

The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms

Antoni Verger, Lluís Parcerisa & Clara Fontdevila
Sociology Department, Universitat Autònoma de Barcelona

Abstract

The Global Education Reform Movement (GERM) is expanding internationally and reaching countries that seemed to be immune to this education reform approach until quite recently. Accordingly, more and more educational systems in the world are articulated around three main policy principles: accountability, standards and decentralisation. National large-scale assessments (NLSAs) are a core component of the GERM; these assessments are increasingly used for accountability purposes as well as to ensure that schools achieve and promote centrally defined and evaluable learning standards. In this paper, we test these premises on the basis of a new and original database on NLSAs, as well as on data coming from the Programme for International Students Assessment (PISA) questionnaires. In the paper we also discuss how different theories on policy dissemination/globalisation explain the international spread of NLSAs and test-based accountability worldwide, and reflect on the potential of a political sociology approach to analyse this globalising phenomenon.

Keywords: standardized tests; educational assessment; test-based accountability; policy instruments; educational policy; globalization

Introduction¹

The Global Education Reform Movement (GERM) is a metaphorical concept that indicates that most educational reforms being currently adopted worldwide respond to similar problems and priorities and address these problems from a very similar managerial approach. Accountability, standards, decentralisation and school autonomy are the main policy principles of the GERM (Sahlberg 2016). As we show in this paper, national large-

¹ An English version of this article has been published in *Educational Review*

scale assessments are the policy instrument that brings these policy principles together in a coherent reform package.

National large-scale assessments (NLSAs) are not new elements in educational systems. They have been historically used for students' certification purposes, or to diagnose the problems and challenges that education systems face (Jones 1996; Kamens and McNeely 2010). However, in the context of the global education reform movement, NLSAs are being increasingly used to monitor the delivery of standardised curricula and to hold schools, principals and teachers more accountable.

The origin of the GERM is usually associated to the neoliberal education reforms adopted since the 1980s in countries such as the US, the UK and Australia (Sahlberg 2016). These reforms – or some of their main components – would be later exported to lower-income countries and to transition economies by international development banks through lending conditionality (Hargreaves 2001; Klees 2008). The GERM is a globalising phenomenon, although according to Sahlberg (2016) not all education systems have become “infected” by the GERM yet – and, in fact, some of them have even expressed reservations on the potential and suitability of this approach. According to this author, this would be the case of East Asian countries such as Japan and South Korea, and European countries such as France, Germany, Denmark, Norway, Belgium and, of course, Finland.

In this paper, we empirically show that, although Anglo-Saxon and aid-dependent countries were the early adopters of the GERM, this approach is currently reaching many other countries, including those European and OECD countries that seemed to be immune to the GERM until quite recently. We test this premise on the basis of a new and original database on national students' learning assessments, as well as on the Programme for International Students Assessment (PISA) questionnaires' data. National assessments are

a core component of and a central instrument in the GERM. As our data shows, NLSAs are used more and more for accountability purposes as well as to ensure that, in increasingly decentralised education systems, schools achieve and comply with centrally-defined learning standards. We dedicate the central parts of the paper to test these arguments. In the last section of the paper, we discuss how different global policy (or policy dissemination) theories explain the international spread of NLSAs and test-based accountability worldwide, and we reflect on the possibilities and potential of a political sociology approach (cf. Lascoumes and Le Galès 2007; and Kassim and Le Galès 2010) to analyse this globalising phenomenon.

The Global Education Reform Movement: the core function of NLSAs

On the origins of the GERM

Globalisation, understood as the increasing interdependence and competition between territories and actors in the economic, political and cultural domains, has contributed to deterritorialise education policy processes, and to national actors and spaces losing their centrality in the definition of educational reforms (Robertson 2012). To a great extent, globalisation implies the redefinition of the scale, space and dynamics through which education policy is being negotiated, formulated and implemented (Verger, Novelli, and Altinyelken 2018). In this context, international organisations and transnational networks of experts have an increasing capacity to determine education agendas and to define country priorities through both their discursive and material resources (Dale 1999).

The globalisation of education policy spaces and the increasing competitive dynamics that gear around educational systems have established the conditions for the emergence and rapid dissemination of the so-called *Global Education Reform Movement* (GERM) (Sahlberg 2005). The GERM is a global education reform package that aims at

strengthening the effectiveness of educational systems and is essentially articulated around three main education policy principles, namely standards, accountability and decentralisation (Gable and Lingard 2015; Sahlberg 2016). According to Sahlberg (2016), the idea of the GERM was first elaborated by Hargreaves and his collaborators in the book *Learning to Change: Teaching Beyond Subjects and Standards* (2001). In this book, the authors point out that:

“[A] new, official orthodoxy of educational reform is rapidly being established in many parts of the world. This is occurring primarily in predominantly Anglo-Saxon countries but through international funding organizations such as the World Bank and the global distribution of policy strategies, elements of this orthodoxy are increasingly being exported in many parts of the less-developed world as well” (Hargreaves et al. 2001, 1; as quoted in Sahlberg 2016, 132).

Originally, the GERM emerged in countries that were embracing the neoliberal paradigm in public services provision, such as the US and the UK (Robertson 2015). This reform approach would be later spread toward Latin America, Eastern Europe and the former Soviet Union with the financial support of international development banks. However, more and more countries are currently embracing the GERM for reasons not directly related to the promotion of pro-market reforms in education. To a great extent, the expansion of the GERM is intimately associated to the student-centred paradigm of learning and to the fact that improving learning achievement has become the main goal of most educational reformers and international organisations (Sahlberg 2005, 2016).

Currently, standardised and measurable learning outcomes are conceived as the most relevant indicator of education quality. Many policy-makers and scholars have conflated education quality to the more specific idea of students' achievement, not necessarily for ideological reasons, but due to the fact that learning outcomes are more

concrete, comparable and measurable than other types of more context-sensitive education quality indicators (Verger and Parcerisa 2018). To respond to these new demands and rationales for quality assurance, many governments have developed learning standards that are aligned and combined with the implementation of NLSAs that focus on the core subjects of the national curriculum (Smith 2016). Nonetheless, the GERM cannot be isolated from broader changes in public sector administration, and from the fact that public sector reforms are increasingly informed by the tenets of New Public Management (NPM). Decentralisation, the fragmentation of public services into more autonomous providers, outcomes-based management and accountability are not principles that have exclusively penetrated the education sector, but all types of policy domains - frequently through NPM reforms (Gunter et al. 2016).

The key role of national assessments within the GERM

Although the GERM is not a monolithic model of education reform and can evolve through different policy manifestations (Ball, Junemann, and Santori 2017), it is usually structured around three main core policy principles, namely standards, decentralisation and accountability. The *standardisation* of education involves the definition of learning standards and the prescription of a national curriculum that is structured around measurable common core standards – at least in relation to key curricular subjects.

The *decentralisation* of education implies a shift “in territorial competences from the central government to regional, state or local authorities” (Weiler 1990, 434). In certain contexts, education decentralisation also implies the devolution of organisational and/or pedagogic responsibilities to schools (e.g., school autonomy and school-based management models) and the introduction of managerial leadership models that aim to make schools “more businesslike and more like businesses” (Ball, Junemann, and Santori

2017, 3). Nevertheless, according to Weiler (1990), administrative decentralisation and school autonomy do not go always hand in hand, and some countries can combine, for instance, a high degree of education decentralisation and a low degree of school autonomy.

The third component of the GERM, *accountability*, implies making educational actors (including schools and teachers) more responsive to and responsible for their actions and results. The assumption behind performative-accountability systems is that the linking of learning achievement to incentives, sanctions and/or reputational implications will lead to the improvement of teaching and learning at both the classroom and the school levels (Sahlberg 2016).

National large-scale assessments are highly functional to the practical deployment of the policy principles that give form to the GERM. First, these assessments represent the main policy instrument in the hands of governments to ensure that school actors meet national targets and adhere to the teaching and learning standards defined in the national curriculum. Secondly, in a context of increasing decentralisation of education systems, NLSAs are a key policy technology to retain the State's power and to “steer at a distance” the complex network of providers and actors that configure the school system (Osborne and Gaebler 1992). Thirdly, NLSAs are at the centre of consequential accountability dynamics and systems. The predominant form of accountability that is spreading in the context of the GERM has a performance-based nature. This accountability approach focuses on students’ learning outcomes and involves the generation of data through NLSAs, which is why it is usually known as test-based accountability, or TBA (Hamilton, Stecher, and Klein 2002). Specifically, NLSAs promote accountability relationships of both an administrative and a market nature. In administrative accountability systems, the results obtained by schools and teachers in NLSAs are attached to consequences and

incentives, which are usually distributed by the educational administration. NLSAs operate also as a device of market accountability when test results are published and used to inform families' school choices. School scores, league tables and rankings are policy tools that directly derive from NLSAs and whose intention is making schools more aware of and responsive to school demand dynamics and competitive pressures (Olmedo and Wilkins 2017).

Table 1. The role of national assessments within the GERM

Table 1 around here

Source: Authors.

The international adoption of NLSAs and test-based accountability

Methodological approach and conceptual considerations

One of the main objectives of this paper is to advance towards a better understanding of the spread of NLSAs and TBA as core components of the GERM. To this purpose, in this section we document and examine two different (although interlinked) phenomena: First, the rise and expansion of NLSAs and, secondly, the changing purposes these assessments serve - with a focus on whether they are being more frequently used for accountability purposes.

The study relies on different data sources: a) Comparative reports from international or regional organisations that provide an overview of existing NLSAs in place (see references with an asterisk in the reference list);² b) self-reported data retrieved

² This information has been in turn complemented by background papers behind the referenced reports. This is especially the case of Country Reports prepared in the context of the OECD's *Review on Evaluation and Assessment Frameworks for Improving School Outcomes* (2013a).

from national archives and repositories, usually stored in the webpages of ministries of education, evaluation institutes and similar public agencies; c) academic and media sources, to provide a more complete understanding of the characteristics and purposes of assessments, especially where there is incomplete administrative data on the characteristics of NLSAs; and d) data from the principals' questionnaire in PISA (2003 2006, 2012 and 2015 reports) on how schools use NLSAs' data.

For systematisation purposes, the data collected from these sources has been compiled and organized into a new and original database on NLSAs covering 34 countries or autonomous administrative divisions³. This systematisation exercise provides an historical overview of NLSAs, especially as concerns the purposes, coverage and scope, tested subjects, targeted grades, and main unit of analysis of these instruments.⁴ The universe of countries considered in our database are limited to OECD member states⁵ – given the very limited availability of data on assessment practices in relation to low- and middle-income countries,⁶ but also due to the fact that we find it especially relevant to study the penetration of the GERM within the OECD policy space. The timeframe of our analysis is from 1995 to 2014⁷.

Our research focuses on NLSAs - i.e. those standardised tests designed and administered externally to schools (usually, by the educational administration or external

³ This is the case of the Flemish-speaking community and the Fédération Wallonie-Bruxelles in Belgium, and the case of England, Wales, Scotland and Northern Ireland in the United Kingdom.

⁴ The specific *uses* of national assessments have been explored with data coming from the PISA schools' questionnaire, given the limits of available information on this particular topic in administrative sources.

⁵ Figures 6 to 10 include also non-OECD countries or economies participating in PISA, as they draw on data from principal questionnaires administered as part of the program.

⁶ The 2007 and 2015 editions of the Global Education Monitoring Report constitute an exception to this.

⁷ 1995 was selected as the starting point given the difficulties in collecting comparable and reliable data for prior years. Nonetheless, as reported in Eurydice (2009), national testing did not become widespread until the 1990s.

evaluation agencies), in order to ensure consistency and comparability (Popham 1991, as cited in OECD 2013a; see also Eurydice 2009). We give also consideration to so-called hybrid forms of assessment, i.e. those exams developed externally, but implemented internally and/or administered at a local level, since these are assessments also suitable to system-level comparative purposes⁸.

When it comes to the *uses/purposes* of the assessments, a series of conceptual refinements should be made at this point. Most of the literature on national assessments distinguishes between formative and summative assessments, however, in our database, we classify NLSAs according to a broader range of categories. These include:

- Assessment for student's certification, streaming and selection purposes: These are standardised examinations that are high stakes for students, but not necessarily for schools. This category would include exit examinations or any kind of standardised assessment that certifies learning progress and/or allows students access to higher education levels or have a formal consequence for students' career (decision on streaming, progression from a school year to the next, etc.).
- Assessment for diagnostic purposes: These are standardised assessments that do not have formal implications either for students or for schools. They have a more formative character and aim at providing students and schools with useful data and feedback so that they can improve their future performance. They are also useful to identify students' learning needs and adapt teaching accordingly.
- Assessment for monitoring and evaluation purposes: These are NLSAs used for monitoring and evaluation purposes (measuring progress and performance against different goals, and at different levels of the education system). Includes any test

⁸ Conversely, school-internal (i.e. non-standardised, non-external) assessments are not considered in our database.

oriented at comparing performance between jurisdictions, schools or providers, evaluating the system or a particular program, or creating the necessary data for accountability purposes.

These categories are not mutually exclusive as many assessment programs are designed to serve multiple purposes. It should also be noted that assessments are categorised according to the uses formally and explicitly reported by responsible authorities, and as they are described in comparative reports. In other words, informal, unintended or unofficial uses given to national assessments are not considered in this analysis. Also, our NLSA database does not capture the intensity of different testing programs – the length and number of tests established by each program.

Finally, an additional attribute we take into consideration in order to understand the uses of NLSAs is their scope – i.e., the distinction between census- and sample-based assessments, administered respectively to the whole universe of schools or to a representative sample. For accountability purposes, census-based assessments are the most appropriate, since they allow performative pressure to be placed in all the schools that make up the educational system.

In the next sections, we explore how NLSAs in OECD countries have evolved in both quantitative and qualitative terms, and show that there is an international trend in using NLSAs for accountability purposes.

The increase and changing nature of NLSAs

The number of countries conducting at least one NLSA has increased significantly since the mid-1990s in the OECD education policy space. Both for primary and lower secondary education, this number has nearly doubled since 1995 (see Figure 1). Nonetheless, the total number of assessment programs in place has grown at a greater

pace than the number of countries adopting NLSAs, which means that more and more countries are administering more than one national assessment program.

Figure 1. Number of countries conducting at least one assessment and total number of assessments

Figure 1a about here

*** Figure 1b about here***

ISCED 1

ISCED 2

Source: Authors

Note: In this and the following figures, ISCED 1 and ISCED 2 correspond respectively to primary education (or first stage of basic education) and to lower secondary education (or second stage of basic education, which usually goes until age 15/16).

While the number of countries with at least one assessment program appears to be plateauing out, this is not the case of the total number of assessments being enacted. National large-scale assessments appear to be in a stage of exponential growth or what the epidemiological model of global dissemination denominates an *explosive phase* (Watts 2003, cited in Steiner-Khamsi 2016). According to this model, the explosive phase is usually followed by a burn-out phase, in which the policy adoption trend slows down. However, this slow-down phase has not been reached in relation to the studied case yet, and the number of NLSAs being adopted in the OECD policy field is likely to continue increasing.

When we look at the specific aims pursued by NLSAs, we observe that their growth is, to a large extent, explained by the rise of a particular variety of assessments, those with monitoring and evaluation purposes (see Figure 2). This is the case for both primary and lower secondary education – although, in primary education, it is also possible to discern a substantive increase in assessments with diagnostic purposes. These changes are indicative of a very significant transformation of the nature of assessments.

Hence, testing appears to be increasingly oriented toward measuring achievement with purposes other than making decisions on students' career, among which school-level accountability uses are likely to feature prominently.

Figure 2. Number of assessments conducted per year – distribution by purpose

Figure 2a about here

Figure 2b about here

ISCED 1

ISCED 2

Source: Authors.

In fact, the particular modalities of assessment that have experienced the greatest growth are precisely those typically associated with accountability goals – namely, assessment programs adopting a census-based approach. As shown in Figure 3, among programs with monitoring, evaluation and diagnostic purposes⁹, there has been a particularly noticeable expansion of census based or full-cohort testing programs. Conversely, sample-based assessment programs have experienced a much more moderate growth. Again, this trend applies both to primary and to lower-secondary education.

Figure 3. Number of assessments conducted per year (evaluation and diagnostic purposes) – distribution by scope

Figure 3a about here

Figure 3b about here

ISCED 1

ISCED 2

Source: Authors.

⁹ The census- vs. sample-based distinction do not apply to tests with certification and selection purposes since these are not designed for survey purposes. The “coverage” effectively achieved by this modality of testing is essentially a function of its compulsory or optional character.

Data on the spread of NLSAs in relation to their main unit of analysis points to a similar direction. As shown by Figure 4, while the absolute number of assessments with a focus on the education system as a whole has increased only moderately, this is not the case of assessments focusing on both the whole education system *and* the school level. The latter variety of assessments, which aims at aggregating data at both the school and the system level, has experienced a more pronounced increase since the late 1990s.

Figure 4. Number of assessments conducted per year (evaluation purposes) – distribution by main unit of analysis

Figure 4a about here

Figure 4b about here

ISCED 1

ISCED 2

Source: Authors

Finally, the curricular areas (or subjects) evaluated by NLSAs have also undergone a significant transformation. As shown in Figure 5, assessments introduced during the last decade tend to focus on a narrow set of so-called basic skills and subject - especially numeracy and literacy. To put it differently, new assessments are less likely to assess subjects other than the language of instruction, mathematics and, to a lesser extent, science and ICT. Such a trend is essentially the consequence of a rise in the share of census-based NLSAs. Given the need to keep the resources and time invested in testing at a manageable level, and to ensure that collected data can be adequately compared, assessments adopting a census approach are more likely to focus on a limited number of areas and, especially, on numeracy and literacy-related subjects. Focusing on this narrow group of subjects is also a way to align NLSAs with what is being predominantly measured in international large-scale assessments such as PISA (see Meyer and Benavot 2013).

Figure 5. Targeted subjects - total number of assessments per year*

Figure 5a about here

Figure 5b about here

ISCED 1

ISCED 2

Source: Authors

*Note the overlapping between blue and green lines (literacy and numeracy skills, respectively)

Ultimately, the growing prevalence of specific modalities of NLSAs, as described so far, is indicative of a change in the *potential* uses of assessments: from student certification to the monitoring of schools and teachers. Large-scale assessments with monitoring purposes may serve a multiplicity of purposes such as informing policy formulation and implementation, identifying or leveraging political priorities, or holding schools and teachers accountable (Tobin et al. 2015). Our NLSAs database does not include specific information on accountability uses, but shows that current NLSAs create the necessary conditions for an eventual advancement of the accountability agenda. In the next section, we use OECD/PISA data to more specifically find out whether the accountability uses of NLSAs are intensifying in OECD countries and to what extent.

The increasing accountability purposes of NLSAs

According to Bovens (2007) and Hatch (2013), the external evaluation of public services and professional practices is a necessary condition, but not sufficient, for enhancing accountability systems. For accountability to happen, some “consequences” need to derive from the evaluation results. In the case of education, these consequences include – but are not limited to - using test data to closely supervise teachers and/or schools, to decide on teachers’ salaries and promotion, to encourage school competition via the

publication of test results, or to intervene the autonomy of underperforming schools. The fact that more and more schools and educational actors perceive performative pressure can also be interpreted as an increase of accountability relationships within educational systems. In this section, in the basis of data coming from several PISA editions (2003, 2006, 2012 and 2015), we examine the accountability purposes and practices associated to NLSAs.

In the 2003 and 2012 editions of PISA, school principals were asked whether NLSAs data is used to monitor teachers' work. As Figure 6 shows, in 2003, seventeen countries (out of a total of thirty-six countries that participated in both PISA editions and have available data for this question) had less than 60% of their students attending schools that use the NLSAs data to monitor teachers' practices, whereas, in 2012, there were only three countries below 60% (Finland, Uruguay and Greece). In fact, in 2012, these practices would have increased more than 10 percentage points with respect to year 2003 in twenty-three OECD countries. On the contrary, the only two countries that show a substantial reduction concerning the percentage of students attending schools where assessments are used to monitor teachers' work are Finland and Latvia (OECD 2013b). Overall, our data shows that NLSAs are becoming increasingly central in the evaluation of teachers' practices. Smith and Kubacka (2017) observe a similar trend and add to this argument that NLSAs are also undermining the presence of other forms of teachers' evaluation (such as self-assessment or teacher observations), and of other sources of data in the feedback that teachers receive.

Figure 6. Change between 2003 and 2012 in using student assessment data to monitor teachers

Figure 6 about here

Source: Adapted from OECD 2004, 2013b

Note: Blue columns represent data from PISA 2012; red dots represent data from PISA 2003.

The percentage of students attending schools that use NLSAs' data to compare their performance against the district or against national benchmarks has increased significantly as well. As shown in Figure 7, the OECD average on this item -which somehow indicates schools' awareness of performative standards and pressures - has raised 22 percentage points between PISA 2003 and PISA 2015. The number of countries where more than 80% of students attend schools that use student assessment data for this purpose has gone from 4 countries (in year 2003) to 15 (in year 2015). The countries with a higher percentage of schools using NLSA data to compare school to district or national performance are the United States, Russia, New Zealand, Turkey, Latvia, Sweden, Thailand, the Netherlands, United Kingdom, Portugal, Poland, Brazil, Canada, Iceland, Indonesia, Ireland, Luxembourg, Mexico, and Italy. In 2015, only Greece had less than 20% of students attending schools using NLSAs to compare school performance.

Figure 7. Change between 2003 and 2015 in using student assessment data to compare the school to district or national performance

Figure 7 about here

Source: Adapted from OECD 2004, 2016

Note: Blue columns represent data from PISA 2015; red dots represent data from PISA 2003.

The increase of competitive pressures for schools derived from the administration of NLSAs is also evident in Figure 8, which shows how national assessments are more and more used to compare the school performance with the performance of other schools. In other words, NLSAs are contributing to raising competitive awareness and pressures within local education markets. The OECD average, which has increased around 20 percentage points from 2003 to 2015, also reflects this global trend. In addition, the

number of countries with more than the 80% of students attending schools where the principals use NLSAs to compare their school with other schools has gone from 1 (in 2003) to 9 (in 2015) countries. By contrast, according to the 2015 PISA edition, there are only two countries with less than the 20% of students enrolled in schools that apply this kind of practices: Austria and Greece.

Figure 8. Change between 2003 and 2015 in using student assessment data to compare the school with other schools

Figure 8 about here

Source: Adapted from OECD 2004, 2016

Note: Blue columns represent data from PISA 2012; red dots represent data from PISA 2003.

Between years 2006 and 2015, accountability pressures among schools in OECD countries have slightly increased, as measured by the publication of school results in NLSAs. In relation to this indicator, the OECD average rose by 6.2 percentage points between 2006 and 2015. Figure 9 shows that the number of countries with more than 60% of the students attending schools in which the achievement data is posted publicly went from 7 (in 2006) to 10 (in 2015). The publication of schools' achievement data is a controversial variable when designing test-based accountability systems. While scholars such as Boarini and Lüdemann (2009) suggest that the public release of school results can contribute to improving academic performance, other scholars find that the publication of school results – especially when adopts the form of rankings – contribute to stigmatising schools, or to undermining cooperation between schools (Jones and Egley, 2004). Nevertheless, the fact that the publication of schools' results translate, or not, into market accountability pressure depends on the configuration of the education system in each country or region. Thus, while in countries with an absence of school choice (e.g. Norway) these type of measures can only trigger non-competitive or, at most, reputational

pressures for school improvement, in countries where the education system combines freedom of school choice with demand-side funding schemes (e.g. Chile), the publication of schools' results have more direct effects in increasing market accountability pressures.

Figure 9. Change between 2006 and 2015 in using achievement data in the following ways: Achievement data are posted publicly (e.g. in the media)

Figure 9 about here

Source: Adapted from OECD 2007, 2016

Note: Blue columns represent data from PISA 2015; red dots represent data from PISA 2006.

Finally, NLSAs results can also be used for strengthening administrative accountability dynamics. In Figure 10, we can observe that between 2006 and 2015 the OECD average regarding the percentage of students enrolled in schools where achievement data are used for administrative accountability purposes has increased 6 percentage points. In 2015, there were 14 OECD countries with 80% or more of their students attending schools where achievement data is tracked over time. In contrast, Japan is the only OECD country where less than 20% of students attend schools where achievement data is tracked by an administrative authority for accountability purposes.

Figure 10. Change between 2006 and 2015 in using achievement data in the following ways: Achievement data are tracked over time by an administrative authority

Figure 10 about here

Source: Adapted from OECD 2007, 2016

Note: Blue columns represent data from PISA 2015; red dots represent data from PISA 2006.

To conclude, PISA data indicates that schools' performance in NLSAs is becoming more and more central for governing school systems and school actors' perceptions and behaviour. Nevertheless, on the basis of this data, we cannot directly infer that there is an

incremental rise in the stakes associated to schools' results in NLSAs, and in particular in the rise of the so-called high-stakes. The evidence presented here indicates that many countries are not only implementing NLSAs for educational diagnosis, but also using them more frequently to intensify administrative, performative and/or market pressures on schools.

Understanding the international spread of NLSAs and TBA

National large-scale assessments are expanding internationally as part of a global education reform movement that promotes standard-based curricula, decentralisation and accountability in educational systems. However, the GERM is more than an *explanans*, it is something that needs to be explained (or an *explanandum*). Predominant theories of policy diffusion and globalisation (namely, rationalism, neo-institutionalism and international political economy) can contribute to capture what macro policy mechanisms are behind the international dissemination of the GERM and, in particular, the growth and spread of both NLSAs and TBA.

According to a rationalistic approach to policy transfer, the policies enjoying widespread international circulation are also those that work best and are more effective. This approach assumes that decision-makers are primarily guided by functional efficiency concerns, so that they are prone to adopt those policies and programs whose effectiveness is supported by other countries' experiences and by empirical evidence. NLSAs and the accountability measures attached to them are perceived by many as an effective way to promote and assure quality in education. However, just as happens with other models of educational reform that have been globalised, the evidence on the positive effects of accountability policies based on NLSAs is, at least, inconclusive. To date, empirical research on accountability and education has obtained very different –and even

contradictory— results regarding the effects of TBA on increasing levels of learning, improving educational processes, or reducing educational inequalities. Also, more and more studies show that certain test-based accountability systems can generate undesired or opportunistic behaviours at the school level that end up undermining the quality of teaching and/or the inclusiveness of schools (Allan and Artiles 2016; Au 2007). Therefore, at least from a theoretical point of view, rationalism would be insufficient to understand the globalisation of test-based accountability in education.

Neo-institutionalism challenges the basic assumptions of rationalism, arguing that legitimacy is a more important driver of policy adoption than functional efficiency (Dobbins and Knill 2009). National large-scale assessments and TBA are seen as legitimate policy instruments due to the fact that are intrinsically linked to the modernisation of public services and to the promotion of transparency and meritocracy in public education.¹⁰ Furthermore, the development and growing presence of international large-scale assessments such as PISA, Trends in International Mathematics and Science Study (TIMSS) or Progress in International Reading Literacy Study (PIRLS) have also encouraged many governments to adopt NLSAs. This is due to the fact that enacting NLSAs is a way for countries to more appropriately engage with the global education policy debates and the ‘educational race’ that international large-scale assessments generate (Meyer and Benavot 2013). Overall, according to neo-institutionalism, national assessments would be such attractive policy instruments because they confer governments with legitimacy in international fora, as well as in front of their own population.

¹⁰ From the perspective of meritocracy, national tests would be instrumental in identifying and rewarding talent, as well as in motivating students to work more diligently. (Hutt and Schneider, 2018).

In contrast, international political economy approaches place more emphasis on factors of a material and economic nature. From this perspective, processes of policy dissemination are related to prevailing economic transformations and pressures, as well as to countries' ambitions of positioning themselves better in an increasingly competitive global economy. In such an economic environment, the adoption of NLSAs and TBA would permit the achievement of a threefold economic objective: firstly, to improve student's learning outcomes, which many see as synonym of promoting countries' economic growth, knowledge-based economies and attracting foreign investment; secondly, to promote competition between school providers and cost-efficiency gains in public education services; and thirdly, to favour the opening of the educational sector to the commercial interests of an emerging educational testing and school improvement industry (Ball 2008; Carnoy 2016; Verger and Parcerisa 2018).

A political sociology approach to policy instruments

The macro approaches to policy transfer sketched above reflect on the global mechanisms (namely international evidence, legitimation and economic competitiveness) that are behind the international spread and growth of NLSAs and TBA. In this section, we reflect on how a political sociology approach to policy instruments can complement these macro theories by bringing meso-level factors into the analysis. The political sociology approach to policy instruments aims to understand, on the one hand "how instruments are chosen, how they develop and how they are operationalized" (Kassim and Le Galès 2010, 2) and, on the other, what are the political implications of policy instruments' choices (Peters 2002). This approach is contextually-grounded, in the sense that is compatible with historical institutional premises on the role of institutions in the mediation of global forces and agendas, but also in the sense that provides actors operating at different scales with

voice and agency in understanding policy adoption. The political sociology approach to policy instruments, due to its embeddedness within a constructivist epistemology, emphasises that meaning-making processes importantly interact with political, institutional and economic factors in the production of policies and policy change.

From this perspective, some of the factors that could explain the spread and growth of NLSAs and TBA are the politically-rewarding, economically-convenient and semiotically-malleable nature of these policy instruments.

- *Politically rewarding.* Enacting quality assurance and accountability systems in education allows politicians to signal to their publics that they are working seriously toward education change and that they are concerned about education quality, learning outcomes and the future of children, while simultaneously putting the reform pressure on schools and teachers rather than on the government. Furthermore, for political actors, it is easier to discuss and agree on the adoption of policy instruments than on deeper or more fundamental policy transformations (Kassim and Le Galès 2010). Overall, adopting new instruments might be a way to avoid addressing controversial political debates, or a way of addressing complex debates (such as the causes of and solutions to educational inequalities) from the perspective of technical and instrumental choices. For instance, it is quite common that, in current policy narratives, external assessments are portrayed as key devices to address *learning gaps* and assure that *all kids* achieve minimum learning standards.
- *Economically convenient.* Policy instruments are easier to adopt if they are administratively and economically viable or, at least, their cost is not apparent to the public (Peters 2002). Accountability reforms are somehow “cheaper and quicker than alternative reforms” (Smith et al. 2004, 50), and in particular than

the adoption of more profound and structural equity reforms. Something similar could be said about the adoption of other components of the GERM, such as curricular standards, which according to Hargreaves and Shirley expand like “wildfire” because are “easy to write, inexpensive to fund” (2009, 9; as quoted in Sahlberg 2016, 131).

- *Semiotically malleable.* Policy instruments spread more quickly if they can be attached to positive symbols (such as “quality assurance”) and politically powerful slogans (such as “addressing the learning crisis” or the “learning gap”). Policy instruments also expand faster when they have a polysemic nature and/or can operate as empty signifiers, i.e. the same instrument can serve very different – even contradicting – goals (Steiner-Khamsi 2016). This is the case of instruments such as NLSAs and TBA, which allow different political groups to advance different objectives through them. For instance, the fact that the same accountability instrument might contribute to meet a diversity of goals such as quality, equity, transparency, school choice, etc. facilitates that governments with different ideological orientations converge in conceiving test-based accountability as an adequate instrument through which to articulate educational change.

The selection of policy instruments is contingent to political, discursive and economic factors, as those just mentioned. Nonetheless, policy instruments are far from being technical and neutral technologies that political actors manipulate at their will (Lascoumes and Le Galès 2007). Despite their socially constructed character, or maybe because of this, policy instruments, once adopted, have the potential of affecting the politics, semiotics and economics of education policy. Some specific premises that can be tested regarding the constitutive role of policy instruments are, *first*, “policy

instruments have an existence independent of the decisions that created them” (Kassim and Le Galès 2010, 11). It is difficult to predict the form that any instrument will assume since policy “instruments create their own structures of opportunity in ways that were typically unforeseen when they were adopted” (Kassim and Le Galès 2010, 12), and generate political activity and political consequences that go beyond the initial instrument ontology (Peters 2002).

Without going any further, our data implies that NLSAs tend to be increasingly used for accountability and monitoring purposes, even in places where they were not initially adopted with these objectives. One hypothesis here is that, NLSAs, at the time of adoption, are not seen as intrusive instruments by school actors due to their education diagnostic and/or student grading purposes; however, once the complex large-scale assessment infrastructure is in place, accountability policies start being adopted and are increasingly attached to the assessment results. This is for example the case of the end-of-primary test in the Netherlands. While the test was originally aimed at providing information on the secondary-school type most appropriate for each student, new uses have been attached to the test over recent years. More specifically, the test has become an instrument for school assessment purposes, as aggregate results are now used by inspectors as indicators of school quality (Nusche et al. 2014; OECD 2013a; Scheerens et al. 2012). There are also actors that, beyond national governments, might start doing accountability policy with NLSAs. In Denmark and Norway, the government stopped producing rankings with school scores due to teachers organisations’ and other educational stakeholders’ contestation. However, the media and some local governments, taking advantage of transparency rules in public administration, have access to NLSA data and produce and disseminate their own school rankings (Hatch 2013; Ydesen and Andreasen 2014).

Secondly, once selected, policy instruments privilege certain actors and their interests by determining resource allocations, access to the policy process, and problem representations (Menon and Sedelmeier 2010). Here, it is important to observe which educational actors are empowered by TBA systems, and which are disempowered, as well as whether these systems contribute to the emergence of new political subjectivities such as the *opt-out* movement, which is increasingly influential in several US localities (Pizmony-Levy and Saraisky 2016; Hutt and Schneider 2018). Economic interests also emerge around new policy instruments, as both third-party producers (for instance, testing companies in charge of NLSAs) and parties supporting policy implementation (for instance, consultants offering data analysis, instructional improvement or test preparation services to schools aiming to improve their scores). When these economic interests are strong, there are more reasons to expect that policy instruments will endure in time, independently of how effective the instruments are.

Finally, the effects produced by policy instruments “depend on how the aims and purposes ascribed to them, and the meanings and representations they carry, are perceived, understood and responded to by key actors” (Skedsmo 2010, 7). Overall, a political sociology approach to policy instruments aims to go beyond the study of the impact of policies (in terms of, for instance, educational access, learning achievement, and so on), and rather aims to analyse the types of social relations, behaviours and responses that instruments trigger among communities of practice (Lascoumes and Le Galès 2007). Policy instruments are forms of power that aim at disciplining actors’ behaviour, but they are also resisted and/or creatively transformed by educational actors. Thus, to understand how NLSAs evolve, we need to go beyond the level of formal policy formulation and observe how practitioners also *do policy* (cf Ball, Maguire, and Braun,

2012) by enacting these instruments and the implicit educational mandates that come with them, as well as whether they do so in unpredicted or unintended ways.

Conclusions

In the “governing by numbers” era (cf. Grek 2009; Ozga 2009), NLSAs have become a fundamental instrument in the promotion of an educational reform approach that is structured around three main policy principles, namely standards, accountability and decentralisation. National large-scale assessments are a core component of this reform approach, qualified by many as the GERM. NLSAs generate the necessary data for the state to hold different educational actors accountable for students’ achievement and their adherence to national curricular standards. Both the number of countries adopting NLSAs and the number of NLSAs being administered in the OECD policy space have grown substantially in the last two decades.

Furthermore, the monitoring and accountability uses of these assessments have intensified as well, even in countries from continental Europe that, until recently, seemed to be immune to the GERM. Our database shows that the dissemination of NLSAs opens a window of opportunity for the adoption of TBA systems (and other core components of the GERM), despite it does not specify the final uses attached to these policy instruments by countries. In fact, schools’ performance data generated by NLSAs can both be used as a formative evaluation tool (i.e. as a tool supportive of instructional improvement), or as a device to trigger competition between schools and to control and sanction school actors. A more precise analysis of specific NLSA uses internationally would require of having access to new sources of secondary data or, due to their constantly changing nature, the conduction of qualitative historical case studies that scrutinize the evolving policy design of NLSA/TBA configurations, and their re-contextualisation at the school level.

Predominant theoretical approaches to policy transfer (namely rationalism, neo-institutionalism and international political economy) tell us about the macro-mechanisms and global drivers behind the international dissemination of NLSAs and TBA. Nonetheless, a political sociology approach to policy instruments is more appropriate to understand what particular TBA configurations are being retained in educational systems, and how they are transformed, re-signified and implemented by educational actors operating at multiple scales. National large-scale assessments and TBA spread because of changing global education mandates and pressures, but also because they represent a politically rewarding, economically convenient and politically malleable education reform model at the domestic level. This research approach also conceives NLSAs and TBA as autonomous devices that - beyond simple or neutral quality assurance instruments - might contribute to alter education actors' configurations, school behaviours and the goals of educational organisations in ways that were unforeseen by their promoters and enactors.

The political sociology approach to policy instruments fits within a cultural political economy ontology, in the sense that it is interested in observing how semiosis and meaning-making are part of the broader political-economy configuration of educational institutions. It fits also with a multi-scalar approach to education policy, in the sense that conceives that factors and actors operating at multiple scales critically interact in the production of policies: from the settlement of global policy agendas to everyday policy practices. National large-scale assessments, quality assurance mechanisms and TBA have started being analysed from a similar perspective (see Gable and Lingard 2015; Kauko, Rinne and Takala 2018; Skedsmo 2011), but more empirical research is necessary to acquire a more complex understanding of how and why these

policies evolve and expand worldwide, and whether their multiple effects in educational systems endorse such an expansion.

References

- Allan, J., & Artiles, A. J. (Eds.). (2016). *World Yearbook of Education 2017: Assessment Inequalities*. New York: Routledge.
- Au, W. 2007. "High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis." *Educational Researcher* 36 (5): 258–267. doi: 10.3102/0013189x07306523
- Ball, S.J., 2008. *The education debate*. Bristol: The Policy Press.
- Ball, S.J., M. Maguire, and A. Braun. 2012. *How schools do policy. Policy enactments in secondary schools*. Abingdon/New York: Routledge.
- Ball, S. J., C. Junemann, and D. Santori. 2017. *Edu.net. Globalisation and Education Policy Mobility*. Abingdon/New York: Routledge.
- *Benavot, A., and E. Tanner, E. 2007. *The Growth of National Learning Assessments in the World, 1995-2006. Background paper prepared for the 2008 Global Monitoring Report*. Paris: UNESCO-GMR. Retrieved from: <http://unesdoc.unesco.org/images/0015/001555/155507e.pdf>
- Boarini, R., and E. Lüdemann. 2009. "The Role of Teacher Compensation and Selected Accountability Policies on Learning Outcomes: An Empirical Analysis for OECD Countries." *OECD Journal: Economic Studies* 2009. doi: 10.1787/eco_studies-v2009-art10-en
- Bovens, M. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal* 13 (4): 447–468. doi: 10.1111/j.1468-0386.2007.00378.x
- Carnoy, M. (2016), 'Educational policies in the face of globalization: Whither the Nation State?', in K. Mundy, A. Green, R. Lingard, and A. Verger (eds), *Handbook of global policy and policy-making in education*, 27–42, West Sussex: Wiley-Blackwell
- Dale, R. 1999. "Specifying globalisation effects on national policy: A focus on the mechanisms." *Journal of Education Policy* 14 (1): 1–17. doi:10.1080/026809399286468
- Dale, R. 2000. "Globalisation and Education: Demonstrating a 'Common World Educational Culture' or Locating a 'Globally structured educational agenda'?" *Educational Theory* 50 (4): 427–448. doi:10.1111/j.1741-5446.2000.00427.x
- Dobbins, M., and C. Knill. 2009. "Higher Education Policies in Central and Eastern Europe: Convergence toward a Common Model?" *Governance* 22 (3): 397–430. doi: 10.1111/j.1468-0491.2009.01445.x

- * EFA-GMR. 2007. *Strong Foundations. Early Childhood Care and Education. Education for All Global Monitoring Report 2007*. Paris: UNESCO Publishing. Retrieved from: <http://unesdoc.unesco.org/images/0014/001477/147794e.pdf>

- * EFA-GMR. 2015. *Education for All 2000-2015: Achievement and Challenges. Education for All Global Monitoring Report 2007*. Paris: UNESCO Publishing. Retrieved from: <http://unesdoc.unesco.org/images/0023/002322/232205e.pdf>

- * Eurydice – EACEA. 2009. *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Brussels: Education, Audiovisual and Culture Executive Agency - Eurydice. Retrieved from: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf

- Gable, A., and B. Lingard. 2015. “NAPLAN Data: A New Policy Assemblage and Mode of Governance in Australian Schooling.”, *Policy Studies* 37(6): 568–582. doi: 10.1080/01442872.2015.1115830

- Grek, S. 2009. “Governing by Numbers: The PISA ‘effect’ in Europe.” *Journal of education policy* 24 (1): 23–37. doi: 10.1080/02680930802412669

- Gunter, H. M., E. Grimaldi, D. Hall, and R. Serpieri. 2016. “NPM and the dynamics of education policy and practice in Europe.” In *New Public Management and the Reform of Education: European Lessons for Policy and Practice*, edited by H. M. Gunter, E. Grimaldi, D. Hall, and R. Serpieri, 3–17. London: Routledge.

- Hamilton, L. S., B. M. Stecher, S. P. Klein, eds. 2002. *Making Sense of Test-Based Accountability in Education*. Santa Monica/Arlington/Pittsburgh: RAND Corporation. Retrieved from: https://www.rand.org/content/dam/rand/pubs/monograph_reports/2002/MR1554.pdf

- Hargreaves, A., L. Earl, S. Moore, and S. Manning. 2001. *Learning to Change. Teaching Beyond Subjects and Standards*. San Francisco: Jossey-Bass.

- Hatch, T. 2013. “Beneath the Surface of Accountability: Answerability, Responsibility and Capacity-Building in Recent Education Reforms in Norway.” *Journal of Educational Change* 14 (2): 113-138. doi: 10.1007/s10833-012-9206-1

- Hutt, E & Schneider, J. 2018. *A Thin Line Between Love and Hate: Educational Assessment in the United States*. Retrieved from: https://www.holycross.edu/sites/default/files/educ/thin_line.pdf

- Jones, L. V. 1996. “A History of the National Assessment of Educational Progress and Some Questions About Its Future.” *Educational Researcher* 25 (7): 15–22. doi: 10.3102/0013189X025007015

- Jones, B. D., & Egley, R. J. 2004. "Voices from the Frontlines: Teachers' Perceptions of High-Stakes Testing." *Education Policy Analysis Archives* 12 (39). doi: 10.14507/epaa.v12n39.2004
- Kamens, D. H., and C. L. McNeely. 2010. "Globalization and the Growth of International Educational Testing and National Assessment." *Comparative Education Review* 54 (1): 5–25. doi: 10.1086/648471
- Kassim, H., and P. Le Galès. 2010. "Exploring Governance in a Multi-level Polity: A Policy Instruments Approach." *West European Politics* 33 (1): 1–21. doi: 10.1080/01402380903354031
- Kauko, J, R. Rinne and T. Takala. 2018. *Politics of Quality in Education: A Comparative Study on Brazil, China, and Russia*. New York: Routledge.
- Klees, S. J. (2008). A quarter century of neoliberal thinking in education: Misleading analyses and failed policies. *Globalisation, Societies and Education*, 6(4): 311-348.
- Lascoumes, P., and P. Le Galès. 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. doi: 10.1111/j.1468-0491.2007.00342.x
- Menon, A., and U. Sedelmeier. 2010. "Instruments and intentionality: civilian crisis management and enlargement conditionality in EU security policy." *West European Politics* 33(1): 75–92. doi: 10.1080/01402380903354106
- Meyer, H.-D., and A. Benavot. 2013. "PISA and the Globalization of Education Governance: Some Puzzles and Problems". In *PISA, Power, and Policy: The Emergence of Global Educational Governance*, edited by H.-D. Meyer and A. Benavot, 9–26. Oxford: Symposium Books.
- Nusche, D., H., Bran, G. Hálás, and P. Santiago. 2014. *OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014*. Paris: OECD Publishing. Retrieved from: <http://www.oecd.org/education/school/OECD-Evaluation-Assessment-Review-Netherlands.pdf>
- OECD. 2004. *Learning for Tomorrow's World. First Results from PISA 2003*. Paris: OECD. Retrieved from: http://www.oecd.org/education/school/programme_for_international_student_assessment_pisa/34002216.pdf
- OECD. 2007. *PISA 2006. Science Competencies for Tomorrow's World*. Paris. OECD. Retrieved from: http://www.oecd.org/education/school/programme_for_international_student_assessment_pisa/pisa2006results.htm#tables_figures_dbase

- * OECD. 2013a. *Synergies for Better Learning. An international perspective on evaluation and assessment. OECD Reviews of Evaluation and Assessment in Education*. Paris: OECD. Retrieved from: <https://www.oecd-ilibrary.org/docserver/9789264190658-en.pdf?expires=1523571951&id=id&accname=guest&checksum=DBDA81CBDD068F947A077B99A8005765>
- OECD. 2013b. *Results: What Makes a School Successful? (Volume IV) Resources, Policies and Practices*. Paris: OECD. Retrieved from: https://www.oecd-ilibrary.org/education/pisa-2012-results-what-makes-a-school-successful-volume-iv_9789264201156-en
- OECD. 2016. *PISA Results (Volume II): Policies for Successful Schools*. Paris: OECD. Retrieved from: https://www.oecd-ilibrary.org/education/pisa-2015-results-volume-ii_9789264267510-en
- Olmedo, A., and A. Wilkins. 2017. "Governing Through Parents: A Genealogical Enquiry of Education Policy and the Construction of Neoliberal Subjectivities in England." *Discourse: Studies in the Cultural Politics of Education* 38 (4): 573–589. doi: 10.1080/01596306.2015.1130026
- Osborne, D., and T. Gaebler. 1992. *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading: Addison-Wesley.
- Ozga, J. 2009. "Governing Education through Data in England: From Regulation to Self-Evaluation". *Journal of Education Policy* 24 (2): 149–163. doi: 10.1080/02680930902733121
- Peters, B. G. 2002. "The politics of tool choice". In *The tools of government: A guide to the new governance*, edited by L. M. Salamon, 552–564. Oxford: Oxford University Press.
- Pizmony-Levy, O., and N. G. Saraisky. 2016. *Who Opt's Out and Why? Results from a national survey on opting out of standardized tests*. New York: Teachers College - Columbia University. doi: 10.7916/D8K074GW
- Popham, J. 1991. "Why standardized tests don't measure educational quality." *Educational Leadership* 56 (6): 8-15. Retrieved from: <http://www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don%27t-Measure-Educational-Quality.aspx>
- Robertson, S. L. 2012. "Researching Global Education Policy: Angles In/On/Out". In *Global Education Policy and International Development: New Agendas, Issues and Practices*, edited by A. Verger, M. Novelli, and H. Altinyelken, 33–53. London: Continuum.

- Robertson, S. L. 2015. "What teachers need to know about the 'Global Education Reform Movement'". In *Global education 'reform'. Building resistance and solidarity*, edited by G. Little, 10-17, Croydon: Manifesto Press.
- Sahlberg, P. 2005. "Education Reform for Raising Economic Competitiveness." *Journal of Educational Change* 7 (4): 259–287. doi: 10.1007/s10833-005-4884-6
- Sahlberg, P. 2016. "The Global Educational Reform Movement and Its Impact on Schooling." In *The Handbook of Global Education Policy*, edited by K. Mundy, A. Green, B. Lingard, and A. Verger, 128–144. West Sussex: Wiley-Blackwell.
- Scheerens, J., M. Ehren, P. Slegers, and R. de Leeuw. 2012. *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Country Background for the Netherlands*. Paris: OECD: Retrieved from: http://www.oecd.org/education/school/NLD_CBR_Evaluation_and_Assessment.pdf
- Skedsmo, G. 2011. "Formulation and Realisation of Evaluation Policy: Inconsistencies and Problematic Issues." *Educational Assessment, Evaluation and Accountability* 23 (1): 5–20. doi: 10.1007/s11092-010-9110-2
- Smith, W. C., and K. Kubacka. 2017. "The Emphasis of Student Test Scores in Teacher Appraisal Systems." *Education Policy Analysis Archives* 25 (86). doi: 10.14507/epaa.25.2889
- Smith, W. C., ed. 2016. "An Introduction to the Global Testing Culture." In *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice*, edited by W. C. Smith, 7–23. London: Symposium Books.
- Smith, M. L., L. Miller-Kahn, W. Heinecke, and P. F. Jarvis. 2004. *Political Spectacle and the Fate of American Schools*. New York: Routledge.
- Steiner-Khamsi, G. 2016. "New Directions in Policy Borrowing Research." *Asia Pacific Education Review* 17 (3), 381–390. doi: 10.1007/s12564-016-9442-9
- Tobin, M., P. Lietz, D. Nugroho, R. Vivekanandan, and T. Nyamkhuu, 2015. *Using Large-Scale Assessments of Students' Learning to Inform Education Policy. Insights from the Asia-Pacific Region*. Melbourne/Bangkok: ACER/UNESCO. Retrieved from: https://research.acer.edu.au/cgi/viewcontent.cgi?article=1020&context=monitoring_learning
- Verger, A., and L. Parcerisa. (2018). Test-based accountability and the rise of regulatory governance in education: A review of global drivers. In A. Wilkins and A. Olmedo (eds). *Education governance and social theory: Interdisciplinary approaches to research*. London: Bloomsbury.

- Verger, A., M. Novelli, and H. K. Altinyelken. 2018. "Global Education Policy and International Development: A Revisited Introduction." In *Global Education Policy and International Development New Agendas, Issues and Policies. 2nd Edition*, edited by A. Verger, M. Novelli, and H. K. Altinyelken, 1–34. London: Bloomsbury.
- Weiler, H. 1990. "Comparative perspectives on educational decentralization: An exercise in contradiction?" *Educational Evaluation and Policy Analysis* 12 (4), 433-448.
- Ydesen, C., and K. E. Andreasen. 2014. "Accountability practices in the history of Danish primary public education from the 1660s to the present." *Education Policy Analysis Archives*, 22 (120). doi: [10.14507/epaa.v22.1618](https://doi.org/10.14507/epaa.v22.1618)

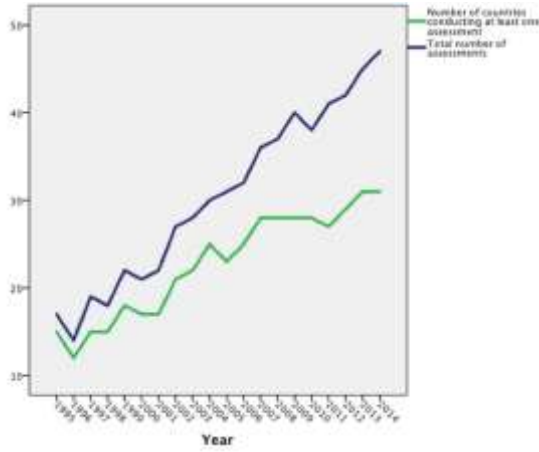
Annex

Table 1. The role of national assessments within the GERM.

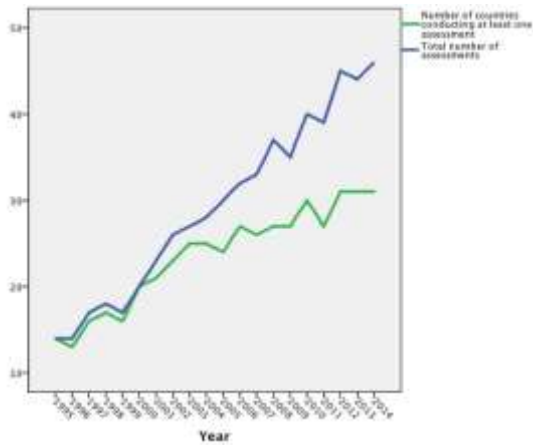
GERM principle	Definition and main policies	Role of national assessments	
Standards	Prescription of a national curriculum and establishment of quality standards	National assessments used to make sure schools meet and adhere to evaluable learning standards	
Decentralisation	Transfer of competences and authority from the central government to lower administrative levels	National assessments used to control state, regional, provincial, and local authorities	
	Devolution of managerial and/or pedagogical responsibilities to principals and schools	National assessments used to govern at a distance a range of autonomous providers through the principles of outcomes-based management	
Accountability	Educational actors made responsible for their actions/results through some form of evaluation linked to consequences.	<i>Administrative accountability</i>	Tests results attached to incentives or sanctions for schools, principals and teachers
		<i>Market accountability</i>	Test results used to inform school choice and promote school competition

Figure 1. Number of countries conducting at least one assessment and total number of assessments

Figure 1a about here



*** Figure 1b about here***



Source: Authors

Figure 2. Number of assessments conducted per year – distribution by purpose

Figure 2a about here

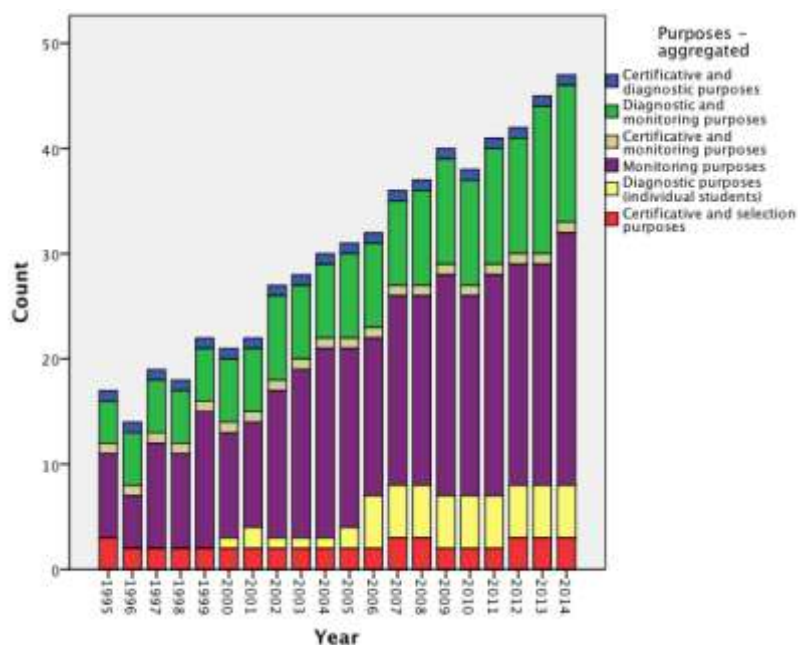
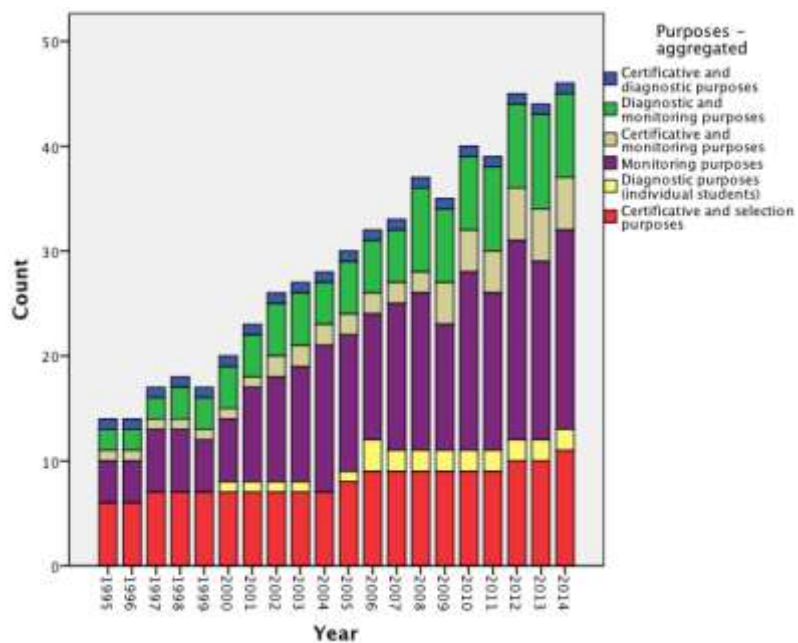


Figure 2b about here



Source: Authors

Figure 3. Number of assessments conducted per year (evaluation and diagnostic purposes) – distribution by scope

Figure 3a about here

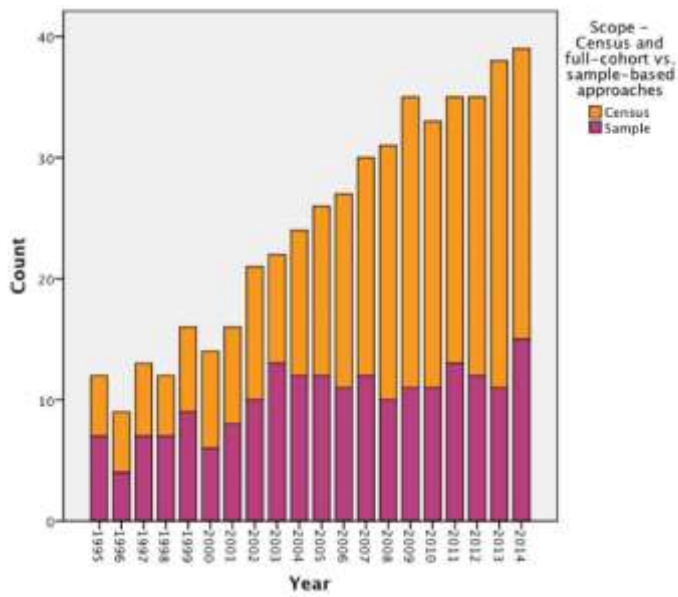
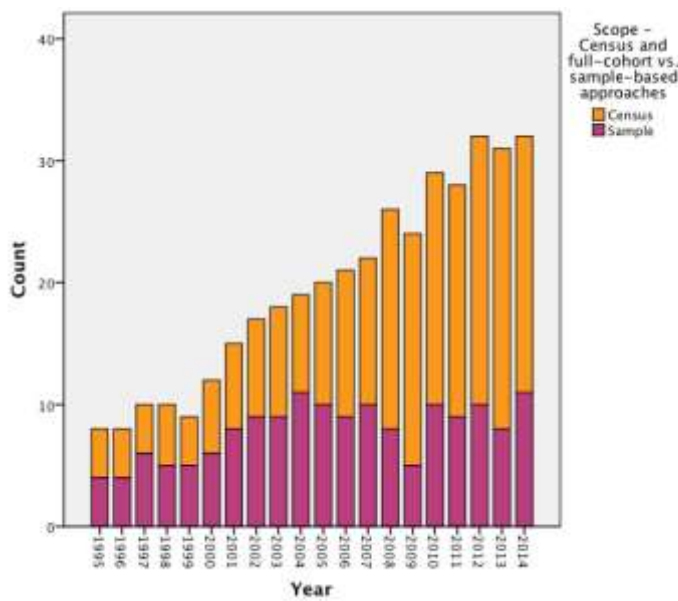


Figure 3b about here



Source: Authors

Figure 4. Number of assessments conducted per year (evaluation purposes) – distribution by main unit of analysis

Figure 4a about here

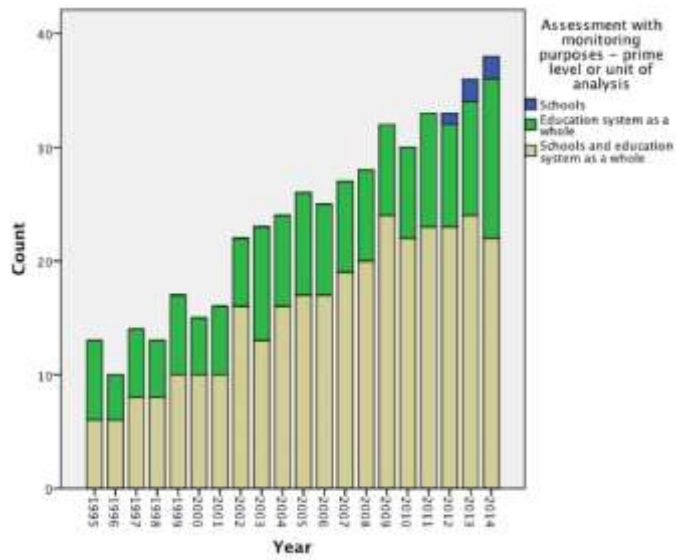
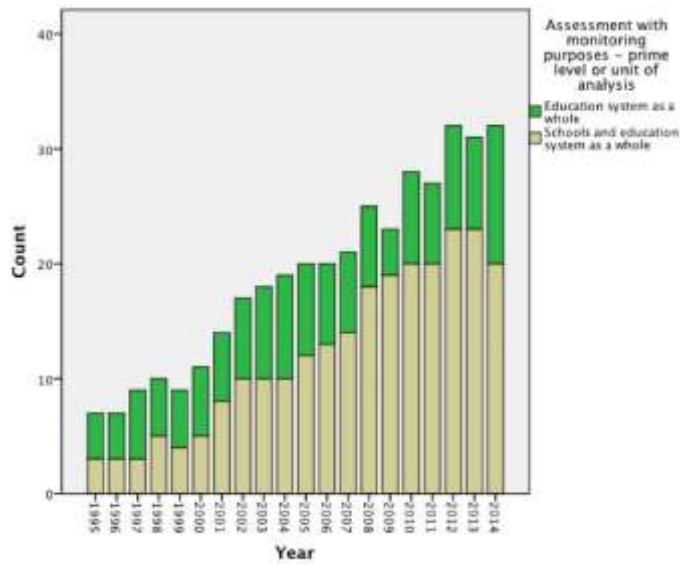


Figure 4b about here



Source: Authors

Figure 5. Targeted subjects - total number of assessments per year*

Figure 5a about here

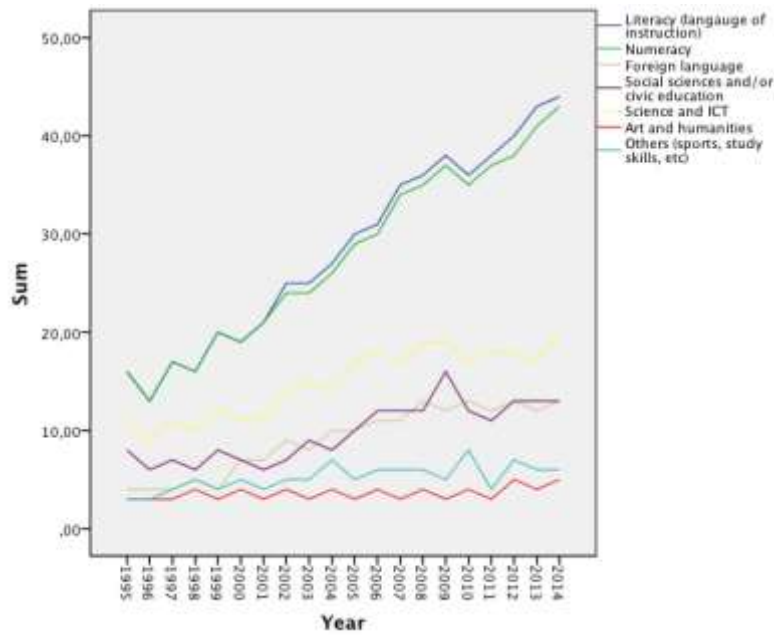
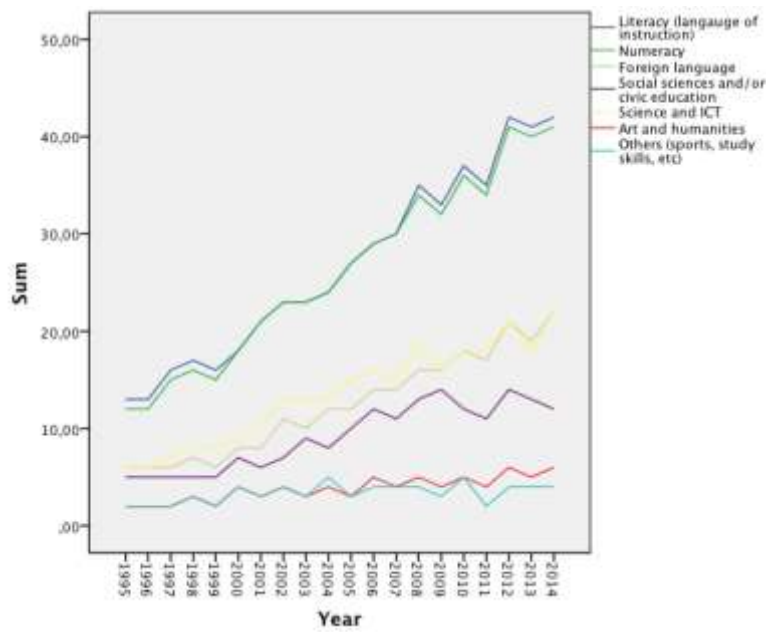
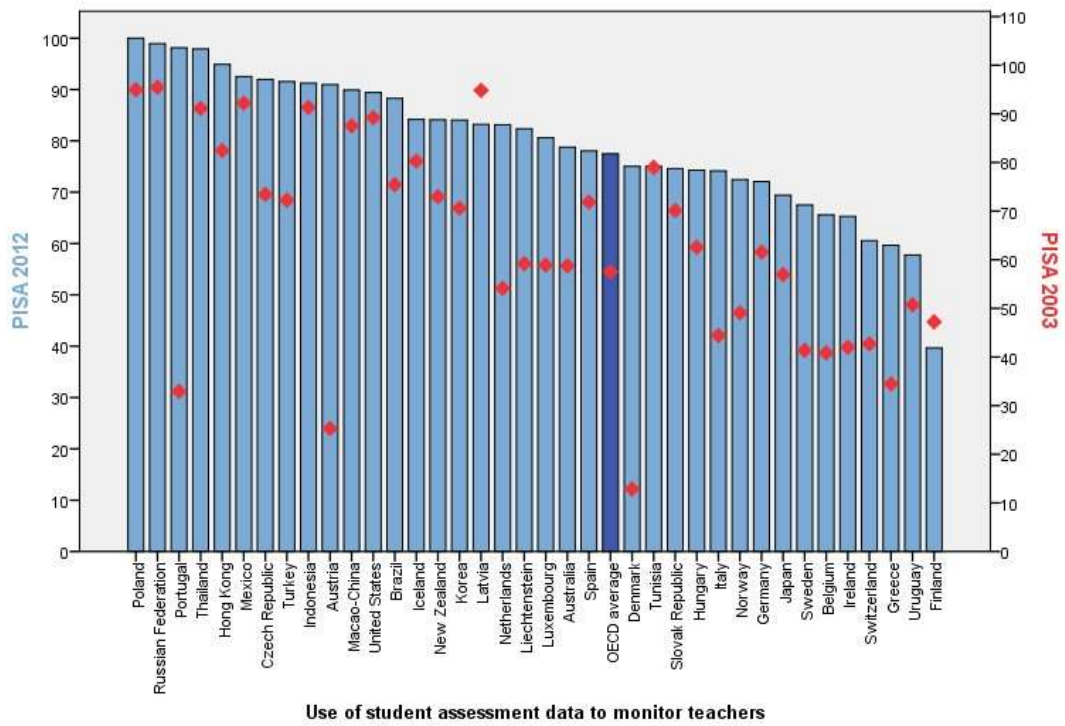


Figure 5b about here



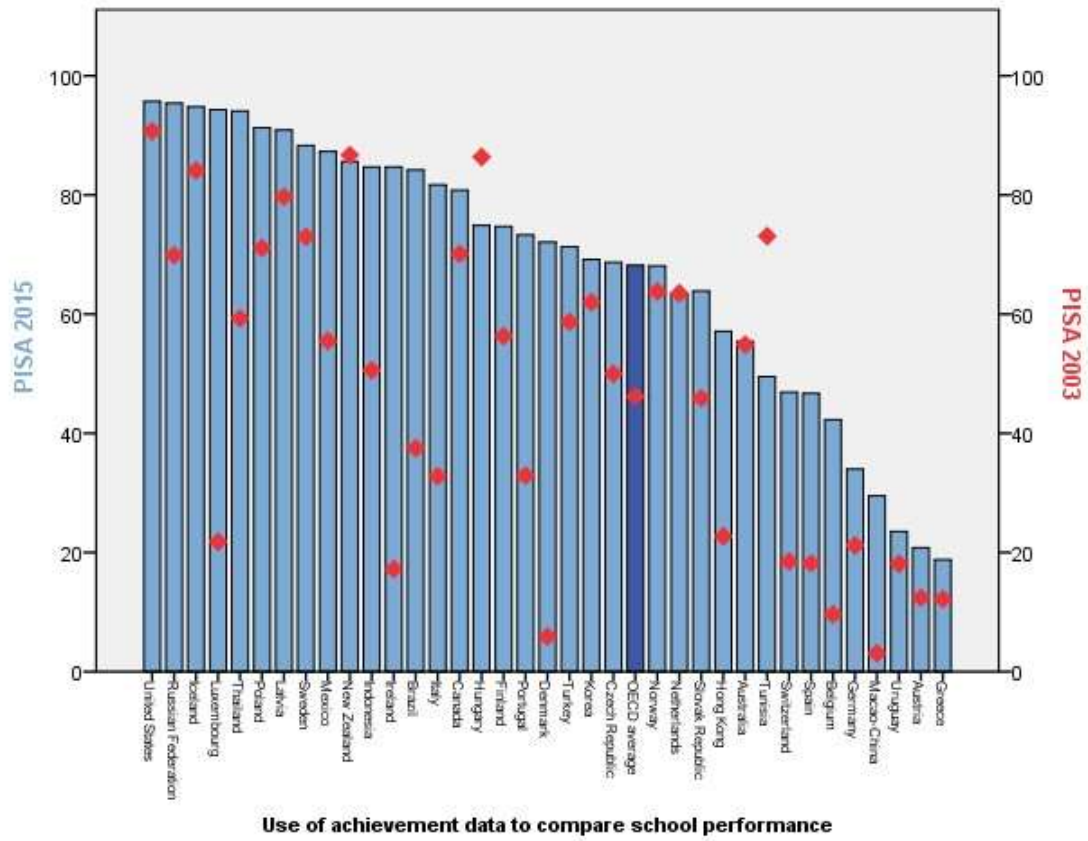
Source: Authors

Figure 6. Change between 2003 and 2012 in using student assessment data to monitor teachers



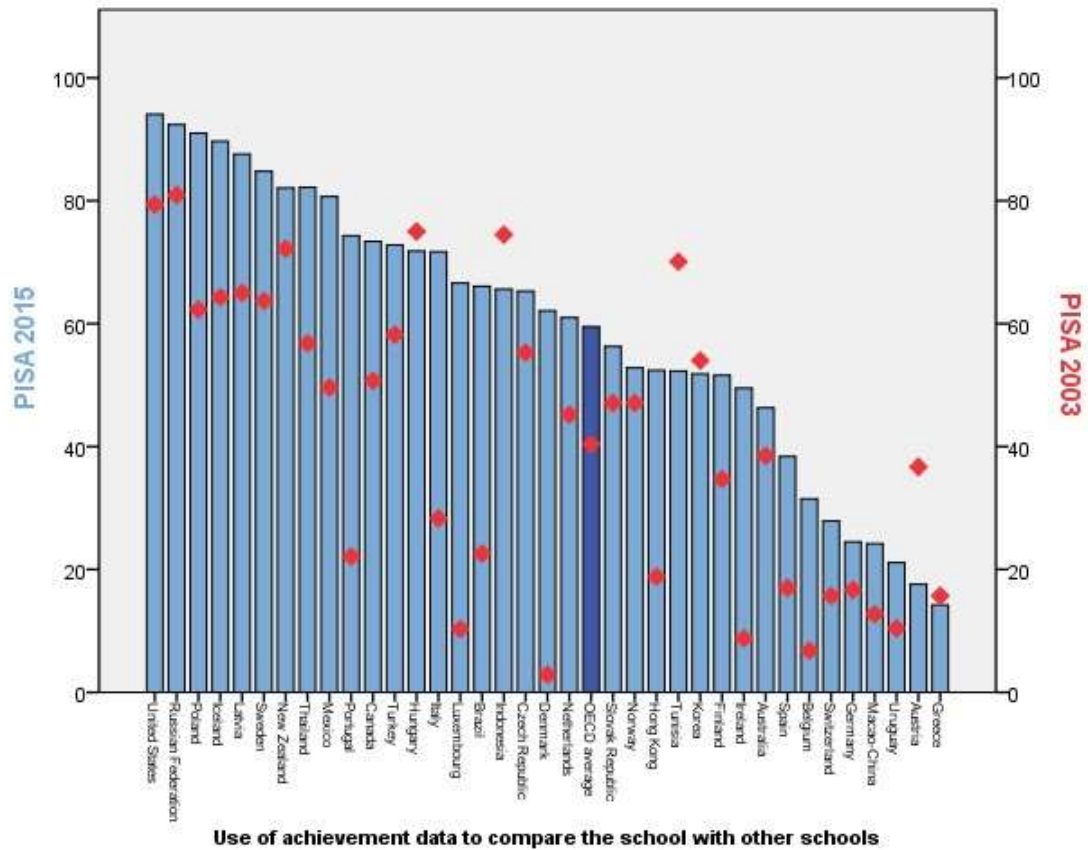
Source: Adapted from OECD 2004, 2013b

Figure 7. Change between 2003 and 2015 in using student assessment data to compare the school to district or national performance



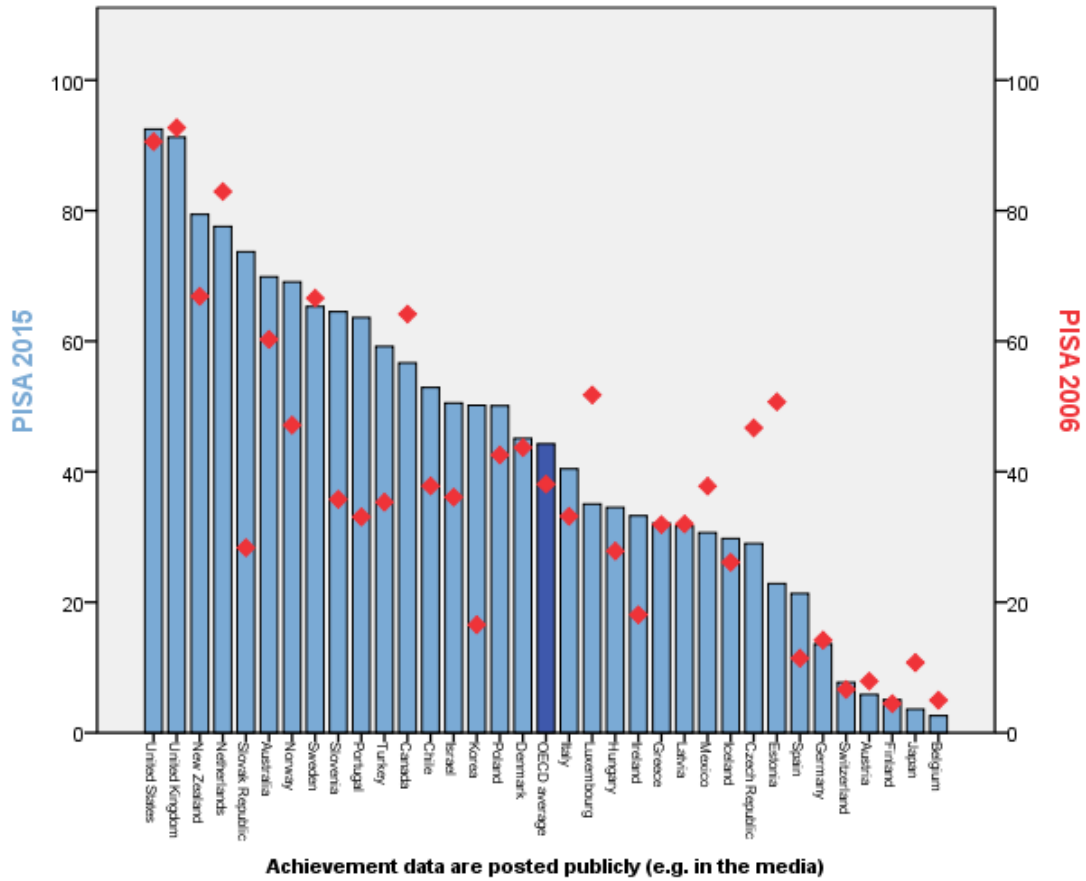
Source: Adapted from OECD 2004, 2016

Figure 8. Change between 2003 and 2015 in using student assessment data to compare the school with other schools



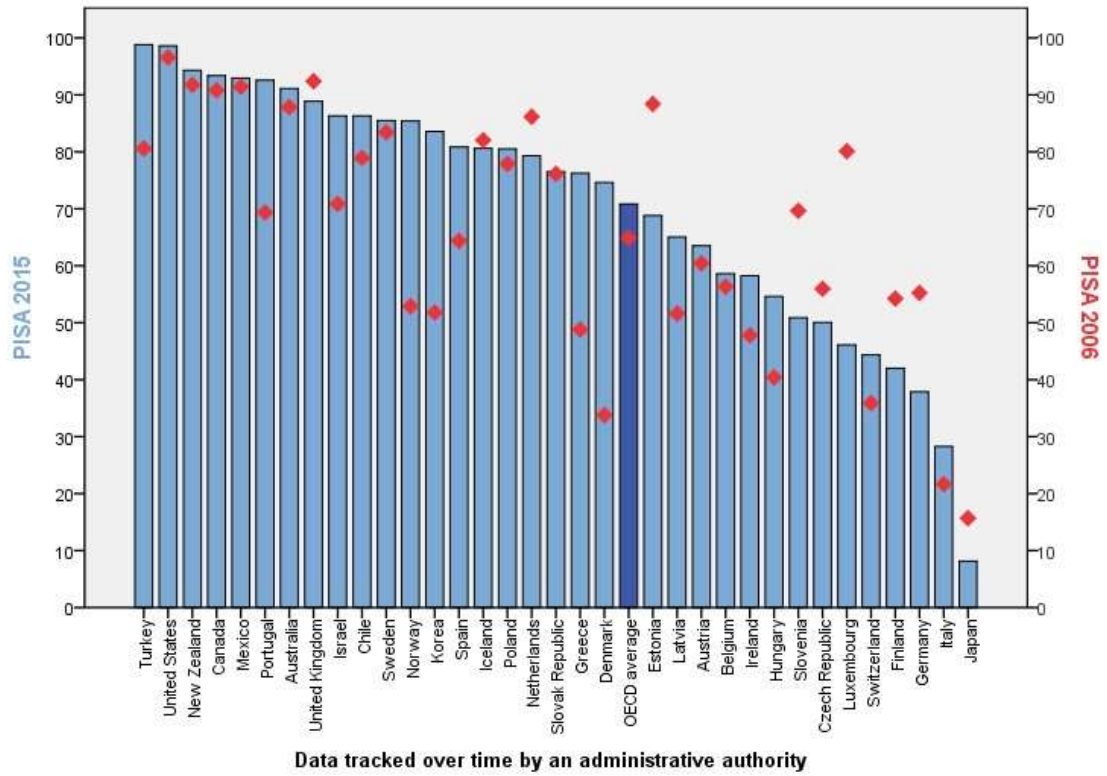
Source: Adapted from OECD 2004, 2016

Figure 9. Change between 2006 and 2015 in using achievement data in the following ways: Achievement data are posted publicly (e.g. in the media)



Source: Adapted from OECD 2007, 2016

Figure 10. Change between 2006 and 2015 in using achievement data in the following ways: Achievement data are tracked over time by an administrative authority



Source: Adapted from OECD 2007, 2016