OECD Education Working Papers No. 250

Does test-based school accountability have an impact on student achievement and equity in education? A panel approach using PISA

Rodrigo Torres

https://dx.doi.org/10.1787/0798600f-en

OECD

Organisation for Economic Co-operation and Development

EDU/WKP(2021)7

Unclassified **English text only**

12 July 2021

DIRECTORATE FOR EDUCATION AND SKILLS

DOES TEST-BASED SCHOOL ACCOUNTABILITY HAVE AN IMPACT ON STUDENT ACHIEVEMENT AND EQUITY IN EDUCATION? A PANEL APPROACH USING PISA

OECD Education Working Paper No. 250

Rodrigo Torres, OECD

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Rodrigo Torres, Rodrigo.Torres@oecd.org

JT03479386

OECD EDUCATION WORKING PAPERS SERIES

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

Abstract

School accountability is one of the most controversial recent reforms taking place in education systems around the world, but evidence of whether and which accountability practices affect equity and performance in academic achievement has been difficult to isolate and establish. By using data available from several cycles of the Programme for International Student Assessment (PISA 2006-2015), this paper assesses the extent to which accountability practices affect equity and performance in academic achievement in high-income-and-low-and-middle-income-countries. We found no conclusive evidence of accountability practices affecting educational outcomes in high-income-countries. However, we found some evidence in low-and-middle-income-countries pointing towards increased performance and increased inequality under accountability regimes in these contexts, although only in mathematics and science, and for one of our preferred specifications. In low-and-middle-income-countries, we found that, under higher levels of accountability, higher school autonomy on curriculum management and assessment could render better academic results in reading, mathematics and science.

Table of Contents

Tables

Figures

1. Introduction

The use of accountability practices is increasingly predominant in education systems. Previous studies have shown that the use of assessments for accountability has increased not only in developed countries, but also in the developing world (Bruns, Filmer and Patrinos, 2011[1]; Smith, 2016[2]; Phelps, 2005[3]). As the use of large-scale standardised tests has increased in previous years, so has the culture of testing for accountability (Benavot et al., 2007[4]; Smith, 2016[2]; European Commission, 2015[5]). Today, school accountability reform is driving education policy in many countries, changing schools' practices and goals (Fitz, 2003[6]; Schnellert, Butler and Higginson, 2008[7]; Smith, 2016[2]), expectations among school communities, parental school choice, and decision making at both local and government levels (Ladd, 2001[8]; Hastings, Kane and Staiger, 2005[9]). On the other hand, the idea that certain types of school accountability could be detrimental for disadvantaged students has been of particular concern. The purpose of this paper is to shed a light on the likely effects of school accountability on educational outcomes, and especially on equity in education, by making use of information from four cycles of PISA data (2006-2015).

What do we understand by school accountability?

The concept of school accountability has been treated differently in the literature and across disciplines. As suggested by Levitt, Janta and Wegrich (2008[10]), accountability can be a somewhat slippery concept in both theory and practice, as it is applied in various ways and under different circumstances. According to these authors, regardless of how accountability is defined, three elements are always present: the definition of who is accountable and for what, the information to stakeholders with regard to the results of what is being assessed (e.g. school academic performance), and a judgement about that information which could lead, directly or indirectly, to specific sanctions or rewards (depending, for instance, on whether the accountability system is a high-stakes or low-stakes system[1]).

A common definition of school accountability is the one elaborated by Figlio and Loeb (2011[11]). They define school accountability as "the process of evaluating school performance on the basis of student performance measures". In this definition of accountability, closely related to economic theory, students' aggregated results are at the heart of what school accountability is. In an education market where parents are allowed to choose the school of their children and/or education officials oversee the education process, the provision of information about schools and students' results would play two important roles: first, lowering the asymmetries of information for decision-makers (e.g. parents, local government, etc.) about the performance of schools. Second, and following the principal-agent theory, the provision of information on students' academic achievement and school performance would be crucial as a means to disregard opportunist behaviour from those in charge of running the educational process on a daily basis (the agents): head teachers, teachers, local administrative authorities. By holding those actors accountable for their results, the principal (parent, head teacher or government, depending on the case) could take the actions needed, sometimes in the form of rewards and sanctions, to exercise

[1] A high-stakes accountability system is that where there are explicit sanctions or rewards to those being assessed/judged, while low-stakes accountability provides information on the stakeholders but there are no direct actions linked to results.

pressure on the agent to change their behaviour in case it is not aligned with the principal's interest. For instance, measuring and reporting school performance and attaching consequences to meeting certain pre-defined standards, would provide incentives to educators to focus on the subjects and contents that are being measured, encouraging them to improve their practice. The sole provision of information could also prove to be itself a measure of pressure, as it could also influence, for instance, parents' behaviour in school communities, or parental school choice. As Figlio and Loeb (2011[11]) also remark, accountability systems could provide the incentives for stakeholders to improve students' results. However, at the same time, accountability systems could also produce perverse incentives, leading to unintended consequences if they are not carefully designed and monitored.

Previous studies have suggested that accountability should not necessarily apply only to learning outcomes, but also to school resources, professional capacity and other school processes (Stanford University/Stanford Center for Opportunity Policy in Education, 2015[12]). The rationale in this case is that in order for school accountability to work, transparency of information about all aspects of the schooling process is needed. Such transparency would lead, for example, to resources being distributed in an efficient and/or equitable way. In this case, standardised assessments results would become relevant only in the case where all other factors (especially resources and the teaching force) in the educational sector are also being overseen, supported and assessed (Bae, 2018[13]). The use of such assessments results at an individual and school level would be mostly focused on informing the generation of strategies to improve learning trajectories and the schooling process (Chapman and Snyder, 2000[14]). They should not, therefore, be used as indicators of compliance, or utilised as indicators for avoiding sanctions from external stakeholders.

In relation to the latter, in a recent report by UNESCO (2017[15]) on the topic, accountability was defined as "a process, aimed at helping actors meet responsibilities and reach goals". In this broad definition, accountability would involve most stakeholders in the education sector: governments, politicians, teachers, head teachers, parents and students. School accountability would be an important instrument by which education systems could improve by setting clear responsibilities and goals to the different actors. This approach would mostly place the policy emphasis on the transparency of information and the inputs of the education process (especially on the material and human resources), as opposed to its outcomes (students' assessment results), as, according to this report, many of the reward and sanctioning mechanisms that make use of measures of learning achievement, when poorly designed, could become harmful, especially for disadvantaged students.

Whether there is an approach towards the use of students' results to evaluate school performance, or whether we understand accountability as a broader concept also involving the use of other types of information, it is undeniable that the concepts of educational assessment, standard-based reform and school accountability are interlinked and have developed side by side in previous years (Kamens and McNeely, 2010[16]). Most accountability systems currently available depend on the existence of information about stakeholders' performance (usually most systems limit themselves to providing information about students' academic results and assessing teachers or head teachers accordingly), and also on the existence of specific goals or benchmarks which need to be ideally met, and against which schools can be assessed (Kornhaber, 2004[17]).

In this regard, previous OECD reports on PISA results (OECD, 2013[18]) provide evidence of a strong association between accountability practices (such as reporting information to parents or posting results publicly) and the use of assessment results at schools for other

purposes. For instance, the higher the use of assessments at schools for curricular or instructional purposes, the more likely those schools are also to report results to parents or to compare their performance with other schools[2].

In this paper, we will focus our analysis on the impact of school accountability on educational outcomes – understanding accountability as the use of students' results to assess schools' and teachers' performance. This approach was chosen due to the fact that the available data on this topic from previous PISA applications is mostly related to this definition. We also acknowledge that although alternative definitions exist, the large majority of the literature on the topic of school accountability makes use of this definition. The importance of this type of accountability is also reflected in the strong link between accountability practices and the widespread utilisation of national standardised assessments and national learning standards.

Evidence of the effect of school accountability policies on learning achievement

There is extensive literature on the effect of school accountability systems on average academic achievement. Usually, these studies exploit policy interventions in specific states or countries attempting to estimate the causal impact of accountability practices in students' and schools' average tests scores. Especially prolific is the evidence from the United States, obtained from the No Child Left Behind Act[3] (NCLB). Several studies there showed that there was an increase in average achievement in schools, at least in one grade and subject, after this policy was enacted (Jacob, 2005[19]; Carnoy and Loeb, 2002[20]; Hanushek and Raymond, 2005[21]; Dee and Jacob, 2011[22]; Chiang, 2009[23]).

In the United Kingdom, Burgess, Wilson and Worth (2013[24]) found that the abolition of school league tables – a very common form of school accountability – in Wales had a negative impact on academic outcomes and had no effect on academic or social segregation at schools. Rosenthal (2004[25]) found an adverse effect of school inspections from the Office of Standards in Education (Ofsted) on educational outcomes in England. Bradley and Taylor (2002[26]) found that the introduction of school league tables and greater parental choice in England led to higher educational outcomes, but also higher social segregation between schools.

The evidence from developing countries is more limited and mixed. In Pakistan, Andrabi, Das and Khwaj (2014[27]) found a significant positive impact on students' results after report cards were being used to inform parents. In a randomised control trial in Uganda, the distribution of scorecards containing information about several school practices and processes (such as teachers' and students' attendance, school teaching materials, and infrastructure and assessment results, among others) led to higher gains in achievement in schools where parents were asked to participate in the scorecards' content definition (Barr et al., 2012[28]). Another programme in India where parents were introduced to student learning levels in their own villages, and trained on how to easily assess their children's reading skills, did not result in any improvement (Banerjee et al., 2010[29]). In Chile, Mizala and Urquiola (2013[30]) showed that the implementation of a national system used to assess schools' performance relative to similar socio-economic status (SES) schools reporting

[2] However, this does not necessarily imply that assessment must lead to accountability. There are cases where assessment systems exist and results are not made public or used to evaluate teachers or schools (Smith, 2016[2]).

[3] The No Child Left Behind Act (NCLB) was a federal law that provided funding for additional educational assistance for poor children in return for improvements in their academic progress. It established learning standards, annual assessments and reports about schools results to the public.

"winning schools" to parents did not have any impact on the schools' social composition or enrolment. Rau and Contreras (2012[31]) found that the associated prize for being a teacher at a "winning school" had a modest positive impact on learning achievement.

Although these results tend to show a positive association between school accountability practices and average learning achievement, the benefits of accountability systems for equity in learning are less clear, and the evidence less conclusive. In an analysis of English schools, Burgess, Propper, Slater and Wilson (2005[32]) found that the publication of league tables measuring pass rates led to worse overall results for low-performing students in secondary education, and this effect increased when there was higher competition for students. In the United States, Hanushek and Raymond (2005[21]) did find larger gaps between black and white students after the No Child Left Behind policy was implemented, with Hispanic students benefiting the most, reducing the gap with white students. There is also evidence that information on students' performance along with greater school choice has an important impact on parental choice, especially for more educated parents, leading to increased social and academic segregation (Bifulco and Ladd, 2007[33]; Hastings and Weinstein, 2008[34]; Hastings, Kane and Staiger, 2005[9]).

These research findings have, however, been contested in previous studies. (Dee and Jacob, 2011[22]) did not find evidence of increased inequality in outcomes according to achievement levels. They show a bigger positive impact of the NCLB policy for Hispanic and Black populations, students under free school meals, and male students for the subject of math in 4th grade. However, they also report a negative effect on equity, in favour of white students in the same grade for reading. In a previous paper examining the effect of this policy in Chicago public schools, Jacob (2005[19]) reports higher achievement gains in low-performing schools, with larger proportions of students not meeting the set minimum standards. This shows that students belonging to the 10th to 50th percentile of previous academic achievement benefitted more from the policy. In a similar study for Texas schools, (Reback, 2008[35]) found that only the lowest-achieving students, and those students who were critical for the school's rating, benefitted from the NCLB program in the subject of mathematics. According to the author, this could be due to short-run incentives to reallocate teaching resources towards those students.

When looking at cross-national studies, evidence is also mixed and quite scarce. Using PISA 2003 data, Schutz, West and Woessmann's (2007[36]) results suggest that the existence of external exit examinations would not affect equity in educational results importantly and would improve overall achievement in education systems. In a similar study, Woessmann (2005[37]) makes use of TIMSS, TIMSS-Repeat and PISA data to estimate the effect of external exit exams on student achievement. He finds a positive effect of external exams in the academic achievement of all students. However, this effect was especially big for high-achieving students, leading to a wider achievement distribution and hence increased inequality.

Analysing a set of school reforms taking place in 24 European countries in the period 1929-2000, Braga, Checchi and Meschi (2013[38]) found an average negative effect of school accountability[4] on educational attainment, measured as years of schooling and highest qualification acquired. They also found increased inequality in attainment after the reform, mainly driven by the fact that low-attainment students would be especially sensitive to it, showing worse than average outcomes.

[4] More specifically, they define school accountability as a composite score dependent of the following reforms: introduction of standardised tests, school evaluation policies, school autonomy and teacher autonomy.

Using PISA 2009 information, Smith (2016[2]) found improved academic achievement in math in those schools posting results publicly, although his results do not take into account possible confounders. This association was found in school systems with both summative, evaluative and sanction/reward-oriented assessments. Finally, previous OECD reports (OECD, 2013[18]) have also explored this association, showing that after taking into account GDP per capita there is no direct link between accountability and higher average achievement. These findings have been partially disputed in previous reports, showing that when accountability systems are in place and schools enjoy high levels of autonomy they show improved student achievement (OECD, 2011[39]).

By using PISA data for years 2006-2015, our analysis will build upon this strand of the literature. We will explore a similar research question to that of Schutz, West and Woessmann (2007[36]), and Woessman (2005[37]) although we will focus on different measures of school accountability. We believe research findings from previous studies in this topic rely on strong assumptions, since their estimates make use of cross-sectional data for only one cohort of students – where it is difficult to isolate school accountability from other related institutional settings of each country – which might also be affecting educational outcomes. In order to tackle this limitation, we will produce panel estimates for the effect of accountability on educational outcomes using several cohorts of students. This work will have a specific emphasis on the effect of accountability on equity in learning outcomes and will also explore the association between accountability and school autonomy.

Unintended consequences of school accountability systems

Although there is currently no agreement on the convenience of the existence of high-stakes accountability systems and on the type of school practices related to accountability that would be more appropriate for educational communities and students' interests (Fitz, 2003[6]; Ravitch, 2010[40]), there is wide agreement on the fact that, if poorly designed, some accountability systems could pose important challenges to school communities and education systems in general (Bifulco and Ladd, 2007[33]; Figlio and Loeb, 2011[11]; Ravitch, 2010[40]).

There is extensive literature on the potential detrimental consequences of some school accountability practices. Some examples of unintended consequences are: increased social segregation at schools due to, for instance, incentivising students' mobility to better off schools via parents' school choice (Davis, Bhatt and Schwarz, 2015[41]); narrowing the curriculum and focusing teachers' attention only on the measured subjects, or focusing teaching activities to preparing students to "sit the test" (Jennings and Rentner, 2006[42]; Supovitz, 2009[43]); larger gaps in academic achievement due to heterogeneous impact of accountability systems across subpopulations (e.g. ethnic minorities vs. white students; (Hanushek and Raymond, 2004[44])) or difficulties recruiting and retaining teachers in low-performing schools (Clotfelter et al., 2004[45]). There is also evidence of schools' increased competition to attract the best students in order to perform well in school league tables, and low incentives to recruit low-performing students, leading to academic segregation in schools. Some studies have even shown that some schools "game" accountability systems, for instance, reclassifying low-performing students to test-excludable categories, such as special education students (Figlio and Getzler, 2002[46]); or simply altering exam results in order to meet specific standards or avoid sanctions (Jacob and Levitt, 2003[47]). This evidence only emphasises the need to produce more research on this topic, to understand possible mechanisms for modifying or removing certain accountability systems.

Relationship between school accountability, school autonomy and educational outcomes

As it has been hypothesised in previous studies, the existence of school autonomy has also been associated with improved student outcomes in certain school environments. According to a previous OECD report, accountability would work when accompanied with autonomy for decision making at schools (OECD, 2011[39]). In contexts with higher autonomy for decision making on resource allocation or curriculum management, school systems would perform better under higher accountability regimes. The importance of school autonomy has also been backed by previous studies on the topic using PISA data, but only in higher development countries (Hanushek, Link and Woessmann, 2013[48]) and for specific practices. Although it makes sense to believe that only in school systems under highly competent school management, autonomy is beneficial, it remains unclear whether autonomy would actually work in high accountability systems. At the same time, the hypothesis of whether autonomy mediates the relationship between accountability and educational achievement is also still to be explored.

2. Data and methods

General description of the data used

This paper uses PISA data available for years 2006-2015, comprising four PISA implementations[5]. The Programme for International Student Assessment (PISA), has been applied to several countries and economies since 2000. Although the countries participating in the application differ from year to year, most countries in our analysis have been present in three or four cycles[6] since 2006. PISA is applied to 15-year-old students attending any type of school and grade (above grade 7), covering three different areas (mathematics, reading and science)[7]. Test results per subject are standardised to have a mean of 500 and a standard deviation of 100 score points across OECD countries[8].

To perform our analysis, we produced a country-level panel dataset, by merging data on students' academic achievement, schools' practices and institutional characteristics[9] and students' socio-economic characteristics for 4 PISA cycles. In total our panel consists of a

[5] PISA 2015 was a computer-based assessment, while implementations in 2006, 2009 and 2012 were paper-based. We assume that plausible values for PISA scores are comparable across cycles for the same countries.

[6] In 2006, 57 countries participated. In 2009, 75 countries. In 2012 and 2015, 65 and 72 countries participated respectively.

[7] PISA covers a set of skills, knowledge and competences defined by OECD as relevant for personal, social and economic well-being, in four domains: Mathematical Literacy, Reading Literacy, Scientific Literacy and Problem Solving Skills.

[8] In order to produce a representative sample of students per country, a two- or three-stage sampling procedure was performed for each application. Although the number of schools and students differs by country and application, usually at least 1 500 students have been assessed in each cycle, with around 35 students per participating school.

[9] These are usually aggregated measures of schools' observed practices and their socio-demographic characteristics.

sample of around 1 150 000 students from 63 different countries. We decided to only include countries with at least two PISA implementations in the period 2006-15, which also had available information about school practices[10]. Our final sample of countries incorporates countries in Latin America, South East Asia, South Asia, Africa, Central and Eastern Europe, North America and Oceania[11].

To produce the final merged dataset, we use information from different sources: the school questionnaire (applied to all head teachers during the examination), the student questionnaire (applied to all students sitting the test) and students results data, containing PISA scores (plausible values) for all students sitting the test in the subject of mathematics. The head teacher's questionnaire contains information about schools' practices (including school accountability and school autonomy practices) and other school characteristics, such as school size, geographical location or proportion of full-time teachers. The student questionnaire contains information about students' demographic characteristics, such as gender, first language or pre-school education; and household characteristics, such as parental education, working status and their occupation, as well as the availability of learning resources at home.

In our specifications we will make use of several individual and family characteristics present in all involved PISA questionnaires. We will also be using school-level characteristics and school practices, for producing aggregated measures at the country level (here our measures of school accountability and school autonomy). A more detailed explanation for some of these measures can be found in the following sections.

In the following table, we present average values for the whole sample of students and family characteristics for the applications in years 2006 and 2015.

Table 1. Students' and school characteristics by PISA cycle (years 2006 and 2015)

	PISA 2006					*PISA 2015*				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Student Level Characteristics										
Maths Score. PISA test.	398 750	471.55	105.97	0.62	921.01	519 334	463.08	103.25	0.00	870.51
Age	398 734	15.78	0.29	15.17	16.33	514 397	15.79	0.29	15.17	16.42
Gender (Female==1)	398 746	0.51	0.50	0	1	519 334	0.50	0.50	0	1
Inmigration background (1st generation==1)	388 458	0.04	0.20	0	1	494 957	0.06	0.24	0	1
Inmigration background (2nd generation==1)	388 458	0.05	0.21	0	1	494 957	0.06	0.23	0	1
Inmigration background (Other language at home==1)	384 488	0.12	0.33	0	1	498 267	0.13	0.34	0	1
Index of Socioeconomic Status (ESCS)	378 651	-0.48	1.15	-6.83	3.80	503 371	-0.26	1.11	-7.26	4.18
School Level Characteristics										
School operation (Publicly operated==1)	369 837	0.82	0.38	0	1	460 819	0.79	0.41	0	1
School Funding (Share of budget paid by government)	360 655	0.81	0.29	0	1	432 523	0.81	0.31	0	1
School Funding (Share of budget paid by fees)	335 216	0.16	0.28	0	1	385 213	0.16	0.31	0	1
School Location (Small town==1) Ref cat: Village or rural area	382 747	0.22	0.42	0	1	505 366	0.20	0.40	0	1
School Location (Town==1)	382 747	0.31	0.46	0	1	505 366	0.28	0.45	0	1
School Location (City==1)	382 747	0.25	0.43	0	1	505 366	0.23	0.42	0	1
School Location (Large city==1)	382 747	0.11	0.31	0	1	505 366	0.14	0.35	0	1
Student teacher ratio	370 005	16.90	17.37	0	348	439 775	14.95	15.97	0	911
Number of students	383 117	838.37	679.99	0	10 000	451 753	897.76	864.20	0	17 805
Share of fully certified teachers	398 750	0.87	0.28	0	1	425 029	0.82	0.39	0	1
Share of full time teachers	365 280	0.85	0.22	0	1	458 845	0.89	0.19	0	1

[10] Here test-based accountability measures, amongst others. In practice this data was present for those countries where school questionnaires were applied to head teachers, and the information was publicly available.

[11] A detailed list of countries is included along with some descriptive statistics in the following section.

Measure of student socio-economic status: the ESCS Index

The OECD produces an index of socio-economic status for each student participating in PISA: the so called "PISA index of Economic, Social and Cultural Status" (ESCS), which is a continuous measure estimated via principal component analysis techniques[12]. This index aims to summarise information about individuals' socio-demographic characteristics, and it is comparable across countries within each test implementation. An adjusted measure of this index which allows for comparability across PISA cycles over time has also been made available by OECD. As we will be mostly concerned with studying changes over time in inequality of PISA results, we will be making use of this adjusted measure during our estimations.

There is important variation in the ESCS index across and within countries (OECD, 2013[49]). In our estimations, we use this index as a proxy for student socio-economic status, studying whether there are systematic differences in the relationship between school accountability practices and student educational outcomes, according to this measure.

Measures of test-based school accountability.

The measure of school accountability that we will be using for this study reflects a measure of test-based external accountability, as it is focused on informing external stakeholders about school performance. Head teachers in participating countries were asked whether school achievement data was posted publicly in their schools. For each PISA cycle separately, we constructed a measure of the proportion of students in each country attending schools declaring to post results publicly. This measure is available for 4 PISA cycles (2006-2015)[13]. When aggregating data at the country level, we used this measure as a proxy of how widespread and common this practice is in each country, also allowing us to assess whether there have been any relevant changes over time in this measure, mostly reflecting changes in policy and national trends. Similar studies have used a similar approach using PISA data when studying education reform.

The following Table 2 and Figure 1 indicate that there are systematic differences across countries for the time-period of interest. We also observe relevant changes over time in this measure for several countries, both in higher- and also lower- and middle-income countries. We focus our analysis on how changes in these measures could be associated to changes in educational performance and equity in education.

[12] More specifically, this index makes use of information derived from measures of occupational status, parents' educational level and home possessions, as well as books available at home. More information on this index and its comparability over time can be found at OECD (2013[49]).

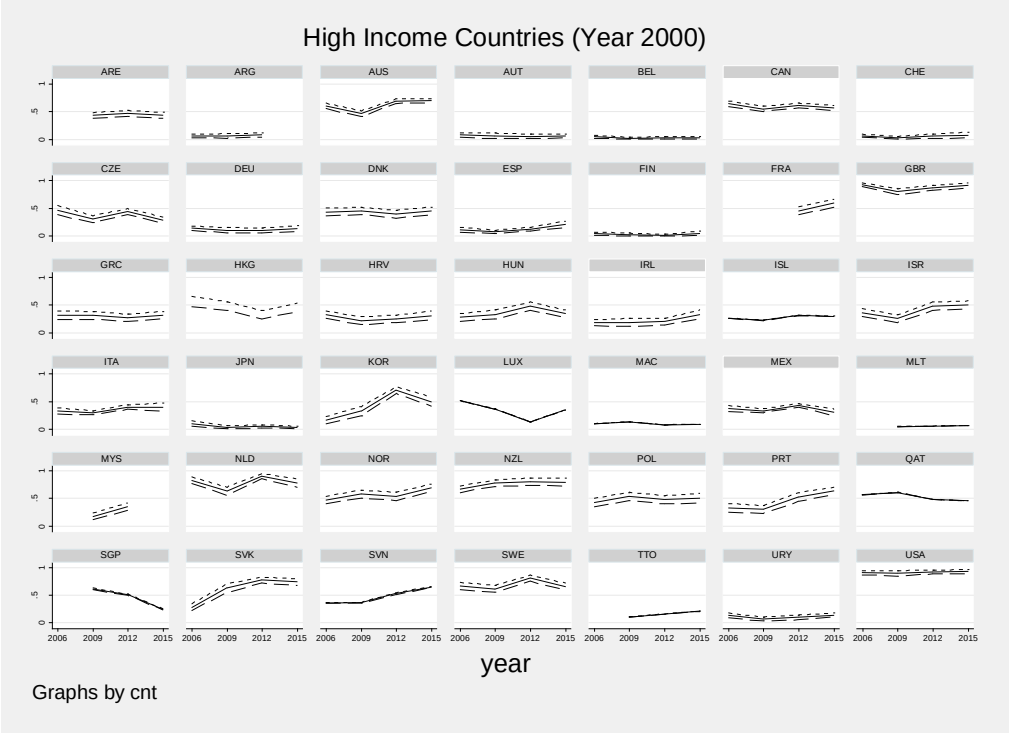
[13] Although there are similar measures of test-based accountability on previous PISA questionnaires, the wording of the related items for years previous to 2006 is quite different, implying that comparability over time could be compromised, what may affect our results importantly. Hence the decision not to use data from previous PISA cycles.

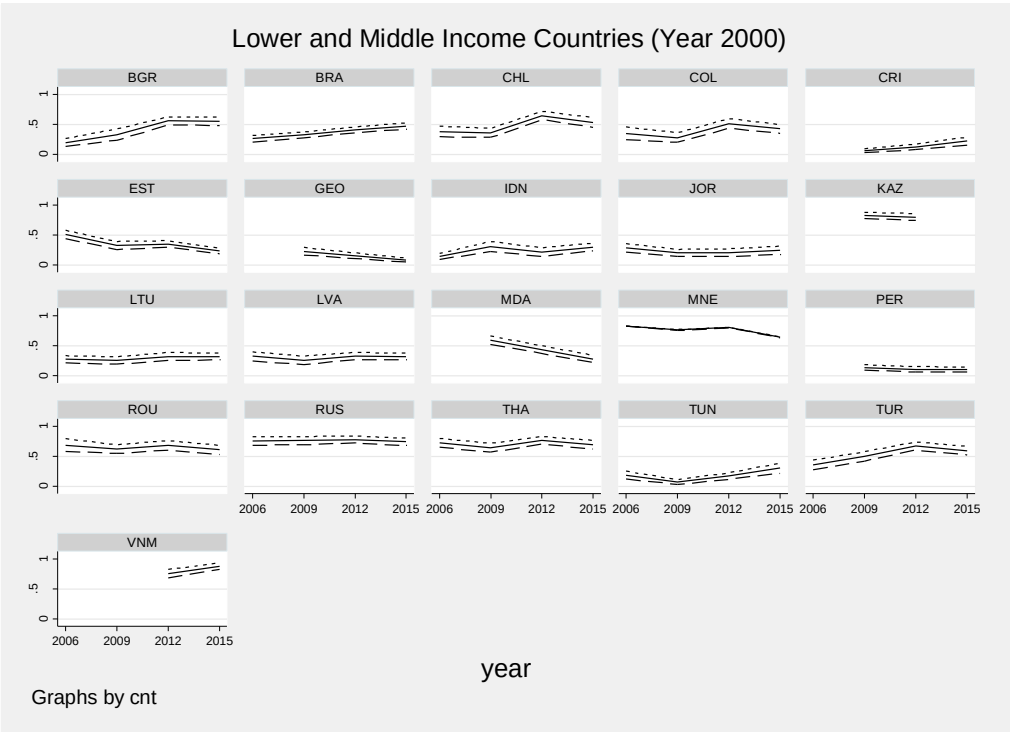
Table 2. Changes in school accountability and achievement in PISA Maths (2006-2015)

	Average Achievement in PISA Maths				School Accountabily - Results Posted		
	2006	2015	chg		2006	2015	chg
Lower and Middle Income Countries							
BGR	438.4	427.9	-10.6	BGR	0.55	0.19	-0.35
BRA	388.5	386.0	-2.4	BRA	0.47	0.26	-0.21
CHL	422.4	420.7	-1.7	CHL	0.53	0.38	-0.15
COL	376.6	380.7	4.2	COL	0.42	0.35	-0.07
CRI	407.2	409.9	2.7	CRI+++	0.22	0.12	-0.10
EST	520.0	512.0	-7.9	EST	0.23	0.51	0.28
IDN	375.5	371.1	-4.4	IDN	0.30	0.14	-0.16
JOR	385.6	386.6	1.0	JOR	0.24	0.29	0.05
KAZ	432.2	405.5	-26.7	KAZ+/+++	0.82	0.80	-0.02
LTU	478.2	476.5	-1.7	LTU	0.32	0.27	-0.05
LVA	490.4	481.5	-9.0	LVA	0.32	0.32	0.00
MDA*	398.5	397.4	-1.1	MDA ++++	0.28	0.58	0.31
MNE	409.5	402.7	-6.8	MNE	0.64	0.83	0.19
PER	368.1	364.8	-3.3	PER +++	0.10	0.10	0.00
ROU	445.3	426.4	-18.9	ROU	0.61	0.69	0.08
RUS	481.9	467.9	-14.0	RUS	0.74	0.75	0.01
THA	426.6	418.6	-7.9	THA	0.69	0.72	0.03
TUN	387.6	371.5	-16.2	TUN	0.30	0.18	-0.11
TUR	447.4	445.7	-1.7	TUR	0.59	0.35	-0.24
VNM**	511.2	510.6	-0.6	VNM++	0.75	0.88	0.13
Average	**429.6**	**421.1**	**-6.4**		**0.44**	**0.44**	**-0.02**
Higher Income Countries							
ARE	433.9	420.7	-13.2	ARE+++	0.43	0.47	0.03
ARG	388.5	387.5	-1.0	ARG+	0.05	0.06	0.01
AUS	503.8	514.6	10.7	AUS	0.70	0.60	-0.10
AUT	505.6	495.4	-10.2	AUT	0.06	0.08	0.02
BEL	515.2	515.7	0.5	BEL	0.03	0.05	0.02
CAN	518.0	526.3	8.3	CAN	0.57	0.64	0.07
CHE	530.5	535.0	4.5	CHE	0.08	0.07	-0.01
CZE	498.8	492.6	-6.2	CZE	0.29	0.47	0.18
DEU	513.7	512.1	-1.6	DEU	0.14	0.14	0.01
DNK	499.9	503.2	3.3	DNK	0.45	0.44	-0.01
ESP	484.6	483.7	-0.9	ESP	0.21	0.11	-0.10
FIN	519.1	540.4	21.3	FIN	0.05	0.04	-0.01
FRA	496.1	496.8	0.6	FRA+++	0.59	0.46	-0.13
GBR	494.0	492.5	-1.5	GBR	0.91	0.93	0.01
GRC	453.5	465.4	12.0	GRC	0.32	0.32	0.00
HKG	561.4	554.7	-6.7	HKG	0.46	0.56	0.10
HRV	471.0	460.6	-10.3	HRV	0.31	0.33	0.02
HUN	477.7	490.0	12.2	HUN	0.35	0.28	-0.07
IRL	501.1	487.3	-13.8	IRL	0.33	0.18	-0.15
ISL	493.1	507.4	14.3	ISL	0.30	0.26	-0.04
ISR	466.8	447.4	-19.5	ISR	0.51	0.36	-0.14
ITA	485.0	483.3	-1.7	ITA	0.40	0.33	-0.07
JPN	536.7	529.2	-7.5	JPN	0.04	0.11	0.07
KOR	554.3	545.9	-8.3	KOR	0.50	0.17	-0.34
LUX	489.6	488.2	-1.5	LUX	0.35	0.52	0.17
MAC	538.1	525.0	-13.1	MAC	0.09	0.10	0.01
MEX	413.2	418.5	5.3	MEX	0.31	0.38	0.07
MLT*	472.9	462.6	-10.3	MLT ++++	0.07	0.08	0.00
MYS	421.3	404.2	-17.1	MYS+/+++	0.19	0.35	0.16
NLD	522.4	525.9	3.5	NLD	0.78	0.83	0.05
NOR	489.7	497.5	7.8	NOR	0.69	0.47	-0.22
NZL	500.3	519.9	19.6	NZL	0.79	0.67	-0.13
POL	517.6	494.2	-23.4	POL	0.50	0.43	-0.08
PRT	486.5	487.3	0.8	PRT	0.64	0.33	-0.31
QAT	376.9	368.5	-8.4	QAT	0.46	0.57	0.11
SGP	573.4	562.6	-10.8	SGP +++	0.24	0.51	0.27
SVK	481.5	496.7	15.2	SVK	0.74	0.28	-0.45
SVN	500.8	501.0	0.3	SVN	0.65	0.36	-0.29
SWE	478.2	493.9	15.7	SWE	0.65	0.67	0.01
TTO*	417.5	414.5	-3.0	TTO	0.21	0.21	0.00
URY	409.1	427.2	18.1	URY	0.14	0.13	-0.01
USA	480.7	487.4	6.6	USA+	0.88	0.91	0.02
Average	**487.4**	**487.2**	**0.3**		**0.39**	**0.36**	**-0.03**

Note 1. School Accountability- Results posted: Percentage of schools where achievement data are posted publicly. Note 2. Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries. Note 3. + Information on school accountability for 2009 instead of 2006. ++ Information on school accountability for 2012 instead of 2006. +++ Information on school accountability for 2012 instead of 2015. ++++ Information on school accountability for 2009 instead of 2015.
Note 4. * Information on achievement for 2009 instead of 2006.

Figure 1. Percentage of students in schools where achievement data are posted publicly





Note: Proportion of schools where student achievement data are posted publicly (0 to 1 scale). Period 2006-2015. Confidence intervals at 95% level.

Measuring school autonomy practices

In order to measure school autonomy, we make use of several items present in the school questionnaire for all PISA cycles involved in our analysis (2006-2015). Following previous OECD reports (OECD, 2010[50]), we classified most school/teaching practices according to two main criteria: whether they referred to management of school resources – including human resources – or curriculum and assessment practices.

Later on, we narrowed down the definition of school resources to human resources, in a dimension we called "teacher management". These sub-divisions were based in both semantic analysis as well as on the estimated correlation between these measures at the country level. Instead of using a single item we decided to construct a score utilising principal component analysis to construct a measure summarising degrees of autonomy at the school level, both for curriculum and assessment and teacher management separately[14] [15].

Empirical approach

The effect of school accountability on educational outcomes

Our empirical approach assumes that students' results in PISA are affected by individual, school- and institutional-level characteristics (here: school accountability). Following a standard educational production function, we assume a linear relationship between these characteristics and individuals' academic achievement. More formally:

$$P_{csti} = \beta_I I_{ct} + \beta_s S_{csti} + \beta_x X_{csti} + \varepsilon_{csti} \qquad (1)$$

where *P* is a vector of PISA plausible values for individual *i* in time *t*, in country *c* and school *s*. S is a vector of school characteristics, *X* is a vector of individual and household characteristics, and *I* is a vector containing observed institutional characteristics in country *c* in time *t*. In this case time t denotes application year, as the PISA assessment is implemented every 3 years.

Since we have one or more PISA implementations per country, it is possible to account for unobserved heterogeneity both at the country level and over time. In this case we assume the error term ε can be expanded into three elements

$$\varepsilon_{csti} = \mu_c + \mu_t + \mu_{csti} \qquad (2)$$

The parameters μ_c and μ_t denote country fixed effects and time fixed effects. The former represents all unobserved institutional characteristics affecting academic achievement that are fixed in time. Time fixed effects, account for cohort effects across PISA assessments. The error term μ_{cti} is assumed to be random and uncorrelated to the measures of interest of school accountability and school autonomy contained in vector *I*.

[14] To establish a measure of autonomy on curriculum and assessment we produced a principal component analysis utilising four different items: establishing student assessment policies; choosing which textbooks are used; determining course content, and deciding which courses are offered. Tetrachoric correlations of the different items at the school level where all above 0.6. The factor analysis produced only one factor.

[15] To produce a measure of autonomy on teacher management, we utilised the following items from the school questionnaire: selecting teachers for hire; dismissing teachers; establishing teachers' starting salaries; determining teachers' salaries increases. Tetrachoric correlations of the different items at the school level where all above 0.6. Factor analysis produced only one factor.

Our main research question is concerned with the effect of accountability in educational outcomes, as well as its association with inequality in PISA results. Hence, we incorporate an interaction term between student SES and time-varying institutional characteristics *I* to the model

$$P_{csti} = \beta_{I1} I_{ct} + \beta_{I2}(I_{ct}\ x\ ESCS) + \beta_s S_{csti} + \beta_x X_{csti} + \mu_c + \mu_t + \mu_{cti} \quad (3a)$$

Where the measure of student SES (ESCS index) is also incorporated in the vector of student characteristics *X*. It is worth underlining that in this case we have constructed a country level pseudo-panel, given the fact that individuals sit PISA only once at age 15, so in practice we are comparing individuals from different cohorts within countries.

As an alternative specification, we explored the possibility of using school SES instead of individual SES as the measure of interest when exploiting inequality in PISA results. In this case we explore whether accountability has a differential impact on educational outcomes according to the average level of students' socio-economic status in the schools of interest[16].

The main advantage of these two alternative specifications is that we can exploit variations on school accountability within countries over time, allowing us to separate those variations from fixed institutional factors also affecting inequality of outcomes. Those fixed institutional factors are usually correlated to school accountability measures and unobserved in most cross-sectional studies.

Our main concern, however, is that we identify the effect from a very small sample of observations, as we measure most country-level variables only up to four times, and we have a limited number of countries available (63). As this renders our estimation more sensitive to the model specifications, we are only able to incorporate few time-varying institutional characteristics related to school accountability at a time. We add other time-varying institutional characteristics in some of our specifications when available[17]. In the following table, we show a correlation matrix between our different measures of school accountability, school autonomy, and other institutional characteristics for PISA 2015. Of particular importance is the high correlation between our different measures of school autonomy (curriculum and assessment versus teacher autonomy). Hence, we decided not to include them all in one model, but to test their association with educational outcomes one at a time.

[16] In this case the specification is $P_{csti} = \beta_{I1} I_{ct} + \beta_{I2}(I_{ct}\ x\ SchoolESCS) + \beta_s S_{csti} + \beta_x X_{csti} + \mu_c + \mu_t + \mu_{cti}$ (3b). Where the vector X contains a student level measure of socioeconomic status (ESCS index).

[17] In the present version of this work, we added school property, school funding and school admission policies in some of our specifications.

Table 3. Correlation between school accountability, school autonomy and other institutional characteristics

PISA 2006-2015 (Observations: 223. 63 countries)	School Accountability - Results Posted	School Autonomy- Curriculm and Assessment	School Autonomy- Teacher Management	School Admission Policies - Prior Achievement	School Admission Policies - Residential Criteria	School Operation (Public)	School Funding (Government)
School Accountability - Results Posted	1.00						
School Autonomy- Curriculm and Assessment	-0.04	1.00					
School Autonomy- Teacher Management	**0.26**	**0.43**	1.00				
School Admission Policies - Prior Achievement	0.02	-0.09	0.02	1.00			
School Admission Policies - Residential Criteria	0.07	**-0.17**	**-0.24**	**-0.31**	1.00		
School Operation (Public)	**0.15**	**-0.26**	0.11	**-0.21**	**0.33**	1.00	
School Funding (Government)	**0.20**	-0.03	**0.38**	**-0.18**	**0.17**	**0.28**	1.00

Note 1. Coefficients in **bold** significant at 5% level.
Note 2. i. School Accountability - Results Posted: Percentage of schools where achievement data are posted publicly. ii. School Admission Policies - Prior Achievement: Percentage of schools using prior achievement as a criterion for student selection. iii. School Admission Policies - Residential Criteria: Percentage of schools using residential area as a criterion of student selection. iv. School Operation: Percentage of schools declaring to be public. v. School Funding: Schools' average proportion of income coming from government sources.

School accountability and school autonomy

As a second research question, our study is interested in investigating whether school autonomy could work as a facilitator for accountability practices to have an impact on educational outcomes. That is, whether under the presence of accountability of students' results, the level of autonomy that school practitioners have in decision making regarding certain school practices, allows them to improve educational outcomes.

To investigate this hypothesis, we explore a specification where the level of accountability and autonomy at the country level interact:

$$A_{csti} = \beta_{I1} I_{ct} + \beta_{I2}(IAcc_{ct}\, xIAut_{ct}) + \beta_s S_{csti} + \beta_x X_{csti} + \mu_c + \mu_t + \mu_{cti} \quad (4a)$$

In this specification both school accountability and school autonomy are measured at the country level. That is, conditional on fixed levels of accountability, we explore whether changes in the levels of autonomy could make a difference explaining educational performance. If there is a differential impact, this would be reflected in the interaction term.

In an alternative specification, we explore whether those schools showing different levels of autonomy in certain countries, show a different performance depending on the overall level of accountability present in the country. This specification allows us to exploit variability in school autonomy across schools for each country, although estimates from it are more subject to bias[18].

[18] By including autonomy measured at the school level, it is possible that other unobserved school characteristics related to school autonomy are captured in this interaction term. In this case the specification is the following: $A_{csti} = \beta_{I1} I_{ct} + \beta_{I2}(IAcc_{ct}\, xIAut_{sct}) + \beta_s S_{csti} + \beta_x X_{csti} + \mu_c + \mu_t + \mu_{cti}$ (4b). Where the vector I contains a measure of accountability at the country level, and the vector S, contains a measure of school autonomy reported at the school level. In this case the coefficient β_{I2} represents the additional number of points in PISA for one additional unit in the autonomy score at the school level (for a fixed level of accountability at the country level).

Estimation methods

We estimated the aforementioned models following standard regression techniques, accounting for the survey design of PISA by using the Stata command *Repest*. The *Repest* Stata command[19] considers the fact that our data comes from four different PISA implementations, each one with a particular survey design. It also accounts for the fact that PISA incorporates 5 or 10 imputed plausible values for each test score. The *Repest* module allows us to work with pooled data from several PISA cycles, estimating consistent coefficients.

As previously mentioned, one of the main risks when using institutional characteristics measured at the country level is that, in practice, regressions are based on a very small number of observations. Moreover, the available estimation command accounts for complex PISA survey designs and sampling variation at the school level, but not for this variability at the country level. This implies that our estimates do not take into account sampling variation according to the countries used to run our models. In order to tackle this issue, we decided to estimate bootstrapped standard errors when reporting coefficients at the country level. In this case our estimates account for country-level sampling variability by estimating the same model several times, taking one country out of the sample at a time, for all 63 countries under analysis[20].

On the interpretation of results and model assumptions.

When estimating the effect of accountability practices on educational performance, and in case the coefficient of interest shows to be significant, the estimated coefficients will correspond to the additional number of units on the PISA score associated with a one percentage point increase in the proportion of students in schools posting results publicly.

For the case of school accountability and equity in PISA scores, if the above models are correctly specified, we can assume that in equation 3a, and for a fixed level of school accountability, the incremental impact of an additional unit in the student ESCS index is β_{I2} additional points in the PISA's mathematics test. Since the individual level ESCS index is standardised to have a zero mean and a standard deviation of one across our sample, in this case this also means that one additional standard deviation of the ESCS index is linked to β_{I2} additional points in PISA. In other words, a student showing one additional standard deviation on the ESCS index score is likely to gain β_{I2} additional points in PISA, when compared to an average student.

A similar conclusion can be derived when using the average level of individual ESCS at the school level – as a proxy of school socio-economic status (instead of the ESCS index,

[19] The *Repest* command estimates statistics using replicate weights, accounting for survey designs in the estimation of the sampling variance. It was especially designed by OECD to be used with PISA data. It also takes into account the fact that plausible values are used. The average estimator across plausible values is reported and the imputation error is added to the variance estimator.

[20] In practice, when reporting standard errors and estimated coefficients we estimate average values as well as standard deviations on the distribution of the estimated coefficients. Country weights were re-weighted by the number of individuals in order to give the same importance to all countries in our sample.

see Equation 3b, Footnote 15). In this case however, the standard deviation of this measure does not equal the value of one[21].

For our second research question, and when interpreting results for the case of the relationship between school accountability, school autonomy and PISA results; in Equation 4a, the coefficient β_{I2} refers to the additional number of points in PISA for an additional percentage point in school autonomy under a set level of school accountability.

3. Results

Effect of school accountability on educational outcomes

Regression results in Table 4, show that for our initial specification (column 3) there is a positive association between posting results publicly and average achievement for the subject of mathematics, although this effect fades out after including other time-varying institutional characteristics to our model (in particular, proportion of school funding provided by the government. See column 4). This finding holds both when measuring socio-economic status at the individual level and at the school level (detailed in Equations 3a and 3b respectively). On the other hand, regression results assessing the association between accountability and inequality in outcomes according to socio-economic status, show that for both specifications (measuring socio-economic status at the school level or individual level), there is no relationship between accountability and inequality in educational performance. When looking at all countries together (63), we achieve similar results when performing the same analysis for the subject of reading (column 4, Table 6 in Annex A). That is, no relationship between accountability measures and educational outcomes. For the case of science, however, we find a mid-sized positive effect of accountability on educational outcomes, but also increased inequality in performance (Table 7 in Annex A). As explained later, this is mainly driven by a subset of countries (lower- and middle-income).

When producing a separate regression analysis for lower- and middle-income countries[22] (21), we find a positive association between accountability and educational outcomes[23] for the subject of mathematics, but also increased inequality across schools from different socio-economic status for one of our specifications (see columns 5 and 6, Table 4). This result holds only for the specification measuring SES at the school level. In this case, and for the model specified in Equation 3b, one additional percentage point in our measure of school accountability is associated to 0.35 additional points in PISA for the subject of mathematics. On the other hand, for each additional percentage point increase in school accountability (measured at the country level), a school one additional unit higher in the School SES index would achieve on average 0.17 extra points in PISA. In other words, a

[21] In this case it is not possible to directly assess the effect of a one standard deviation increase in this measure on educational outcomes. This will be discussed in the results section. A one standard deviation of the school SES proxy (average ESCS per school) equals 0.46 points in the ESCS index measure for the sample of interest.

[22] Following 2006 World Bank's definition, we assume high-income countries to be those countries showing a GDP above USD 6 400 per capita in that year.

[23] One standard deviation is associated with X additional points in PISA for the subject of mathematics.

school from better off background would better benefit from school accountability in these countries, although the impact of this practice would be modest. This trend holds for the subject of science (0.62 and 0.12 extra points respectively. See Table 6, column 6 in Annex A). For the subject of reading, however, we find no impact of accountability on performance or inequality in outcomes for this group of countries.

For higher-income countries (42) we find no association between accountability and educational performance or equity in education for both mathematics and science. This is the case for all specifications including country fixed effects. For reading, results are overall similar, although we find a weak negative association between accountability and average performance (only significant at 10% level. See Table 6, column 8 in Annex A).

When interacting GDP per capita with both accountability and school SES (see column 9, Table 4) for the same original specifications, we find a decreasing relationship between accountability and educational outcomes in mathematics according the countries' GDP per capita. That is, the lower the GDP, the stronger the association. As expected, this finding holds for the subject of science, but not for reading, where there is no association.

Table 4. The effect of school accountability on educational outcomes (mathematics)

Student Maths Performance

	Across-country	Across-country	Within-country	Within-country	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Interaction with country's GDP per capita in 2000.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual SES									
Results Posted Publicly (country level)	18.16	12.5	***23.35***	17.1	**29.38**	24.57	7.16	7.31	**34.7**
	12.14	24.84	10.31	17.33	13.75	19.76	10.25	11.21	20.09
Results Posted Publicly (country level) x SES	**-28.68**	**-21.46**	4.05	*4.55*	2.72	3.56	5.39	5.39	***-8.3***
	14.54	11.6	6.24	5.8	3.41	3.94	11.68	11.21	2.77
Results Posted Publicly (country level) x GDP per capita									***-1.76***
									0.630
School SES									
Results Posted Publicly (country level)	*13.64*	-5.45	***24.82***	13.2	***35.36***	***35.31***	8.81	5.32	*29.22*
	6.98	3.82	8.94	14.98	14.73	14.05	16.17	11.46	19.72
Results Posted Publicly (country level) x School SES	**-35.46**	*-6.07*	*4.32*	*1.94*	***12.52***	***17.39***	-4.43	-8.10	2.37
	15.75	3.49	3.31	3.6	2.88	5.31	27.55	5.31	3.63
Results Posted Publicly (country level) x GDP per capita									*-0.01*
									0.00
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics		Yes		Yes		Yes		Yes	Yes
Country fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	21	21	42	42	63
Number of observations	1 073 978	1 051 517	1 148 640	1 148 640	353 279	353 279	795 361	795 361	1 148 640
R-Squared	0.27	0.31	0.40	0.40	0.34	0.34	0.30	0.30	0.40

Notes: Dependent variable: PISA maths test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (math). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Does school autonomy actually make a difference, allowing for better results? School autonomy levels and the effect of accountability on educational outcomes.

Table 5 shows the association between accountability practices, school autonomy in curriculum and assessment and educational outcomes in mathematics for different specifications. When looking at figures for all 63 countries under analysis, results indicate that, on average, and under the presence of higher levels of accountability, there is a positive association between school autonomy on curriculum and assessment and educational outcomes, when measuring school autonomy at the country level (see column 4, for specification in Equation 4a). That is, for a fixed level of autonomy in curriculum and assessment, those countries showing higher levels of accountability show better results. On the other hand, and when accountability levels are very low, higher levels of autonomy on curriculum and assessment are associated to worse academic results in mathematics. We find no association between these two variables when measuring autonomy at the school level (Equation 4b).

When performing the same analysis for separate groups of countries according to their levels of income, we find that overall results are driven by low- and middle-income countries (see column 10). For these countries there is a strong association between autonomy, accountability and educational performance in mathematics. According to our results, for a set level of school autonomy on curriculum and assessment, one additional percentage point on school accountability is associated to 1.4 additional points in PISA. Low levels of accountability however, imply a negative association between curriculum and assessment autonomy and educational performance. We found very similar figures for the subjects of reading and science (see column 10 in Tables 9 and 10 in Annex A).

For the case of high-income countries, no association is found for the subjects of mathematics and science. Accountability practices nor school curriculum and assessment practices, separately or jointly, have no association with educational outcomes. For the case of reading we see a very weak association for one of our specifications (only significant at 10% level. See Table 9 column 7 in Annex A).

For all countries and academic subjects analysed in this study, when measuring school autonomy at the school level, we found no association between autonomy, accountability and educational outcomes. This implies that variation in school autonomy within countries does not predict changes on educational outcomes for a set level of school accountability.

Table 8 in Annex A shows that a different pattern can be found for teacher autonomy. In this case we do not find any association between teacher autonomy at the school or country level and mathematics outcomes. Our analyses utilising measures of school autonomy both at the school and country level, for both low- and middle-income countries or high-income countries do not show any significant results. These findings remain the same when estimating the same models incorporating all countries in our analysis.

Table 5. The effect of test-based school accountability on maths performance under school autonomy (curriculum and assessment)

Student Maths Performance

	Within-country	Within-country	Within-country	Within-country	Within-country. High income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COUNTRY LEVEL										
Results Posted Publicly (country level)		***21.09******	***20.7******	5.96	7.92	5.59	5.53	16.02	10.32	-4.92
		8.10	8.0	9.05	9.59	11.18	12.94	12.75	20.53	26.95
Curriculum and Assesment Autonomy (country level)	**-8.12***	-5.91	*-6.91**	***-18.15******	-4.24	-6.71	-6.79	-11.16	***-41.81******	***-60.3******
	3.46	3.67	4.15	6.78	7.37	12.10	14.14	8.85	12.12	20.08
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (country level)			*2.50*	*15.96**		8.48	8.57		***55.64******	***76.04******
			8.7	8.92		18.05	20.13		16.13	3.72
SCHOOL LEVEL										
Results Posted Publicly (country level)		***23.43******	***25.97******	13.68	-7.32	7.24	7.37	18.28	***27.8****	18.22
		8.68	10.9	13.94	7.02	10.31	11.21	12.66	13.58	15.94
Curriculum and Assesment Autonomy (school level)	-1.44	-1.404	-1.4	-1.98	1.09	-0.45	-0.44	-1.93	*-2.1**	-3.16
	0.95	0.95	0.95	1.62	1.19	1.30	1.29	1.22	1.23	2.32
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (school level)			4.20	0.89		-0.28	5.39		2.73	2.18
			5.5	3.06		3.3	11.7		3.4	3.72
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics				Yes			Yes			Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	42	42	42	21	21	21
Number of observations	1 148 640	1 148 640	1 148 640	1 148 640	795 361	795 361	795 361	353 279	353 279	353 279
R-Squared	0.40	0.40	0.40	0.40	0.30	0.30	0.30	0.34	0.34	0.34

Notes: Dependent variable: PISA math test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (math). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Limitations of the study

When interpreting our results, we acknowledge that they could be subject to bias due to several reasons.

First, if our regression models are not correctly accounting for all factors influencing educational achievement, the true relationship between accountability and educational outcomes could remain unidentified. In our models this could be the case, for instance, if there is a strong association between test-based accountability or school autonomy practices and other unobserved time-varying institutional characteristics not included in our models, also related to educational results; such as teacher quality or school leadership. We believe that in most countries these institutional characteristics should remain stable in a short period of time[24], and changes on them should also be uncorrelated to changes in our measures of school accountability. If this was not the case, then our estimates cannot have a causal interpretation. In our models, we included several time-varying institutional characteristics to check for this possibility when available data allows[25]. This approach, however, is also not free of criticism, since regression estimates can become unstable when adding several additional time-varying regressors at the same time in our models. We also accounted for countries' initial levels of performance in PISA as well as initial GDP levels before PISA was implemented in our models.

Second, the low number of countries and time-points in PISA data makes working with measures of institutional characteristics rather problematic. As Bryan and Jenkins (2016[51]) show, the identification of any association utilising few observations is highly affected by outliers and sensitive to any misspecification.

We try to tackle this issue indirectly by bootstrapping standard errors for country level estimates. We also gave all countries the same weight in practice, regardless of their number of examined students[26]. However, it is clear that our estimates from those analyses utilising observations from a low number of countries are more prone to bias. In this regard, our results for separate estimations for low- and middle-income countries as well as for high-income countries need to be taken with caution.

Third, the identification of the associations of interest will be mostly based on those countries showing significant variations in school accountability and/or school autonomy over time (at least for our preferred specifications). Although we identified an important proportion of countries showing statistically significant changes during the period under study, some countries show stable levels, and do not necessarily contribute to identify the associations of interest.

Finally, and although unlikely, changes in accountability at country level could be due to systematic misreporting at one point in time, not reflecting a real change. Also, reverse

[24] The inclusion of country-fixed effects provides an important contribution for tackling the omitted variable bias problem, but it does not account for unobserved time-varying country-level characteristics (possible confounders).

[25] As previously mentioned, we included school operation (property), school funding and school admission policies in some of our specifications.

[26] An alternative approach would be to identify those countries having a more important impact on our final estimates, also reporting how results would change after taking out those countries from the estimation sample. This analysis was not incorporated in the current version of this paper.

causality is possible. For instance, it could be the case that higher school accountability is a product of an improvement in educational achievement and not the opposite.

4. Conclusions and Policy Implications

This study aimed at identifying the impact of test-based accountability practices in educational outcomes, with a special interest on equity in education. Our results suggest that across most OECD countries test-based accountability does not relate to academic achievement, nor has a substantial impact on educational inequality for the subject of mathematics. With some small variations we achieved similar results for the subjects of reading and science.

These findings have important implications for policy makers, as test-based accountability has taken an important role on school reform for high-income countries in previous years. As formerly discussed, the intensive competition across schools that test-based accountability promotes could be disruptive in some educational contexts, producing unintended consequences in school communities. The fact that these reforms do not seem to render the expected results implies important questions about the convenience of these policies.

Despite these results, simultaneously our study indicates that, for low- and middle-income countries, test-based accountability could have a positive average impact on student performance in mathematics and science, although there would be an important trade-off between overall better performance and increased inequality in educational outcomes across schools. These results however must be taken with caution, as they are based on a relatively small number of countries. Moreover, they do not hold when measuring socio-economic status at the individual level. Given the magnitude of the estimated effect size however, it is worth further investigating this trade-off by means of single country case studies which could provide additional information about the suitability of these practices in low- and middle-income countries.

When looking at the relationship between school autonomy and school accountability in high-income countries, accountability does not seem to make any difference, no matter the levels of school autonomy that countries show. Moreover, we do not find any association between school autonomy in curriculum and assessment or teacher management and academic results. This finding contradicts that of recent studies using PISA data for previous applications.

On the other hand, and for the case of lower and middle-income countries, again we found mixed results pointing out that accountability could have a mild positive impact on academic performance in mathematics, reading and science when accompanied by higher levels of school autonomy, particularly on curriculum and assessment practices, such as establishing student assessment policies, choosing which textbooks are used, determining course content, and deciding which courses are offered. These findings however should be taken with caution, as they hold for only one of our specifications and are based on a small number of countries.

From these results it seems to be crucial to study how accountability policies and school autonomy practices could interact. Agent-principal theory would back the idea that higher accountability levels put pressure on school principals and teachers to improve their practices, and that these agents would be able to change their practice only if they have enough autonomy in their respective institutions. On the contrary, low levels of accountability accompanied by high levels of autonomy could be detrimental for school communities. Further investigation on this topic is needed, as school autonomy has increased over time in many countries under different regimes of accountability. Exploiting data from national-level reforms in this regard could add additional evidence about the convenience of these practices.

References

Andrabi, T., J. Das and A. Khwaja (2014), "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.2538624. [27]

Bae, S. (2018), "Redesigning systems of school accountability: A multiple measures approach to accountability and support", *education policy analysis archives*, Vol. 26, p. 8, http://dx.doi.org/10.14507/epaa.26.2920. [13]

Banerjee, A. et al. (2010), "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India", *American Economic Journal: Economic Policy*, Vol. 2/1, pp. 1-30, http://dx.doi.org/10.1257/pol.2.1.1. [29]

Barr, A. et al. (2012), *Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda.*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.6834&rep=rep1&type=pdf. [28]

Benavot et al. (2007), *The Growth of national learning assessments in the world, 1995-2006;*, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000155507 (accessed on 25 September 2020). [4]

Bifulco, R. and H. Ladd (2007), *School choice, racial segregation, and test-score gaps: Evidence from North Carolina's charter school program*, http://dx.doi.org/10.1002/pam.20226. [33]

Bradley, S. and J. Taylor (2002), "The Effect of the Quasi-market on the Efficiency-equity Trade-off in the Secondary School Sector", *Bulletin of Economic Research*, Vol. 54/3, pp. 295-314, http://dx.doi.org/10.1111/1467-8586.00154. [26]

Braga, M., D. Checchi and E. Meschi (2013), "Educational policies in a long-run perspective", *Economic Policy*, Vol. 28/73, pp. 45-100, http://dx.doi.org/10.1111/1468-0327.12002. [38]

Bruns, B., D. Filmer and H. Patrinos (2011), *Making Schools Work: New Evidence on Accountability Reforms*, World Bank, Washington. [1]

Bryan, M. and S. Jenkins (2016), "Multilevel modelling of country effects: A cautionary tale", *European Sociological Review*, Vol. 32/1, pp. 3-22, http://dx.doi.org/10.1093/esr/jcv059. [51]

Burgess, S. et al. (2005), "Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools", *CMPO Working Paper Series*, No. 05/128, CMPO, The University of Bristol, http://www.bris.ac.uk/Depts/CMPO/. [32]

Burgess, S., D. Wilson and J. Worth (2013), "A natural experiment in school accountability: The impact of school performance information on pupil progress", *Journal of Public Economics*, Vol. 106, pp. 57-67, http://dx.doi.org/10.1016/j.jpubeco.2013.06.005. [24]

Burgess, S., D. Wilson and J. Worth (2011), *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting*, The Centre for Market and Public Organization, University of Bristol. [53]

Burns, T. and F. Köster (eds.) (2016), *Governing Education in a Complex World*, Educational Research and Innovation, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264255364-en. [2]

Carnoy, M. and S. Loeb (2002), "Does External Accountability Affect Student Outcomes? A Cross-State Analysis", *Educational Evaluation and Policy Analysis*, Vol. 24/4, pp. 305-331, http://dx.doi.org/10.3102/01623737024004305. [20]

Chapman, D. and C. Snyder (2000), "Can high stakes national testing improve instruction: reexamining conventional wisdom", *International Journal of Educational Development*, Vol. 20/6, pp. 457-474, http://dx.doi.org/10.1016/s0738-0593(00)00020-1. [14]

Chiang, H. (2009), "How accountability pressure on failing schools affects student achievement", *Journal of Public Economics*, Vol. 93/9-10, pp. 1045-1057, http://dx.doi.org/10.1016/j.jpubeco.2009.06.002. [23]

Clotfelter, C. et al. (2004), "Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers?", *Journal of Policy Analysis and Management*, Vol. 23/2, pp. 251-271, http://dx.doi.org/10.1002/pam.20003. [45]

Contreras, D. and T. Rau (2012), "Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile", *Economic Development and Cultural Change*, Vol. 61/1, pp. 219-246, http://dx.doi.org/10.1086/666955. [31]

Davis, T., R. Bhatt and K. Schwarz (2015), "School Segregation in the Era of Accountability", *Social Currents*, Vol. 2/3, pp. 239-259, http://dx.doi.org/10.1177/2329496515589852. [41]

Dee, T. and B. Jacob (2011), "The impact of no Child Left Behind on student achievement", *Journal of Policy Analysis and Management*, Vol. 30/3, pp. 418-446, http://dx.doi.org/10.1002/pam.20586. [22]

European Commission (2015), *Comparative study on quality assurance in EU school education systems. Policies, procedures and practices : final report*, https://op.europa.eu/en/publication-detail/-/publication/1428f97f-b048-4465-8f5b-36e920875ce4. [5]

Figlio, D. and L. Getzler (2002), "Accountability, Ability and Disability: Gaming the System?", in *Advances in Applied Microeconomics, Improving School Accountability*, Emerald (MCB UP), Bingley, http://dx.doi.org/10.1016/s0278-0984(06)14002-x. [46]

Figlio, D. and S. Loeb (2011), "School accountability", in *Handbook of the Economics of Education*, Elsevier B.V., http://dx.doi.org/10.1016/S0169-7218(11)03008-5. [11]

Fitz, J. (2003), "The Politics of Accountability: A Perspective from England and Wales", *Peabody Journal of Education*, Vol. 78/3, pp. 230-241, https://www.jstor.org/stable/1492964?seq=1. [6]

Hanushek, E., S. Link and L. Woessmann (2013), "Does school autonomy make sense everywhere? Panel estimates from PISA", *Journal of Development Economics*, Vol. 104, pp. 212-232, http://dx.doi.org/10.1016/j.jdeveco.2012.08.002. [48]

Hanushek, E. and M. Raymond (2005), "Does school accountability lead to improved student performance?", *Journal of Policy Analysis and Management*, Vol. 24/2, pp. 297-327, http://dx.doi.org/10.1002/pam.20091. [21]

Hanushek, E. and M. Raymond (2004), "The effect of school accountability systems on the level and distribution of student achievement", *Journal of the European Economic Association*, Vol. 2/2-3, pp. 406-415, http://dx.doi.org/10.1162/154247604323068096. [44]

Hastings, J., T. Kane and D. Staiger (2005), *Parental Preferences and School Competition: Evidence from a Public School Choice Program*, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w11805. [9]

Hastings, J. and J. Weinstein (2008), "Information, School Choice, and Academic Achievement: Evidence from Two Experiments*", *Quarterly Journal of Economics*, Vol. 123/4, pp. 1373-1414, http://dx.doi.org/10.1162/qjec.2008.123.4.1373. [34]

Jacob, B. (2005), "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools", *Journal of Public Economics*, Vol. 89/5-6, pp. 761-796, http://dx.doi.org/10.1016/j.jpubeco.2004.08.004. [19]

Jacob, B. and S. Levitt (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating", *The Quarterly Journal of Economics*, Vol. 118/3, pp. 843-877, http://dx.doi.org/10.1162/00335530360698441. [47]

Jennings, J. and D. Rentner (2006), "Ten Big Effects of the No Child Left behind Act on Public Schools", *Phi Delta Kappan*, Vol. 88/2, pp. 110-113, http://dx.doi.org/10.1177/003172170608800206. [42]

Kamens, D. and C. McNeely (2010), "Globalization and the Growth of International Educational Testing and National Assessment", *Comparative Education Review*, Vol. 54/1, pp. 5-25, http://dx.doi.org/10.1086/648471. [16]

Kornhaber, M. (2004), "Appropriate and Inappropriate Forms of Testing, Assessment, and Accountability", *Educational Policy*, Vol. 18/1, pp. 45-70, http://dx.doi.org/10.1177/0895904803260024. [17]

Ladd, H. (2002), "School Vouchers: A Critical View", *Journal of of Economic Perspectives*, Vol. 16/4, pp. 3-24, http://dx.doi.org/10.1257/089533002320950957. [52]

Ladd, H. (2001), "School-Based Educational Accountability Systems: The Promise and the Pitfalls", *National Tax Journal*, Vol. 54/2, pp. 385-400, https://www.jstor.org/stable/41789554. [8]

Levitt, R., B. Janta and K. Wegrich (2008), *Accountability of teachers: Literature review*, RAND Corporation. Santa Monica, CA, https://www.rand.org/pubs/technical_reports/TR606.html. [10]

Mizala, A. and M. Urquiola (2013), "School markets: The impact of information approximating schools' effectiveness", *Journal of Development Economics*, Vol. 103, pp. 313-335, http://dx.doi.org/10.1016/j.jdeveco.2013.03.003. [30]

OECD (2013), *PISA 2012 Results: Excellence through Equity (Volume II): Giving Every Student the Chance to Succeed*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264201132-en. [49]

OECD (2013), *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264201156-en. [18]

OECD (2011), "School Autonomy and Accountability: Are They Related to Student Performance?", *PISA in Focus*, No. 9, OECD Publishing, Paris, https://dx.doi.org/10.1787/5k9h362kcx9w-en. [39]

OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264091559-en. [50]

Phelps, R. (2005), "Trends in Large-Scale Testing Outside the United States", *Educational Measurement: Issues and Practice*, Vol. 19/1, pp. 11-21, http://dx.doi.org/10.1111/j.1745-3992.2000.tb00018.x. [3]

Ravitch, D. (2010), *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*, Basic Books: New York. [40]

Reback, R. (2008), "Teaching to the rating: School accountability and the distribution of student achievement", *Journal of Public Economics*, Vol. 92/5-6, pp. 1394-1415, http://dx.doi.org/10.1016/j.jpubeco.2007.05.003. [35]

Rosenthal, L. (2004), "Do school inspections improve school quality? Ofsted inspections and school examination results in the UK", *Economics of Education Review*, Vol. 23/2, pp. 143-151, http://dx.doi.org/10.1016/s0272-7757(03)00081-5. [25]

Schnellert, L., D. Butler and S. Higginson (2008), "Co-constructors of data, co-constructors of meaning: Teacher professional development in an age of accountability", *Teaching and Teacher Education*, Vol. 24/3, pp. 725-750, http://dx.doi.org/10.1016/j.tate.2007.04.001. [7]

Stanford University/Stanford Center for Opportunity Policy in Education (2015), "Meaningful Learning in a New Paradigm for Educational Accountability: An Introduction", *Education Policy Analysis Archives*, http://dx.doi.org/10.14507/epaa.v23.1982. [12]

Supovitz, J. (2009), "Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform", *Journal of Educational Change*, Vol. 10/2-3, pp. 211-227, http://dx.doi.org/10.1007/s10833-009-9105-2. [43]

UNESCO (2017), *Accountability in education: Meeting our commitments.*, UNESCO Publishing, https://unesdoc.unesco.org/ark:/48223/pf0000259338. [15]

Woessmann, L. et al. (2007), "School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003", *OECD Education Working Papers*, No. 13, OECD Publishing, Paris, https://dx.doi.org/10.1787/246402531617. [36]

Wößmann, L. (2005), "The effect of heterogeneity of central examinations: Evidence from TIMSS, TIMSS-repeat and PISA", *Education Economics*, Vol. 13/2, pp. 143-169, http://dx.doi.org/10.1080/09645290500031165. [37]

Annex A. Tables

Table A A.1. The effect of school accountability on educational outcomes (Reading)

Student Reading Performance

	Across-country	Across-country	Within-country	Within-country	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Interaction with country's GDP per capita in 2000.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual SES									
Results Posted Publicly (country level)	**23.56*** **	30.4	19.36	12.0	22.72	19.75	-14.26	**-15.01***	43.93
	8.04	21.12	22.30	25.97	18.49	22.63	12.32	8.83	30.52
Results Posted Publicly (country level) x SES	-25.36	-20.07	5.19	5.53	5.44	6.08	4.90	4.89	5.7
	23.31	22.8	11.88	12.2	7.32	7.94	15.36	15.41	12.13
Results Posted Publicly (country level) x GDP per capita									**-.002** **
									0.001
School SES									
Results Posted Publicly (country level)	**22.80*** **	3.83	19.2	11.23	**28.91** **	24.38	-13.67	**-14.46***	22.49
	8.20	8.95	18.90	23.50	13.87	15.48	10.39	8.12	22.75
Results Posted Publicly (country level) x School SES	-23.68	*-3.34*	5.8	*3.85*	**12.64***	**10.99***	-1.94	-2.27	4.12
	16.62	2.22	6.74	6.95	7.29	6.33	25.39	24.55	6.81
Results Posted Publicly (country level) x GDP per capita									-0.001
									0.001
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics		Yes		Yes		Yes		Yes	Yes
Country fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	21	21	42	42	63
Number of observations	1 148 639	1 148 639	1 148 639	1 148 639	353 279	353 279	795 360	795 360	1 148 640
R-Squared	0.24	0.34	0.35	0.35	0.29	0.29	0.27	0.27	0.35

Notes: Dependent variable: PISA reading test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (reading). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Table A A.2. The effect of school accountability on educational outcomes (Science)

Student Science Performance

	Across-country	Across-country	Within-country	Within-country	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Interaction with country's GDP per capita in 2000.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual SES									
Results Posted Publicly (country level)	**28.52*** **	3.88	**38.53*** **	27.33**	**62.17*** **	*52.92**	-11.42	-12.23	**67.74*** **
	6.69	8.72	7.82	12.78	11.51	27.62	10.07	12.59	13.02
Results Posted Publicly (country level) x SES	-35.19	-2.32	5.70	6.23	4.79	5.84	7.06	7.05	5.99
	27.49	7.60	11.90	12.3	6.30	10.62	16.09	16.11	7.17
Results Posted Publicly (country level) x GDP per capita									**-.003*** **
									.001
School SES									
Results Posted Publicly (country level)	**27.5*** **	3.75	**38.96*** **	**26.78*** **	**72.31*** **	**62.85*** **	-10.73	-11.52	**67.74*** **
	6.78	8.59	6.18	10.37	15.19	11.25	11.47	13.30	13.03
Results Posted Publicly (country level) x School SES	-32.01	-1.87	7.99	4.99	**16.1** **	*12.64**	0.98	0.66	5.99
	21.20	2.30	7.20	7.53	7.58	6.55	26.55	25.80	7.18
Results Posted Publicly (country level) x GDP per capita									**-.003*** **
									.001
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics		Yes		Yes		Yes		Yes	Yes
Country fixed effects			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	21	21	42	42	63
Number of observations	1 148 639	1 148 639	1 148 639	1 148 639	353 279	353 279	795 360	795 360	1 148 639
R-Squared	0.23	0.37	0.38	0.38	0.33	0.33	0.27	0.28	0.38

Notes: Dependent variable: PISA science test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (science). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Table A A.3. The effect of test-based school accountability on mathematics performance under school autonomy (teacher management)

Student Maths Performance

	Within-country	Within-country	Within-country	Within-country	Within-country. High income countries	Within-country. High income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
COUNTRY LEVEL												
Results Posted Publicly (country level)		***23.47***	***23.46***	***23.6***		8.95	11.51	12.00		**25.13**	*24.45*	21.82
		9.02	8.32	8.08		10.49	12.31	11.93		12.72	12.65	18.47
Teacher Autonomy (country Level)	0.79	0.43	-5.83	2.91	0.267	-0.11	4.03	3.98	-5.23	-4.82	-3.85	-0.15
	2.32	2.25	5.09	2.63	3.81	3.70	6.42	6.58	6.68	7.02	5.90	10.51
Results Posted Publicly (country level) x Teacher Autonomy (country Level)			-5.84	-5.98			-9.81	*-10.34*			*-1.89*	*4.16*
			5.09	4.38			11.17	11.54			7.94	6.93
SCHOOL LEVEL												
Results Posted Publicly (country level)		**21.27**	19.36	12.05		16.42	16.18	15.92		***30.00***	***29.58***	17.86
		9.14	12.6	13.86		24.93	24.51	24.52		13.38	13.36	15.79
Teacher Autonomy (school Level)	0.69	0.66	-.094	-0.23	1.54	1.51	0.82	0.84	-0.07	-0.10	0.61	0.49
	0.88	0.88	2.47	1.49	1.19	1.21	2.43	2.40	0.82	0.82	1.56	1.51
Results Posted Publicly (country level) x Teacher Autonomy (school Level)			1.60	1.67			1.12	1.09			-1.51	-1.22
			3.6	2.26			3.72	3.69			2.74	2.69
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics				Yes				Yes				Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	41	41	41	41	22	22	22	22
Number of observations	1 148 640	1 148 640	1 148 640	1 148 640	795 361	795 361	795 361	795 361	353 279	353 279	353 279	353 279
R-Squared	0.38	0.40	0.40	0.4	0.30	0.30	0.30	0.30	0.34	0.34	0.29	0.34

Notes: Dependent variable: PISA math test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (math). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Table A A.4. The effect of test-based school accountability on reading performance under school autonomy (curriculum management)

Student Reading Performance

	Within-country	Within-country	Within-country	Within-country	Within-country. High income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COUNTRY LEVEL										
Results Posted Publicly (country level)		18.6	15.0	3.70	-13.75	-17.25	**-18.64****	28.63	**22.7****	13.79
		18.69	15.8	15.3	12.41	13.65	9.51	23.80	11.4	14.25
Curriculum and Assesment Autonomy (country level)	0.79	1.96	**-5.86****	***-11.27*****	-0.28	-2.80	-4.03	7.90	***-20.61*****	***-27.35*****
	2.10	3.22	2.83	3.08	3.67	3.52	3.01	7.94	10.48	5.95
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (country level)			***19.5*****	***25.94*****		8.68	*9.93**		***51.76*****	***59.21*****
			4.3	4.41		5.27	5.14		11.38	8.67
SCHOOL LEVEL										
Results Posted Publicly (country level)		16.7	16.6	8.98	-13.63	-13.54	-14.30	15.97	15.98	12.38
		16.13	16.0	19.51	13.36	12.77	9.17	12.50	12.4	15.09
Curriculum and Assesment Autonomy (school level)	-1.37*	-1.35*	-1.6	-1.83	-0.76	-0.59	-0.62	-1.58	-2.68	-2.85
	0.78	0.79	1.47	1.48	1.11	2.23	2.38	1.18	2.06	2.07
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (school level)			0.43	0.77		-0.32	-0.28		2.44	2.7
			2.9	2.88		4.7	4.9		3.76	3.82
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics				Yes			Yes			Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	41	41	41	22	22	22
Number of observations	1 148 639	1 148 639	1 148 639	1 148 639	795 360	795 360	795 360	353 279	353 279	353 279
R-Squared	0.35	0.35	0.35	0.35	0.27	0.27	0.27	0.29	0.29	0.29

Notes: Dependent variable: PISA reading test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (reading). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.

Table A A.5. The effect of test-based school accountability on science performance under school autonomy (curriculum management)

Student Science Performance

	Within-country	Within-country	Within-country	Within-country	Within-country. High income countries	Within-country. High income countries	Within-country. High income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries	Within-country. Lower and Middle income countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COUNTRY LEVEL										
Results Posted Publicly (country level)		**34.69*** **	**34.50*** **	**17.57*** **	-10.77	-14.04*	**-15.31** **	**52.76*** **	**49.29** **	32.76
		6.02	6.10	6.03	8.29	8.39	12.41	12.34	24.38	31.04
Curriculum and Assesment Autonomy (country level)	*-3.51**	-1.33	-1.78	*-9.72**	2.56	0.20	-0.92	-2.95	**-19.64*** **	**-32.18*** **
	1.94	2.31	2.51	5.19	1.99	3.40	6.73	5.90	10.32	9.61
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (country level)			1.11	*10.62**		8.12	9.27		***30.31*** **	**44.14*** **
			4.8	5.74		6.80	10.18		12.8	8.90
SCHOOL LEVEL										
Results Posted Publicly (country level)		**35.3*** **	**35.36*** **	**23.77*** **	-10.57	-10.75	-11.58	**55.74*** **	**55.74*** **	**48.73*** **
		6.00	6.1	7.69	9.11	9.85	12.29	14.50	14.43	11.07
Curriculum and Assesment Autonomy (school Level)	*-1.46**	*-1.41**	-1.24	-1.66	-0.75	-1.09	-1.13	-1.73	-1.69	-2.02
	0.82	0.81	1.67	1.66	1.08	2.87	3.00	1.06	2.06	2.09
Results Posted Publicly (country level) x Curriculum and Assesment Autonomy (school level)			-0.32	0.20		0.64	0.68		-0.08	0.41
			2.9	2.86		5.2	5.3		3.26	3.31
Student and family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other country characteristics				Yes			Yes			Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015	2006-2015
Number of countries	63	63	63	63	42	42	42	21	21	21
Number of observations	1 148 639	1 148 639	1 148 639	1 148 639	795 360	795 360	795 360	353 279	353 279	353 279
R-Squared	0.38	0.38	0.38	0.38	0.28	0.28	0.28	0.33	0.33	0.33

Notes: Dependent variable: PISA science test plausible values. Coefficients in bold and italics significant at 1% level (***), coefficients only in bold significant at 5% level (**), coefficients only in italics significant at 10% level (*). Standard errors clustered at country level. Observation weights re-estimated so each country contributes with the same weight to the least squares regression final estimates. Student and family characteristics: age, gender, immigration status, language at home and ESCS score. School characteristics: Student/teacher ratio, school size, school location, school property (private or public), school funding (pctg. government), school average ESCS. Other country characteristics: school funding (average pctg. of government funding that year), country's initial average score in PISA (science). Countries with GDP per capita over 10 000 USD in 2000 classified as high income countries.