229–240.
- Le Galès, P. (2016). Performance measurement as a policy instrument. *Policy Studies*, 37(6), 508–520.
- Lim, M. A. (2020). Impact case studies: What accounts for the need for numbers in impact evaluation? *International Studies in Sociology of Education*, 29(1–2), 107–125. <https://doi.org/10.1080/09620214.2019.1624591>
- Lynch, M. (2019). The narrative of the number: Quantification in criminal court. *Law & Social Inquiry*, 44(1), 31–57.
- Mau, S. (2017). *Das metrische Wir: Über die Quantifizierung des Sozialen*. Suhrkamp.
- Mau, S. (2020). Numbers matter! The Society of Indicators, Scores and Ratings. *International Studies in Sociology of Education*, 29(1–2), 19–37. <https://doi.org/10.1080/09620214.2019.1668287>
- Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45(1), 223–245.
- Merry, S. E., & Wood, S. (2015). Quantification and the paradox of measurement. *Current Anthropology*, 56(2), 205–217.
- 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 Dobrochinski H.H. (2018). Transnational policy borrowing and national interpretations of educational quality in Russia, China, and Brazil. In M. Chankseliani, I. Silova & (Eds.), *Comparing post-socialist transformations: purposes, policies, and practices in education* (pp. 27–44). Oxford Studies in Comparative Education. Symposium Books.
- Morita, A. (2014). The ethnographic machine: Experimenting with context and comparison in Strathernian ethnography. *Science, Technology, & Human Values*, 39(2), 214–235.
- Mugler, J. (2015). By their own account: (Quantitative) accountability, numerical reflexivity, and the national prosecuting authority in South Africa. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The world of indicators: The making of governmental knowledge through quantification* (pp. 76–101). Cambridge University Press.
- Phoenix, A. (2013). Analysing narrative contexts. In M. Andrews, C. Squire, & M. Tamboukou (Eds.), *Doing narrative research* (pp. 72–87). SAGE Publications, Ltd. <https://doi.org/10.4135/9781526402271>
- 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.
- Piattoeva, N. (2018). How can Transnational Connection Hold? An Actor-Network Theory Inspired Approach to the Materiality of Transnational Education Governance. In A. Wilkins, & A. Olmedo

- (Eds.), *Education governance and social theory: interdisciplinary approaches to research* (pp. 103–119). Bloomsbury
- Piattoeva, N., & Saari, A. (2018). The infrastructures of objectivity in standardized testing. In B. Maddox (Ed.), *International large-scale assessments in education* (pp. 53–68). Bloomsbury
- Piattoeva, N., & Boden, R. (Eds.). (2020). Escaping numbers? The ambiguities of the governance of education through data. *International Studies in Sociology of Education*, 29(1–2).
- Piattoeva, N., Centeno, V. G., Suominen, O., & Rinne, R. (2018). Governance by data circulation? The production, availability, and use of national large-scale assessments data. In J. Kauko, R. Rinne, & T. Takala (Eds.), *Politics of quality in education: a comparative study on Brazil, China, and Russia* (pp. 115–136). Routledge
- 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). Duke University Press.
- Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. *Human Relations*, 57(6), 765–783.
- 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 University Press.
- Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale Univ. Press.
- Sellar, S. (2015). A feel for numbers: Affect, data and education policy. *Critical Studies in Education*, 56(1), 131–146.
- Shore, C., & Wright, S. (2015). Governing by numbers: Audit culture rankings and the new world order. *Social Anthropology/anthropologie Sociale*, 23(1), 22–28.
- Squire, C., Andrews, M., & Tamboukou, M. (2013). What is narrative research?. In Andrews, M., Squire, C., & Tamboukou, M. (Eds.), *Doing narrative research* (pp. 1–26). SAGE Publications, Ltd. <https://doi.org/10.4135/9781526402271>
- Starodubtsev, A. (2018). *Federalism and regional policy in contemporary Russia*. Routledge.
- Strathern, M. (Ed.). (2000). *Audit cultures: Anthropological studies in accountability, ethics, and the academy*. Routledge.
- Taubman, P. M. (2009). *Teaching by numbers: Deconstructing the discourse of standard and accountability in education*. Routledge.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.