ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-741

How Context Shapes the Relationship between School Autonomy and Test-Scores: An Explanatory Analysis using PISA 2015

Christian Buerger
Indiana University-Purdue University Indianapolis

Jane Arnold Lincove
University of Maryland, Baltimore

Catherine Mata
Brown University

School autonomy has been and continues to be one of the most important education reform strategies around the world despite ambiguity about its theoretical and empirical effects on students learning. We use international data from PISA to test three country-level factors that might account for inconsistent results in prior literature: (1) the selective implementation of school autonomy based on school performance; (2) differential influence on high-risk subgroups; and (3) the presence of accountability policies to prevent opportunism by autonomous schools. We find that the relationship between autonomy and student test performance varies both across countries and within countries across subgroups in both magnitude and direction. Similar results are observed if decentralization is coupled with accountability policies. All of three tested factors influence country-level associations between school decentralization and student learning, which suggests that autonomy is effective only when contextual factors and other policies are aligned.

VERSION: March 2023

Suggested citation: Buerger, Christian, Jane Arnold Lincove, and Catherine Mata. (2023). How Context Shapes the Relationship between School Autonomy and Test-Scores: An Explanatory Analysis using PISA 2015. (EdWorkingPaper: 23-741). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/jjbs-t526

How Context Shapes the Relationship between School Autonomy and Test-Scores: An Explanatory Analysis using PISA 2015

Christian Buerger, Jane Arnold Lincove, and Catherine Mata*

February 28, 2023

Abstract

School autonomy has been and continues to be one of the most important education reform strategies around the world despite ambiguity about its theoretical and empirical effects on students learning. We use international data from PISA to test three country-level factors that might account for inconsistent results in prior literature: (1) the selective implementation of school autonomy based on school performance; (2) differential influence on high-risk subgroups; and (3) the presence of accountability policies to prevent opportunism by autonomous schools. We find that the relationship between autonomy and student test performance varies both across countries and within countries across subgroups in both magnitude and direction. Similar results are observed if decentralization is coupled with accountability policies. All of three tested factors influence country-level associations between school decentralization and student learning, which suggests that autonomy is effective only when contextual factors and other policies are aligned.

JEL Classification: I21, I24, I28
Key words: school autonomy, school accountability, student achievement gaps

* **Buerger**: O'Neill School of Public and Environmental Affairs, Indiana University-Purdue University Indianapolis, 801 West Michigan Street, 6202 Indianapolis, Indiana (email: cbuerge@iupui.edu). **Lincove**: School of Public Policy, University of Maryland, Baltimore County (email: jlincove@umbc.edu). **Mata**: Annenberg Institute, Brown University (email: catherine_mata@brown.edu)

1. Introduction

Over several decades, decentralization and school autonomy have emerged as one of the most common elements of education reform (Faguet 2014; Faguet and Sánchez 2008; Neeleman 2019; Suggett 2015). By the end of the 1990s, 80 percent of developing and transitioning countries of the former Eastern Bloc and Soviet Union experimented with forms of school decentralization (Manor 1999). In the following years, OECD and the World Bank promoted school autonomy around the globe, often funding decentralization efforts (Lincove 2006) and giving practical advice about how to create autonomous schools (OECD 2011, 2013, 2016; Arcia, Macdonald, Patrinos, and Porta 2011). School decentralization remains part of the toolkit advocated by school reformers as a strategy to improve student performance (Bulkley, Henig, and Levin 2010; Lincove and Bulkley 2020; Yuki and Igei 2020).

Despite numerous and continued efforts, both the conceptual and empirical foundation for relationships between school autonomy and student performance remain ambiguous. Supporters argue that local decision makers know the capacity of their school and the needs of their students better than a centralized authority and thus, make better decisions regarding the use of resources (Barankay and Lockwood 2007; Oates 1972, 1999; Ouchi 2003; Hoxby 1999). Decentralized schools are framed as subject to less red tape than traditional schools, allowing them to focus more on student learning (Burian-Fitzgerald, Luekens, and Strizek 2004). Decentralization is also hypothesized to increase parent involvement, which is expected to increase both student motivation and school oversight (Ammermüller 2005; Gertler, Patrinos and Rubio 2006; Jimenez and Sawada 1999; Schuetz, West, and Woessmann 2007).

There are also numerous theoretical arguments against the superiority of decentralized over centralized education systems. Among these arguments are the potential lack of decision-

making capacity at the local level (Bardhan and Mookherjee 2005; Galiani, Gertler, and Schargrodsky 2008), the loss of economies of scale when formerly centralized tasks must be performed by autonomous schools (Buerger and Harris 2020; Levin 2012), vulnerability to local corruption (Fan, Lin, and Treisman 2009; Oates 2005), and the lack of oversight resulting in reduced incentives and pressure to provide optimal schooling services (Ingersoll 2009; Siddle and Koelble 2012). There are also concerns regarding equity in communities where local power differentials might amplify existing inequalities and undermine broader goals of inclusion and integration (Bardhan and Mookherjee 2005).

Substantial empirical literature testing the effects of decentralization does little to resolve these theoretical conflicts. Findings of both positive and negative relationships between school autonomy and student performance are common in the literature, often with little attention to underlying sources of conflicting results (see for overviews Faguet and Sánchez 2008; Jeong, Lee and Cho 2017; Kameshwara et al. 2020). Taken as a whole, the current body of evidence reveals that the impacts of decentralization vary substantially across contexts, settings, and even individual students for reasons that have yet to be explained.

In this study, we move beyond asking whether decentralization is helpful or harmful to understand how and why the empirical findings are inconsistent. We do this by testing whether results vary due to variation in the context of implementation or variation in the design of policies granting school autonomy. We test several theories of why associations vary across systems and students. The first theory relates to the selective implementation of decentralization across schools within a system and whether positive or negative selection of schools into autonomy explains observed differences in student performance. The second theory relates to differential effects of school autonomy for students from within and outside dominant subgroups.

Here we ask whether autonomy is differentially associated with achievement for girls and boys and students with varying socio-economic background. Finally, we examine differences in policy design, testing for differences based on whether schools have autonomy over different areas of decision-making and whether school autonomy is coupled with school accountability measures.

To investigate these questions across a broad range of contexts, we use data from the Programme for International Student Assessment (PISA).[1] We first estimate global and country-level associations between attending an autonomous school and students' PISA 2015 math scores. We find estimates close to null globally but both positive and negative coefficients for individual countries. Within countries, we also find that autonomy has both positive and negative affiliations with achievement gaps for some disadvantaged subgroups. This suggests that estimates based on international comparisons mask heterogeneity across contexts. In our analyses, we differentiate between several types of autonomy that might be granted to schools (human resources, budgeting, course content, and admissions) and compare achievement gaps for multiple at-risk subgroups (girls, students in poverty, and students with less educated parents).

Analyzing the effects of autonomy, we find that autonomy is often associated with student and school characteristics. Both positive and negative regression coefficients for autonomy indicators are mostly made insignificant when we add controls for individual and aggregate student demographics, indicating that both negative and positive selection is present in the PISA countries. Similarly, we find that school decentralization can be related to both

[1] Our preferred specifications include the following countries. For East Asia and the Pacific Region: Indonesia, South Korea, Thailand, and Vietnam. For Europe and Central Asia: Belgium, Croatia, Denmark, Finland, France, Germany, Hungary, Ireland, Kosovo, Moldova, North Macedonia, Norway, Portugal, Romania, Slovenia, Spain, and Switzerland. For Latin America and the Caribbean: Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Mexico, Peru, and Uruguay. For the Middle East and North Africa: Lebanon, Qatar, and United Arab Emirates. For North America: Canada and the United States. See also Appendix A1 for more information.

increasing and decreasing achievement gaps, and the direction of this relationship depends on both the form of autonomy granted to schools and student characteristics. Implementing the same form of autonomy (e.g. budget autonomy) may decrease the achievement gap for one group of students, while increasing it for another. We also find no consistent evidence regarding coupling autonomy with accountability. Internationally, there is no association between measures of autonomy and accountability on average, but country analyses reveal both positive and negative relationships between autonomy and accountability.

These findings provide important insights into the literature on school autonomy. First, our results neither confirm nor refute the theoretical arguments of decentralization proponents and opponents but show that the relationship between school autonomy and student achievement is complex and varied. Student and school characteristics, existing achievement gaps, and accountability policies are factors that influence the relationship between school autonomy and student learning. Second, our research gives insight into why research on autonomy and test scores shows such variation between studies. Not controlling for factors that influence the implementation of autonomy, neglecting the interaction between decentralization and existing achievement gaps, and insufficiently controlling for complementary accountability policies might have biased estimates in earlier studies. Additionally, methods that average estimates across contexts likely miss important nuances in the relationship between decentralization and test scores.

2. Literature Review

School autonomy refers broadly to the authority school managers have over planning, operations, and policies at their school. Autonomy is typically determined by the degree of decentralization in a national school system, and thus autonomy is granted by education

authorities to school managers (Parry 1997). More comprehensive definitions involve the freedom to make decisions about the inner and outer relationships of a school (Wohlstetter and Chau 2004; Wohlstetter, Wenning, and Briggs 1995). These definitions almost always involve the following three criteria: administrative decisions, resource allocation, and educational framework (Bulkley and Fisler 2003; Finnigan 2007; Gawlik 2007; Hanushek, Link, and Woessmann 2013; OECD 2011).

Administrative decisions are concerned with the way different actors such as principals, teachers, parents, students, etc. interact with each other. They also include all choices with regard to hiring or firing of staff. Resource allocation, which has been highlighted by some researchers as the most important criterion of autonomy (e.g. Ouchi 2004), incorporates control over revenues and expenditures in a school's budget, but in some settings, it might also include responsibility for raising revenue (Parry 1997). The educational framework involves all questions regarding the curriculum and instructional techniques a school uses.

What is generically called school autonomy can describe control over some or all of these areas (Ainley and McKenzie 2000; Bulkley and Fisler 2003; Suggett 2015). In this study, we are able to differentiate four types of autonomy through school-level reports of schools' control over human resources, budgeting, course content, and student admissions.

Empirical evidence on the relationship between autonomy and student performance is mixed and inconclusive (Faguet and Sánchez 2008; Jeong, Lee and Cho 2017). Using cross-country data from PISA, two studies find that school autonomy and student performance are positively correlated (Diaz-Serrano and Meix-Llop 2012; Woessmann 2003). This result is replicated by other researchers for individual countries such as Norway (Naper 2010), Switzerland (Barankay and Lockwood 2007), Argentina (Galiani and Schargrodsky 2002), and

Colombia (Faguet and Sanchez 2008). Other studies provide evidence that increased autonomy does not lead to improved student learning. Prawda (1993) analyzes the impact of decentralization on the efficiency and quality of schools in Argentina, Chile, Colombia, and Mexico and does not find a relationship between autonomy and student performance. Focusing on single countries, Parry (1997) suggests that decentralization efforts had a mixed impact on education quality in Chile, while Merrouche (2007) does not find a positive relationship between school autonomy and literacy rates or educational attainment in Spain.

Several factors potentially explain the variation in empirical findings. First, samples, data, and methods differ widely between studies (Faguet and Sánchez 2008; Jeong, Lee and Cho 2017). If the nature of school autonomy policies and practices vary across settings, research findings might reflect local variation in implementation. In this case, ambiguous results suggest that autonomy must be implemented in specific ways to achieve positive results. Second, comparisons of schools with and without autonomy might be subject to selection bias if education authorities grant autonomy to schools based on performance. Third, if autonomy benefits groups in power more than underserved groups, differential effects might be explained by the distribution of student subgroups across schools (Lincove 2006). Finally, institutional settings such as accountability rules also differ between countries and might interact with autonomy in ways that could either increase or decrease student performance (Fuchs and Woessmann 2007; Levin 1974, 2012; OECD 2013; Waslander et al. 2020).

Differential Implementation of Autonomy

The way higher-level authorities grant autonomy to schools varies between and even within countries (Daramola and Mulfinger 2020; Parry 1997). In some cases, authorities decentralize schools universally. This approach embodies the theory that decentralization helps

educators to allocate resources more efficiently and effectively. In some settings, autonomy has to be earned with a track record of high performance. This strategy responds to concerns that managers of schools with low student achievement lack skills to improve, or that educators could use their additional freedom in ways that harm student learning (Hanushek, Link, and Woessmann, 2013; World Bank 2004). Conversely, schools might be decentralized as a response to poor performance, based on the theory that top-down administrative control hinders improvement. The latter is a common justification for the privatization of public schools in low-performing urban school systems in the US through charter school contracts (Bulkley 2005; Chubb and Moe 1990; Stoddard and Corcoran 2007). For our empirical analysis, it is important to note that in some settings the implementation and retention of autonomy is positively correlated with achievement, while in others the association is negative. Differential selection into autonomy would likely lead to varied observations of the association between autonomy and student performance across settings, as well as variation in the direction of bias in estimates that ignore selection.

Differential Effects of Autonomy on Student Subgroups

How decentralization interacts with already existing achievement gaps is another likely reason for the varying results in the literature, as school autonomy could either enhance or diminish these gaps. Decentralization negatively influences achievement gaps if protections for vulnerable students are reduced or reversed, or local elites gain control over resource allocations. A competing hypothesis suggests that autonomous schools with high enrollment from underserved subgroups will allocate resources in ways that reduce existing achievement gaps. Which one occurs in practice is likely a complex result of the design of autonomy policies and incentives that are generated for managers.

In this study, we focus on three subgroup achievement gaps that are reported in the literature as frequent sources of educational inequality. Our first achievement gap is based on gender. Gender gaps in academic achievement result from stereotype threats (Radulović et al. 2022; Reardon et al. 2019; Schippers, Scheepers, and Peterson 2015), gendered participation in the labor force (Sundaram and Vanneman 2008), and gender-based variation in returns to education (Becker 1964). As a result, in countries like Bolivia, China, Ethiopia, India, Mexico, Peru, Senegal, Sierra Leone, and Paraguay boys are more likely to go to school and achieve higher than girls, while in other settings, as for instance in Vietnam, the reverse is true (Azam and Kingdon 2013; Dercon and Singh 2013; Masterson 2012; Yao and You 2018). Decentralization could increase existing gender based achievement gaps if, for instance, teachers and administrators use their additional discretion to reduce the already sparse educational resources for girls even further. On the other side, it is also possible that school personnel employs autonomy to overcome existing disparities in resources between male and female students and thus decreases prior achievement gaps. To find out in which of these two pathways prevails is one of the main goals in our empirical analysis.

The second and third achievement gaps are based on socio-economic differences between parents such as wealth and education, respectively. Both gaps follow similar theoretical predictions, which suggest, as a first mechanism, that parents with higher socio-economic status are likely to invest more into the human capital of their children relative to parents with lower a socio-economic background (Becker 1964). A second mechanism implies that more affluent parents give their children a better understanding of the dominant culture and an ability to succeed within it than more impoverished mothers and fathers (Bourdieu 1977, 1984). In line with these theoretical predictions, Bouhlila (2017) finds a positive association between parents'

socio-economic status and students' achievement for Middle Eastern and North African countries. Similarly, Kafle, Jolliffe, and Winter-Nelson (2018) show the importance of household assets for student performance in Tanzania. In a comprehensive review of the literature on the topic, Huisman and Smits (2009) provide substantial evidence of a strong correlation between educational outcomes on one side and parental income and education on the other. The authors further find, in their own empirical analysis, a strong relationship between parent characteristics and student achievement for 30 developing countries. Similar to the achievement gap based on gender, additional school autonomy could channel resources for students of different socio-economic status in ways that either decrease or increase existing gaps in achievement.

Interaction between Autonomy and Test-Based Accountability

Test-based accountability policies seek to make school managers directly accountable for student performance. The theoretical framework implicitly underlying test-based accountability policies is a principal-agent-model where the incentives of the agent (teachers and administrators) are imperfectly aligned with the goals of the principals (e.g. parents, taxpayers, and education authorities). Because it is costly for the principal to directly oversee the agent, accountability systems seek to align the agent's incentives with the goals of the principal by publicly reporting student achievement, attaching rewards to high achievement, and attaching sanctions to low achievement (Figlio and Loeb 2011; West and Woessmann 2012; Woessmann 2007). Without accountability, there is a threat that autonomous school managers will neglect or even harm students by directing resources towards personal gain instead of collective educational goals.

Since the 1980s, international organizations such as the OECD and World Bank started to promote the simultaneous implementation of autonomy and accountability policies, based on the rationale that accountability policies create incentives for schools to use their additional autonomy to elevate student learning (Bollen 1996; Demas and Arcia 2015; Hopkins and Lagerweij 1996; Hopkins and Reynolds 2001; OECD 2011). Both organizations supported their recommendation with results from comparative studies suggesting that countries with accountability and autonomy policies have greater test-scores relative to countries implementing only autonomy policies (see Demas and Arcia 2015 for a comprehensive literature review). There are several studies, however, suggesting that accountability policies in practice can harm student learning and undermine performance gains from decentralization efforts. The negative effects of accountability policies include cheating on standardized tests (Jakob and Levitt 2003), strategically excluding some students from testing (Figlio 2006), narrowing of curriculum and resources to tested subjects (Koretz and Barron 1998), greater difficulty to retain teachers in low-performing schools (Clotfelder et al. 2004), pre-emptive grade retention of students and placing students in programs outside accountability systems (Jacob 2005), and a focus on students close to accountability thresholds (Springer 2007). To test whether accountability mediates the relationship between autonomy and test scores, we interact autonomy and accountability policies in our empirical models to investigate if the presence of both policies further increases or decreases student performance.

3. Empirical Methods

Our empirical goal is to estimate global and country-level relationships between school autonomy and student performance while testing the potential role of selection bias, the distribution of student subgroups, and the implementation of accountability policies. To illustrate the importance of variation across settings, we first run aggregate models for all PISA countries

and then separately for each country in our sample. In alignment with these goals, our empirical analysis focuses on the association between the variables of interest, but not on how one variable causes a change in the other.

Our first research question asks whether selective implementation of autonomy explains why some settings see a positive relationship between autonomy and learning while others do not. Here we compare autonomy coefficients from models with and without controls in a pooled sample of students from PISA countries. First, we estimate the following equation:

$$Y_{isc} = \alpha + \beta \cdot Autonomy_s + \varepsilon_{isc} \quad (1)$$

where Y_{isc} are test scores for student i in school s in country c. The variable of interest is $Autonomy$, which indicates a school's local decision-making authority. We test four constructs of autonomy: hiring, budget, curricular content, and admission. We also test models that include all four autonomy indicators. These variables are explained in more detail in the data section. In the next step, we add vectors of student and school controls and country fixed effect ($\boldsymbol{\pi_c}$) to Equation (1). The country fixed effect allows us to control for unobserved national education policy and context and to isolate differences in autonomous and non-autonomous schools operating in the same county. We estimate:

$$Y_{isc} = \alpha + \beta \cdot Autonomy_s + \gamma \cdot Student_i + \delta \cdot School_s + \pi_c + \varepsilon_{isc} \quad (2)$$

By comparing the size and significance of β's from equation (1) and equation (2), we discern if the association between student performance and school autonomy is driven by student and school characteristics, autonomy, or performance. Three outcomes of this analysis are possible. First, coefficients are positive or negative in equation (1), but when student and school controls are added in equation (2), estimates of autonomy (β) are null. In this case, decentralization is associated with higher or lower performance but likely due to selection bias

on student and school characteristics rather than direct interrelations. Second, the addition of controls might change the direction or magnitude of β. Here, autonomy is associated with student characteristics, but also has a unique association with achievement. In these first two instances, the omission of control variables leads to biased estimates of autonomy. The final outcome occurs if the β's in equation (1) are robust to the addition of controls in equation (2). This finding would suggest that decentralization is associated with students' achievement in a way that is independent of the characteristics of students in autonomous schools.

When equation (1) and (2) are run for all observations in our data pooled across countries, we potentially mask important differences between countries. We next estimate these equations (1) and (2) separately by country, estimating unique β's for each setting. This comparison at the country-level allows us to observe first, whether autonomy has a predictable association with student performances internationally or whether there are different coefficients in different contexts, and second, whether the differences in the strategy of selective implementation of autonomy explain these potential differences.

For our second research question, we model the relationship between autonomy and test scores for high-need student subgroups in pooled regressions as:

$$Y_{isc} = \alpha + \beta_1 \cdot Autonomy_s + \beta_2 \cdot Subgroup_i + \beta_3 \cdot Autonomy_s \cdot Subgroup_i + \delta \cdot School_s + \pi_c + \varepsilon_{isc} \quad (3)$$

where $Subgroup_i$ is a dummy variable indicating whether the student belongs to a disadvantaged subgroup (described in the next section). We exclude additional individual control variables from equation (3) to avoid any multicollinearity problems between the $Subgroup_i$ variables and other student characteristics. As in equation (2), we include school-level covariates and country fixed effects.

Our analysis focuses on β_2, which measures the relationship between subgroup membership and achievement for schools without autonomy, and β_3, which estimates how autonomy influences this relationship. A negative and significant β_2 indicates that an achievement gap exists where the average subgroup member has lower average performance than other students. Added to a negative β_2**,** a negative and significant β_3 indicates that autonomy increases this performance gap, while a positive β_3 indicates that autonomy decreases the gap. As above, we estimate equation (3) both pooled for all countries and separately by country. The latter allows us to observe if the affiliation between autonomy and inequities is similar or different across countries.

Our third research question asks whether differential relationships of autonomy on subgroups are explained by the presence or absence of accountability measures. Here we restrict our sample to schools with autonomy, and test whether accountability policies influence achievement gaps by estimating:

$$Y_{isc} = \alpha + \beta_1 \cdot Accountable_s + \beta_2 \cdot Subgroup_i + \beta_3 \cdot Accountable_s \cdot Subgroup_i + \delta \cdot School_s + \pi_c + \varepsilon_{isc} \quad (4)$$

where $Accountable$ is a dummy variable equal to 1 if schools are required to make achievement data available to the public. The coefficients of interest are β_2 and β_3, which measure the relationship between test scores for different subgroups in autonomous schools without and with accountability, respectively. The interpretation of these estimates is similar to the interactions in the previous equation. Once more, we estimate equation (4) including all countries in the analysis with country fixed effects, and then we run models for each country separately to discern if accountability coupled with autonomy has common associations across countries.

As recommended by the survey designers, we estimate standard errors (ε) using the Balanced Repeated Replication (BRR) with Fay's adjustment as variance method.[2]

4. Data and Sample

For these analyses, we use student-level data from the 2015 wave of the Programme for International Student Assessment (PISA), conducted by the Organisation for Economic Co-Operation and Development (OECD). PISA 2015 is the most recent study available including questions regarding school autonomy and accountability. The data comprises information from 73 countries of East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, the Middle East and North Africa, and North America. Our analysis focuses on 34 countries from across all these five regions with sufficient data for the analysis of autonomy. We excluded from the analysis countries where the information was representative only for parts of the country, the autonomy data were missing, or there was not enough within-country variation in autonomy to conduct the analysis.[3] To ensure sufficient within-country variation in schools' autonomy, we include only countries in which at least 20 percent but no more than 95 percent of schools have hiring autonomy. We must also exclude student and school observations with missing values for any critical variables in the models.[4] This results in the loss of up to 18 percent of schools and 22 percent of all student observations across the 34 included countries. Our final sample comprises a total 7,524 schools and approximately 205,000 students. Robustness checks, discussed below, tested alternative rules for inclusion.

[2] For more information on PISA replication methods, see PISA (2009) and Stata's *Repest* help manual.
[3] The full list of countries included in our sample is presented in Table A1 of the Appendix.
[4] About five percent of the schools in our sample did not respond the survey questions on autonomy. This percentage varies across countries in a range that goes from 0.4 percent in one country to 25 percent of schools in another country.

Our dependent variable measures student performance using the first plausible value on the PISA math exam.[5] Prior research suggests that math test scores can be most easily compared across countries (Hanushek, Link and Woessmann 2013). Observations are weighted according to their probability of selection to reflect the population of 15-year-old students in each country.[6] Decisions regarding plausible values and weighting were also tested in robustness checks, described below.

Autonomy and Accountability Indicators

PISA provides common measures of autonomy and accountability based on school principals' responses to a set of questions about school management and policy. Thus, our measures are based on school manager reports of their own authority and accountability.[7] For autonomy, we constructed four dichotomous autonomy indicators focused on: 1) hiring decisions, 2) budget formulation, 3) curricular content, and 4) student admissions. We identify a school as autonomous in a category if the school leader indicated that school managers and/or teachers have control over decision-making in that area. We identify a school as accountable if the principal reported that the school's achievement data is made available to the public. This variable permits us to measure within country variation and includes only few missing values for

[5] Plausible values can be described as a representation of the range of abilities that a student might reasonably have. This is, instead of obtaining a point estimate of the student's ability, a range of possible values for the student's ability is estimated with an associated probability for each of these values (see OECD, 2009 for more details). PISA 2015 reports ten plausible values for each student tested, reporting a range of student's abilities. The probability associated to each of these plausible values is built to prevent biased inferences occurring as a result of measuring an unobservable underlying ability through a test that uses a relatively small number of items. Even though using only one plausible value, instead of the 10 reported by PISA, could prevent estimating the imputation error that reflects the influence of test unreliability for the parameter estimation, the estimates of population parameters are still unbiased in large datasets like ours where this imputation error tends to be relatively small (see OECD, 2009). Our results are robust to the use of any of the ten plausible values.

[6] See PISA Data Analysis Manual 2015 for more information on test scores and weights.

[7] Note that our autonomy and accountability variables, as all self-reported measures, allow for some interpretation by the survey taker.

the countries in our analysis.[8] Our results are robust to alternative definitions of autonomy where the school leader and teachers must have full control that is not shared with a higher authority.

Figure 1 illustrates the frequency of autonomy with and without accountability for the 34 countries in our primary sample, and Table 1 provides summary statistics for the 7,524 participating schools in our analytic sample. All countries in the sample include some autonomous schools of each type and, in nearly all cases, countries include autonomous schools both with and without accountability. Internationally, 40% of schools in the sample are accountable through publicly available achievement data. Figure 1 reveals broad variation both across and within countries in the presence of different types of autonomy and the complementary use of accountability measures. Table 1 shows that the most common autonomy type is autonomy over course content, which is reported at 70% of schools, followed by student admission at 58% of schools, hiring decisions at 56% of schools, and budget formulation at 48% of schools. Although most schools have at least one type of autonomy, only 24% of schools report having all four types of autonomy. Table 1 also reports school-level correlations between the four autonomy types with values between 0.23 and 0.33. This suggests a wide international and intranational variation in the scope and types of authority that are decentralized to schools.

Student Subgroups

Our research objective includes testing whether autonomy increases or decreases performance gaps for disadvantaged students. From prior literature, we identify three student subgroups where achievement gaps are a concern in many countries and where school autonomy might have either positive or negative associations with performance. Specifically, we investigate achievement gaps based on gender, family resources, and parent education. The

[8] Note that accountability measures used in other studies, such as exit exams, often do not vary within most countries. We also found large numbers of missing values for the PISA school survey question on exit exams.

gender gap is measured by comparing female students to males. The family resources gap is measured by the PISA wealth index[9]. For each country, we divide students in wealth quintiles and compare students in the lowest quintile to students in the top four quintiles. The parent education gap is measured by comparing children of mothers who never attended college to those with college-educated mothers.

Control Variables

Our final sets of variables include student- and school-level control variables to test selective sorting of schools into autonomy based on the relative advantage or disadvantage of the student population. At the school level, we control for school size (student enrollment), the share of students at the bottom wealth quintile, the share having mothers without a college degree, whether the school is publicly funded, and if the school is located in a rural area. Student-level controls include gender, age, nativity, mother's education, and families' wealth level. Table 2 provides summary statistics for student and school-level covariates in the pooled sample of 34 countries.

Note that in selecting control variables we face a tradeoff between a parsimonious and more saturated model. We decided for the parsimonious model, as adding a large number of controls increases the number of observations with missing values in at least one school or student characteristic. We acknowledge that this strategy omits some variables from our models, but emphasize, once more, that our goal is to provide evidence for the selective implementation of school autonomy, which can be shown with only few controls that have been conceptually

[9] In PISA, five indices are derived from students reports of the availability of 16 household items at home including three country-specific household items that were seen as appropriate measures of family wealth within the country's context and the number of books at home. The WEALTH index is constructed to reflect a family wealth possession and is based on the availability at home of an own room for the student, a link to the internet, televisions, cars, rooms with bath or shower, cellphones with internet access, computers, tablets, E-book readers, and three country-specific wealth items (OECD, 2017).

linked to student achievement in other studies (e.g. Woessmann et al. 2009; Hanushek, Link, and Woessmann 2013). Nonetheless, we provide several sensitivity checks demonstrating that our main results are robust to changes in the control variables (see Table 3).

5. Results

Autonomy and Selection Bias

We begin by estimating the relationship between four types of school autonomy and student performance while ignoring student, school, and country context. Table 3 displays the results for pooled regressions that include all students with complete data in all 34 included countries. Columns 1-4 include the results of bivariate regression predicting math performance with each of the four autonomy indicators. PISA math scores that are provided by the OECD are normalized with respect to the OECD mean (mean =500) and standard deviation (sd=100). Thus, a coefficient estimated at +50 means students in autonomous schools scored 0.50 standard deviations higher on average on the PISA math test than students at non-autonomous schools. Columns 1-4 show substantial variation in the coefficients for different autonomy types. We estimate that scores are approximately 31 points higher at schools with hiring autonomy, 8 points higher at schools with content autonomy, 8 points lower at schools with admissions autonomy, and no difference for schools with budget autonomy.

In columns 5-8, we replicate these results adding student characteristics, school aggregate characteristics, and country-level fixed effects. Coefficients for all four autonomy types are substantially changed by the addition of control variables. The formerly significant coefficients for hiring, content, and admissions autonomy are all null, while budget autonomy, which was insignificant without controls, is associated with a small, significant increase in student performance of approximately 5 points. This suggests that models without controls are

problematic in multiple ways as they both overstate and understate autonomy coefficients. The substantial change in the coefficient for hiring autonomy from +32 to -1 also suggests that there is substantial selection bias in the assignment of autonomy to schools.

In columns 9-12 we include all four autonomy indicators in a single estimate of equation (1) and incrementally add student controls (column 10), school controls (column 11), and country fixed effects (column 12). The uncontrolled results in column 9 show positive estimates of hiring autonomy, but negative coefficients for budget and admissions autonomy. Comparing to columns 1-4, this suggests that coefficients attached to autonomy variables are sensitive to the types of autonomy granted to schools and the researcher definition of how autonomy is measured. Adding student controls leads to substantive shifts in all four coefficients. The hiring and admission estimates are null, the budget estimate switches its sign, and the estimate for content autonomy is now statistically significant. This again suggests substantial selection bias into different autonomy types based on student characteristics. Adding school characteristics (column 11) produces more modest changes in coefficients. Finally, in column 12, we add country fixed effects. The only autonomy type that remains significant is budget autonomy, which has a small positive coefficient of 4.62 points – although it was significant and negative in column 9. Taken as a whole, this leads to three conclusions regarding empirical estimates of school autonomy on test scores. First, coefficients are highly sensitive to how autonomy is defined and measured. Second, autonomy is endogenous to student sorting and therefore estimates are subject to omitted variable bias. Finally, coefficients in specifications with country fixed effects suggest that the relationship between autonomy and student performance varies by context.

We next estimate equations (1) and (2) separately for each of the 34 countries in the analytic sample and each type of autonomy. The results are presented in Figure 2 and summarized in Table 4. Figure 2 displays countries on the x-axis and coefficient sizes on the y-axis. Black circles display autonomy coefficients with no controls (*β's* from equation 1). Red circles display coefficients with student and school controls (*β's* from equation 2). For both sets of estimates, we display statistically significant results ($p<0.05$) with solid circles, and null results as hollow circles. Table 4 summarizes the same estimates by tabulating countries with negative, positive, and null (i.e., not statistically significant) *β's* for equation 1 (columns) against the *β's* for equation 2 (rows). The diagonal from the upper left to the lower right corner counts the number of countries with results that are similar in direction and significance with the inclusion of covariates – i.e., estimates that are robust to the inclusion of covariates.

All four graphs in Figure 2 show many countries with positive coefficients for specifications without student and school controls (solid black circles) and few or one with negative estimates. Hiring and content autonomy show particularly large, positive coefficients in many settings in uncontrolled regressions. The size of these estimates suggests that students in schools with autonomy have test scores that are 5 to 78 points higher than students in schools without autonomy (up to 0.78 standard deviations). However, including student and school characteristics in the estimation (red circles), nearly all previously significant coefficients are closer to zero and few remain statistically significant. Only six cases across 34 countries and four autonomy types show statistically significant positive estimates. In two cases, significant positive coefficients attached to hiring autonomy become significant negative estimates if control variables are added. Negative coefficients are robust in only three countries for content autonomy and one country for admission autonomy. This suggests that while the relationship

between autonomy and test scores is mostly positive, it is also substantially explained by relatively advantaged students sorting into autonomous schools. Our results suggests that estimates of autonomy without student controls are subject to substantial selection bias and might even be estimated in the wrong direction.

Autonomy and Performance Gaps

In the next step, we analyze whether school decentralization decreases or increases achievement gaps for commonly underserved groups. We present results for all pooled regressions in Table 5 for gender gaps (columns 1-4), wealth gaps (columns 5-8), and parent education gaps (columns 9-12). The coefficients for the subgroup variable measure the association between student characteristics and test scores for schools without autonomy (β_2 in equation 3). This is the PISA performance gap for the subgroup relative to other students at non-autonomous schools. Estimates on the interaction terms for subgroup and autonomy (β_3 in equation 3) reflect whether attending an autonomous school increases or decreases these achievement gaps on average.

In the pooled sample, we estimate significant subgroup gaps by gender, wealth, and parent education at non-autonomous schools ranging from 9 to 19 points. The autonomy coefficients, controlling for subgroup membership are mostly positive but not statistically significant. We estimate significant coefficients only for budget autonomy when controlling for student wealth and for budget and content autonomy when controlling for parent education, all in a positive direction. The autonomy estimates on subgroup achievement (β_3) are ambiguous with both negative and positive coefficients. Significant coefficients for β_3 suggest that budget and content autonomy *decrease* gender gaps, while content autonomy *increases* parent education gaps.

As in our first analysis, pooled regressions potentially disguise differences between countries. Thus, once more, we run specifications separately for each country in our sample. The results are displayed graphically in Figures 3-5 and numerically in Table 6. The coordinate systems show again country names on the x-axis and the size of coefficients on the y-axis. Black circles report estimates for schools without autonomy (β_2 in equation 3) and red circles for schools with autonomy (linear combination of $\beta_2 + \beta_3$ from equation 3). Once more, we show statistically significant coefficients as filled circles. Table 6 shows the frequency of country-level performance gaps at schools without autonomy (rows), tabled against the presence of significant gaps at schools with autonomy (columns). If autonomy eliminates gaps, we would see significant gaps at non-autonomous schools only. If autonomy increases or causes gaps, we would see significant gaps at autonomous schools as well.

Figure 3 and Table 6, Panel A present findings for the association between gender gaps and autonomy, we identify a gender gap where boys significantly outperform girls. Across autonomy types, between 21-23 countries have significant gender gaps where boys out-perform girls (i.e., β_2 is significant and negative), on average, at non-autonomous schools. These differences range between 20 to 40 points on PISA math test. In countries with significant gender gaps at non-autonomous schools, we typically find smaller gender gaps at autonomous schools or, less often, no gaps at all at autonomous schools. However, there are also select cases where gender gaps are larger at autonomous schools or where gender gaps exist only at autonomous schools. Figure 4 and Table 6, Panel B display similar results for performance gaps based on family wealth, and Figure 5 and Table 6, Panel C display results for parent education gaps. While autonomy often appears to reduce gaps in student performance, there are also countries

where gaps are larger at autonomous schools. In several cases wealth or parent education gaps exist only in autonomous schools, with non-autonomous schools having more equal outcomes.

Interaction between Autonomy and Test-Based Accountability

Our final analysis investigates if accountability influences student achievement in schools with different forms of autonomy. Here, we limit our analytic sample to students at autonomous schools and estimate subgroup performance gaps at schools with and without accountability, where accountability is defined as a requirement to make student performance data public. The estimates for the pooled regression are presented in Table 7. From equation (4), β_1 is the coefficient of accountability on non-subgroups members, β_2 is the subgroup performance gap at schools without accountability, and β_3 is the differential subgroup performance gaps at schools with accountability. We again find significant and substantial achievement gaps ranging from 6-9 points for gender, 16-19 points for wealth, and 15-18 points for parent education (β_2). In pooled regressions, we find only small estimates that are not statistically significant of the interactions between accountability and subgroups (β_3). This suggests that while accountability may be associated with small performance gains, accountability at autonomous schools neither reduces nor increases achievement gaps internationally.

Figures 6-8 and Table 8 display results for this accountability analysis disaggregated by country. In Figures 6-8, black circles represent subgroup achievement gaps at autonomous schools without accountability (β_2), and red circles represent gaps at schools with autonomy and accountability ($\beta_2+\beta_3$), with sold circles showing results that are statistically different than zero. Again, we find substantial inconsistency in estimates across countries. We find cases where gaps exist only at schools with accountability, only at schools without accountability, and at all types of schools regardless of accountability.

Robustness Checks

Notably, our primary finding is that any observed relationships between autonomy and student outcomes are subject to selection bias. As this likely applies to our own estimates, as well as those of prior studies, we conducted several robustness checks to ensure that conclusions regarding the research questions were not influenced by any particular analytic choices. We find that while individual coefficients may vary with different choices, there are no specifications or definitions that conflict with the general findings that decentralization coefficients are subject to selection bias, that associations vary across settings, and that associations with accountability and achievement gaps vary by context. Our first robustness checks address the identification of autonomy and accountability. We first altered the selection of 34 countries with 20 to 95% autonomous schools. Ranges of 5 to 95% (35 countries) and 10 to 90% (32 countries) yielded similar coefficients for autonomy indicators and similar evidence of selection bias when control variables were added. Next, we tested a stricter definition of autonomy that identified schools as autonomous only if school leaders and teachers had full control, instead of possibly sharing control with higher authority. This led to fewer than 50 schools (out of over 7,500) shifting status from autonomous to not autonomous and relevant coefficients were statistically unchanged. Finally, we tested the robustness of our decision to exclude observations with any missing values for covariates from estimates without covariables. Again, results for uncontrolled regressions were similar for the full sample and sample without missingness.

Our second set of robustness checks relates to choices made about PISA and the functional form of the specifications. In addition to our reported regressions with controls, we estimated other specifications. We replicated the longer list of control variables from Hanushek, Link, and Woessman (2013) and also added fixed effects for groupings of similar students in

similar schools.[10] Estimated coefficients for autonomy in the pooled regression with extra controls and matching were similar to what is reported above, again suggesting that uncontrolled estimates for autonomy are subject to selection bias. Finally, we tested the importance of our decision to include PISA weights and to use only the first of ten plausible values for the math score. Both issues are vigorously debated in literature (e.g. Jerrim et al. 2017, Arikan et al. 2020, Mang et al. 2021). Our results were not substantively different if sample weights were excluded or if estimates were run on other plausible values.

6. Discussion and Conclusion

The results in the figures and tables show that accountability decreases the number of countries with statistically significant achievement gaps for at-risk students for almost all forms of school autonomy. These reductions include between six and 53 percent of the countries in our sample, depending on the type of autonomy and achievement gap under observation. While the pooled analysis did not show that accountability changed achievement gaps between students, there is some evidence that autonomy coupled with accountability increases or decreases these gaps in the country level analysis. School autonomy emerged as one of the most important education reform policies over the last three decades. The theoretical discussion and empirical evidence on the topic, however, remain mixed and inconclusive. Grounded in conceptual arguments for and against education decentralization, we formulate three theoretical mechanisms that potentially explain the varying results in the existing empirical studies. These mechanisms include: (1) the ways additional autonomy is granted to schools; (2) how decentralization

[10] Students were matched based on these school characteristics: upper or lower secondary, school size deciles, rural location, school poverty rate, school rate of mother's who did not attend college; and these individual characteristics: gender, household wealth quintile, mother's education level, and student nativity.

influences already existing achievement gaps between students; and (3) the relationship between autonomy and accountability policies.

Regarding the first mechanism, we find that the relationship between school autonomy and student performance depends on the definition and measure of education decentralization. We further observe that central authorities tend to implement and retain additional autonomy in schools with characteristics that are associated with high performance. Once empirical models control for student and school characteristics, the relationship between student achievement and autonomy is only in few countries of considerable importance. In these countries, the association between decentralization and student performance is mainly positive and, only in few countries, negative. As our analysis shows, the direction of this relationship depends on country specific context factors.

The findings for the second and third mechanism demonstrate that decentralization can either be negatively or positively correlated with existing equity-based achievement gaps or accountability policies. While autonomy often reduces performance gaps between student groups, it also increases them in several countries. Similarly, for accountability policies, we find decreases in achievement gaps for at-risk students in some countries, but there are also countries where differences in student performance grow. Overall, there is no generally observable pattern that autonomy reduces achievement gaps between student groups or that decentralization and accountability together diminish learning discrepancies between student groups.

We add several novel insights to the scholarly work on school autonomy. For instance, most literature, investigating the relationship between autonomy and student achievement, builds on abstract theoretical concepts such as principal-agent approaches (Hanushek, Link and Woessmann 2013; Woessmann et al. 2007). We add to these considerations by introducing a

conceptual framework that is also grounded in pathways of student achievement, mechanisms of policy implementation, and unintended policy effects. Furthermore, while prior research uses country level proxies for governance and human capital (Hanushek, Link and Woessmann 2013), we measure these characteristics directly at the school and student level. As a consequence of these conceptual and empirical contributions, we are able to establish and measure mechanisms describing the relationship between autonomy and student achievement that have not been discussed previously. Finally, our analysis uses fine grained definitions of autonomy and accountability, similar to authors developing categories for international comparisons (Arcia et al. 2011), but in contrast to these studies, we also put our definitions of autonomy and accountability to an empirical test.

Our findings also have several implications for practitioners. First and foremost, it is important to note that the impact of autonomy on students' test scores is context dependent. Relevant context factors include, according to our analysis, different forms of decentralization (e.g., hiring, budget, content, and admission autonomy), selective implementation of school autonomy, existing achievement gaps between students, and accountability policies. School autonomy (and accountability) may not have the desired policy effects if these factors are not considered and could even lead to decreases in student achievement, for all or some groups of students. The application of additional background information will make overarching policy recommendations more difficult but has the potential of formulating autonomy programs that are more successful.

For policy makers and scholars interested in measuring the effect of autonomy on student achievement, we recommend the following three strategies to overcome bias in the evaluation of decentralization policies across countries. First, researchers need to account for different forms

and measures of school autonomy, as both influence student performance differently. Second, studies need to control for the selective implementation of decentralization policies, which is often based on school and student characteristics. Finally, school autonomy interacts with other education policies and existing inequities in education. These nuances need to be carefully analyzed to fully grasp the extent to which decentralization affects student learning.

Having emphasized several contributions, our results are, as any empirical findings, subject to caveats. For instance, our results could be influenced by the selective implementation of school autonomy. Future studies could address this issue by identifying mechanisms for how school decentralization is granted or leveraging external shifters of school autonomy in a single country or subsets of countries. Moreover, our variables measure school and student characteristics in a quantifiable and internationally comparable way. Additional research could employ qualitative methods to establish pathways that are more country, school, or student specific.

References

Ainley, J., & McKenzie, P. (2000). School governance: Research on educational and management issues. *International education journal*, *1*(3), 139-151.

Ammermüller, A. (2005). Educational opportunities and the role of institutions. *ZEW-Centre for European Economic Research Discussion Paper*, (05-044).

Anderson, P. M., & Butcher, K. F. (2006). Reading, Writing, and Refreshments Are School Finances Contributing to Children's Obesity?. *Journal of Human Resources*, *41*(3), 467-494.

Arcia, G., Macdonald, K., Patrinos, H. A., & Porta, E. (2011). School autonomy and accountability. SABER;. World Bank.

Azam, M., & Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, *42*, 143-164.

Arikan, S., Ferah, Ö, Şeker, V., & Ertas, G. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(1), 43-60.

Barankay, I., & Lockwood, B. (2007). Decentralization and the productive efficiency of government: Evidence from Swiss cantons. *Journal of public economics*, *91*(5-6), 1197-1218.

Bardhan, P., & Mookherjee, D. (2005). Decentralizing antipoverty program delivery in developing countries. *Journal of public economics*, *89*(4), 675-704.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press. Chicago, IL.

Bollen, R. (1996). School effectiveness and school improvement: The intellectual and policy context. In D. Reynolds, R. Bollen, B. Creemers, D. Hopkins, L. Stoll, & N. Lagerweij (Eds.), Making good schools. Linking school effectiveness and school improvement (pp. 1–19). London: Routledge

Bouhlila, D. S. (2017). Parents' education and literacy skills: Evidence on inequality of socioeconomic status in Arab countries. *World Development Perspectives*, *5*, 34-43.

Bourdieu, P. (1977). The economics of linguistic exchanges. *Social science information*, *16*(6), 645-668.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard university press. Cambridge, Mass.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences?. *Public Finance Review*, *36*(1), 88-111.

Branisa, B., Klasen, S., & Ziegler, M. (2013). Gender inequality in social institutions and gendered development outcomes. *World Development*, *45*, 252-268.

Buerger, C., & Harris, D. N. (2020). The impact of government contracting out on spending: The case of public education in New Orleans. *The American Review of Public Administration*, 2021;51(2):139-154.

Bulkley, K. (2005). Understanding the charter school concept in legislation: The cases of Arizona, Michigan and Georgia. *International Journal of Qualitative Studies in Education*, 18(4), 527-554.

Bulkley, K., & Fisler, J. (2003). A Decade of Charter Schools. *Educational Policy*, *17*, 317-342.

Bulkley, K. E., Henig, J. R., & Levin, H. M. (2010). *Between Public and Private: Politics, Governance, and the New Portfolio Models for Urban School Reform*. Harvard Education Press. Cambridge, MA.

Burian-Fitzgerald, M., Luekens, M. T., & Strizek, G. A. (2004). Less red tape or more green teachers: Charter school autonomy and teacher qualifications. *Taking account of charter schools: What's happened and what's next*, 11-31.

Daramola, J. and Mulfinger,L.S. (2020). Autonomy: flexibility in response to student needs. In: Bulkley, K.E., Marsh, J. A., Strunk, K.O., Harris, D. N. and Hashim, A.K, Challenging the one best system : the portfolio management model and urban school governance. Harvard Education Press. Cambridge, MA.

Demas, A., & Arcia, G. (2015). What Matters Most for School Autonomy and Accountability. A Framework Paper. Systems Approach for Better Education Results SABER Working Paper Series. http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/SAA/SABER_School_Autonomy_and_Accountability_What_Matters_Framework_Paper.pdf

Dercon, S., & Singh, A. (2013). From nutrition to aspirations and self-efficacy: Gender bias over time among children in four countries. *World Development*, *45*, 31-50.

Diaz-Serrano, L., & Meix-Llop, E. (2012). Do fiscal and political decentralization raise students' performance? A cross-country analysis.

Faguet, J. P. (2014). Decentralization and governance: a special issue of World Development, 2013. *World Development*, *53*.

Faguet, J. P., & Sánchez, F. (2008). Decentralization's effects on educational outcomes in Bolivia and Colombia. *World development*, *36*(7), 1294-1316.

Fan, C. S., Lin, C., & Treisman, D. (2009). Political decentralization and corruption: Evidence from around the world. *Journal of Public Economics*, *93*(1-2), 14-34.

Finnigan, K. S. (2007). Charter school autonomy: The mismatch between theory and practice. *Educational Policy*, *21*(3), 503-526.

Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, *90*(4-5), 837-851.

Figlio, D., & Loeb, S. (2011). School accountability. In: Almlund, M., Duckworth, A. L., Heckman, J. J., Kautz, T., Hanushek, E. A., Machin, S., & Woessmann, L. (Eds.) *Handbook of the Economics of Education*, *3*, 383-421.

Fuchs, T., & Woessmann, L. (2008). What accounts for international differences in student prformance? A re-examination using PISA data. In *The economics of education and training* (pp. 209-240). Physica-Verlag HD.

Galiani, S., Gertler, P., & Schargrodsky, E. (2008). School decentralization: Helping the good get better, but leaving the poor behind. *Journal of public economics*, *92*(10-11), 2106-2120.

Galiani, S., Schargrodsky, E. (2002). Evaluating the impact of school decentralization on educational quality. *Economia*, *2*(2), 275-314.

Gawlik, M. A. (2007). Beyond the charter schoolhouse door: Teacher-perceived autonomy. *Education and Urban Society*, *39*(4), 524-553.

Gertler, P., Patrinos, H., & Rubio-Codina, M. (2006). Empowering Parents to Improve Education: Evidence from Rural Mexico. Impact Evaluation Series Number 4. World Bank Policy Research Working Paper 3935. *World Bank Publications*.

Hannaway, J. (1993). Decentralization in two school districts: Challenging the standard paradigm. In J. Hannaway & M. Carnoy (Eds.), Decentralization and school improvement – Can we fulfil the promise? (pp. 135-162). San Francisco: Jossey-Bass.

Hanushek, E. A., Link S., & Woessmann L. (2013). Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA. *Journal of Development Economics* 104 (September): 212–32.

Hopkins, D., & Lagerweij, N. (1996). The school improvement knowledge base. In D. Reynolds, R. Bollen, C. Bert, D. Hopkins, L. Stoll, & N. Lagerweij (Eds.), Making good schools. Linking school effectiveness and school improvement (pp. 61–93). New York: Routledge.

Hopkins, D., & Reynolds, D. (2001). The past, present and future of school improvement: Towards the third age. British Educational Research Journal, 24(7), 459–475.

Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. *Journal of Public Economics*, *74*(1), 1-30.

Huisman, J., & Smits, J. (2009). Effects of household-and district-level factors on primary school enrollment in 30 developing countries. *World development*, *37*(1), 179-193.

Ingersoll, R. M. (2009). *Who controls teachers' work?: Power and accountability in America's schools*. Harvard University Press.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of public Economics*, *89*(5-6), 761-796.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, *118*(3), 843-877.

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*, 51-58.

Jeong, D. W., Lee, H. J., & Cho, S. K. (2017). Education decentralization, school resources, and student outcomes in Korea. *International Journal of Educational Development*, *53*, 12-27.

Jimenez, E., & Sawada, Y. (1999). Do community-managed schools work? An evaluation of El Salvador's EDUCO program. *The world bank economic review*, *13*(3), 415-441.

Kafle, K., Jolliffe, D., & Winter-Nelson, A. (2018). Do different types of assets have differential effects on child education? Evidence from Tanzania. *World Development*, *109*, 14-28.

Kameshwara, K. K., Sandoval-Hernandez, A., Shields, R., & Dhanda, K. R. (2020). A false promise? Decentralization in education systems across the globe. *International Journal of Educational Research*, *104*, 101669.

Kazeem, A., Jensen, L., & Stokes, C. S. (2010). School attendance in Nigeria: Understanding the impact and intersection of gender, urban-rural residence, and socioeconomic status. *Comparative education review*, *54*(2), 295-319.

Koretz, D. M., & Barron, S. I. (1998). *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)*. Rand Education https://www.rand.org/content/dam/rand/pubs/monograph_reports/2009/MR1014.pdf

Lüdemann, E., Schütz, G., West, M. R., & Woessmann, L. (2007). School Accountability, Autonomy, Choice and the Level of Student Achievement: International Evidence from PISA 2003. *Books*.

Levin, H. M. (1974). A conceptual framework for accountability in education. *The School Review*, *82*(3), 363-391.

Levin, H. M. (2012). Some economic guidelines for design of a charter school district. *Economics of Education Review*, *31*(2), 331-343.
Lincove, J. A. (2006). Efficiency, equity and girls' education. *Public Administration and Development: The International Journal of Management Research and Practice*, *26*(4), 339-357.
Lincove, J.A. & Bulkley, K.E. (2020). Chapter 8: School supports: Administration, instruction, compliance. In K.E. Bulkley, J.A. Marsh, K.O. Strunk, D.N. Harris, & A. K. Hashim, *Challenging the One Best System: The portfolio management model and urban school governance.* Harvard Education Press.
Lounkaew, K. (2013). Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review*, *37*, 213-225.
Mang, J., Küchenhoff, H., Meinck, S., & Prenzel, M. (2021) Sampling weights in multilevel modelling: an investigation using PISA sampling structures. *Large-scale Assessments in Education*, 9, 1-39.
Manor, J. (1999). The political economy of democratic decentralization. The World Bank.
Masterson, T. (2012). An empirical analysis of gender bias in education spending in Paraguay. World Development, 40(3), 583-593.
Merrouche, O. (2007). *The long term effect of education spending decentralization on human capital in Spain* (No. 2007: 3). Working Paper. https://www.econstor.eu/bitstream/10419/82650/1/wp2007-003.pdf
Meunier, M., De Coulon, A., Marcenaro-Gutierrez, O., & Vignoles, A. (2013). A longitudinal analysis of UK second-generation disadvantaged immigrants. *Education Economics*, *21*(2), 105-134.
Mont, D., & Nguyen, C. (2013). Does parental disability matter to child education? Evidence from Vietnam. *World Development*, *48*, 88-107.
Naper, L. R. (2010). Teacher hiring practices and educational efficiency. *Economics of Education Review*, *29*(4), 658-668.
Neeleman, A. (2019). The scope of school autonomy in practice: An empirically based classification of school interventions. *Journal of Educational Change*, *20*(1), 31-55.
Oates, W. E. (1972). *Fiscal federalism.* New York: Harcourt Brace Jovanovich. .
Oates, W. E. (1999). An essay on fiscal federalism. *Journal of economic literature*, *37*(3), 1120-1149.
Oates, W. E. (2005). Toward a second-generation theory of fiscal federalism. *International tax and public finance*, *12*(4), 349-373.
Organisation for Economic Co-operation and Development (OECD). (2009). PISA Data Analysis Manual. SAS, Second Edition. OECD Publishing.
Organisation for Economic Co-operation and Development (OECD). (2011). *School autonomy and accountability: Are they related to student performance?*. OECD Publishing.
Organisation for Economic Co-operation and Development (OECD). (2013). School Governance, Assessments and Accountability. *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices*, 127-64. OECD Publishing.
Organisation for Economic Co-operation and Development (OECD). (2016). PISA 2015 Results (Volume II). Policies and Practices for Successful Schools. OECD Publishing.
Organisation for Economic Co-operation and Development (OECD). (2017). PISA 2015 Technical Report. OECD Publishing.

Ouchi, W. G. (2003). Making schools work: The seven keys to success. *Education Week*, *23*(1), 56.
Ouchi, W. G. (2004). Academic freedom. Education Next 4(1): 21–25.
Parry, T. R. (1997). Achieving balance in decentralization: A case study of education decentralization in Chile. *World development*, *25*(2), 211-225.
Prawda, J. (1993). Educational decentralization in Latin America: lessons learned. *International journal of educational development*, *13*(3), 253-264.
Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender achievement gaps in US school districts. *American Educational Research Journal*, *56*(6), 2474-2508.
Lüdemann, E., Schütz, G., West, M. R., & Woessmann, L. (2007). School Accountability, Autonomy, Choice and the Level of Student Achievement: International Evidence from PISA 2003. OECD Education Working Papers, No. 13.
Radulović, B., Županec, V., Stojanović, M., & Budić, S. (2022). Gender motivational gap and contribution of different teaching approaches to female students' motivation to learn physics. *Scientific Reports*, 12(1), 18224.
Schippers, M. C., Scheepers, A. W., & Peterson, J. B. (2015). A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap. *Palgrave Communications*, 1(1), 1-12.
Siddle, A., & Koelble, T. A. (2012). *The failure of decentralisation in South African local government: Complexity and unanticipated consequences*. Cape Town: UCT Press.
Springer, M. G. (2007). Accountability incentives: Do failing schools practice educational triage. *Education Next*, *8*(1), 74-79.
Stoddard, C., & Corcoran, S. P. (2007). The political economy of school choice: Support for charter schools across states and school districts. *Journal of Urban Economics*, 62(1), 27-54.
Suárez-Orozco, C., Gaytán, F. X., Bang, H. J., Pakes, J., O'Connor, E., & Rhodes, J. (2010). Academic trajectories of newcomer immigrant youth. *Developmental psychology*, *46*(3), 602.
Sundaram, A., & Vanneman, R. (2008). Gender differentials in literacy in India: The intriguing relationship with women's labor force participation. *World Development*, *36*(1), 128-143.
Suggett, D. (2015). School autonomy: Necessary but not sufficient. *Evidence base: a journal of evidence reviews in key policy areas*, *2015*(1), 1.
Taş, E. O., Reimão, M. E., & Orlando, M. B. (2014). Gender, ethnicity, and cumulative disadvantage in education outcomes. *World Development*, *64*, 538-553.
Ver Hoef, J. M. (2012). Who invented the delta method?. The American Statistician, 66(2), 124-127.
Waslander, S., Hooge, E. H., Theisens, H. C., & Pater, C. J. (2020). Steering in complex education systems. Why similar aims can have dramatically different results. *Journal of Education Policy*, *35*(2), 168-187.
West, M. R., & Woessmann, L. (2012). How School Choice, Autonomy, and Accountability Impact Student Achievement: International Evidence. Wolf Legal Publishers.
Woessmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford bulletin of economics and statistics*, *65*(2), 117-170.
Woessmann, L. (2007). International evidence on school competition, autonomy, and accountability: A review. *Peabody journal of education*, *82*(2-3), 473-497.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2007). School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003. https://www.oecd-ilibrary.org/content/paper/246402531617

Wohlstetter, P., & Chau, D. (2004). Does autonomy matter? Implementing research-based practices in charter and other public schools. In Bulkley, K. E., & Wohlstetter, P. (Eds.). (2004). *Taking account of charter schools: What's happened and what's next?*. Teachers College Press. 53-71.

Wohlstetter, P., Wenning, R., & Briggs, K. L. (1995). Charter schools in the United States: The question of autonomy. *Educational Policy*, *9*(4), 331-358.

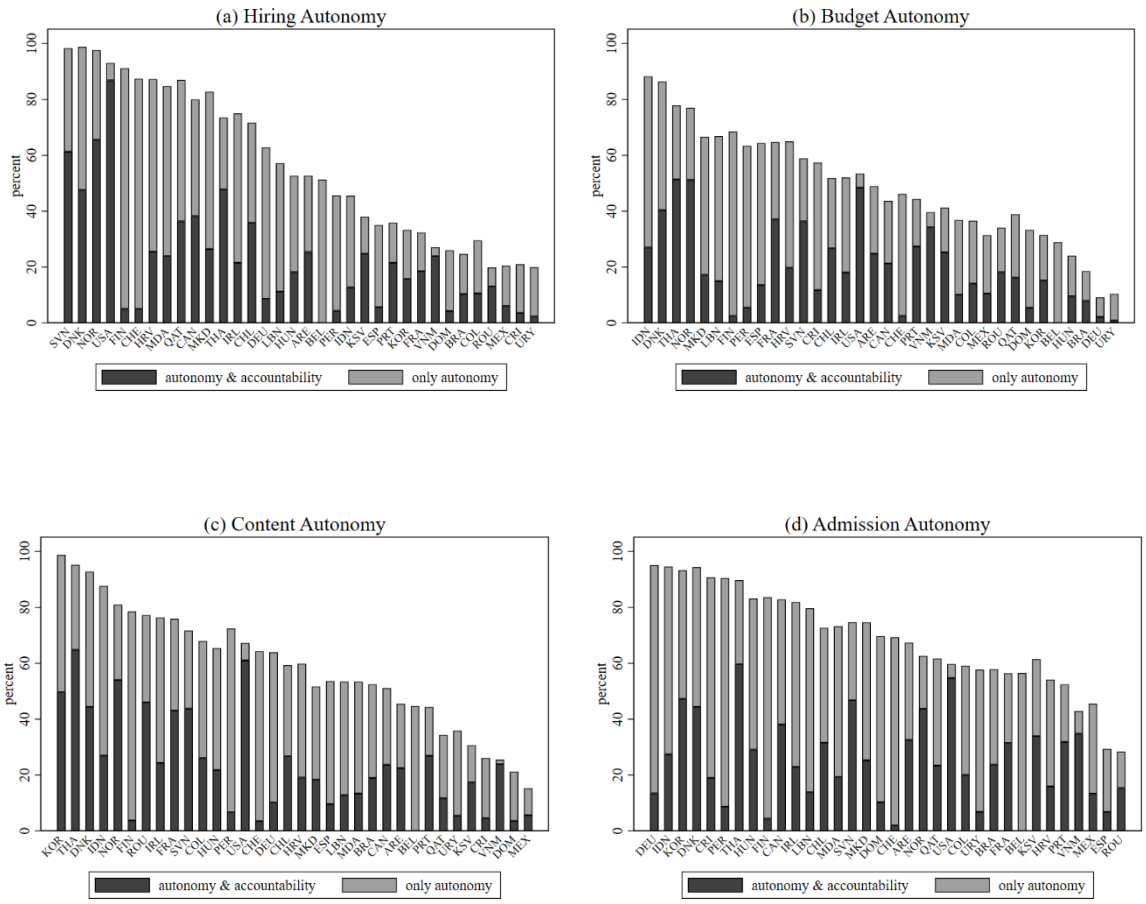
World Bank. (2003). World development report 2004: making services work for poor people. The World Bank.

Yao, Y., & You, W. (2018). Women's political participation and gender gaps of education in China: 1950–1990. World Development, 106, 220-237.

Yuki, T., & Igei, K. (2020). School autonomy and learning outcomes in Burkina Faso and Senegal. In: Nishimura, M. *Community Participation with Schools in Developing Countries: Towards Equitable and Inclusive Basic Education for All*, Routledge. New York, NY. 184-197.

Zhang, D., Li, X., & Xue, J. (2015). Education inequality between rural and urban areas of the People's Republic of China, migrants' children education, and some implications. *Asian Development Review*, *32*(1), 196-224.

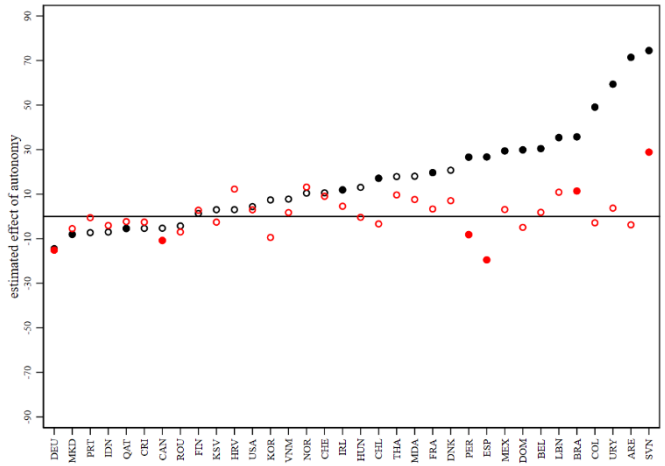
Figure 1. Frequency of School Autonomy and Accountability in PISA Countries



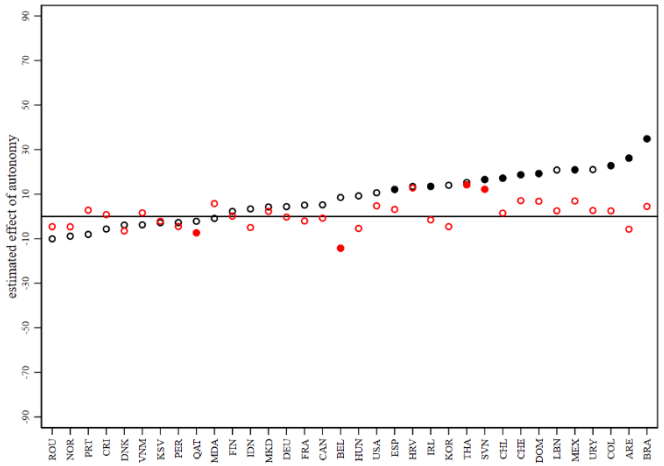
Notes: Author calculations based on school managers' responses to the 2015 PISA school survey. A school is coded as having an autonomy type if the manager responds that the school has authority in that area. A school is coded as having accountability if it is required to share performance data with the public.

Figure 2: Estimated effects of autonomy on test scores with and without controls

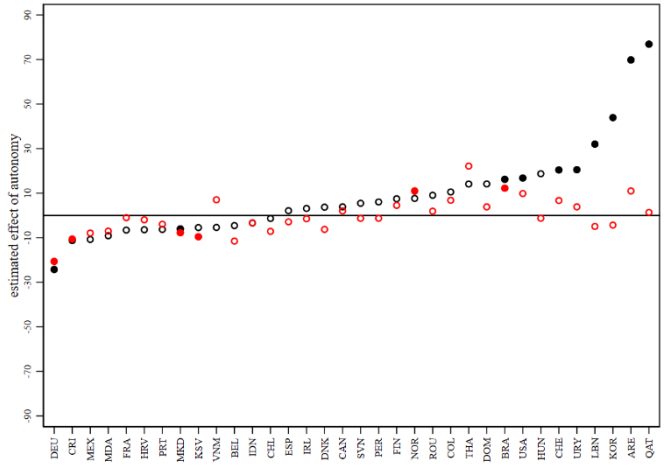
(a) Hiring autonomy



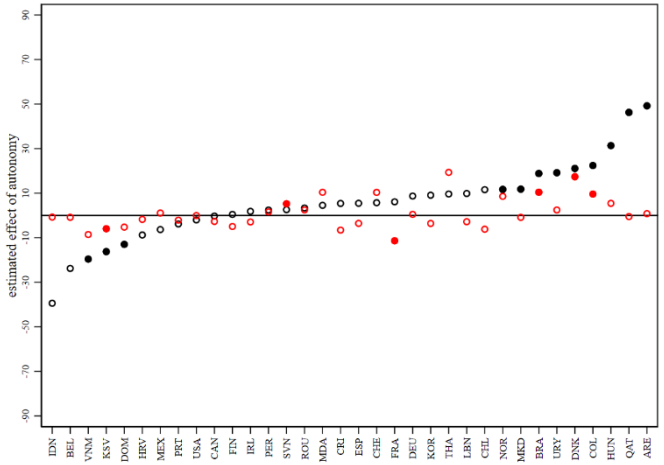
(b) Budget autonomy



(c) Content autonomy



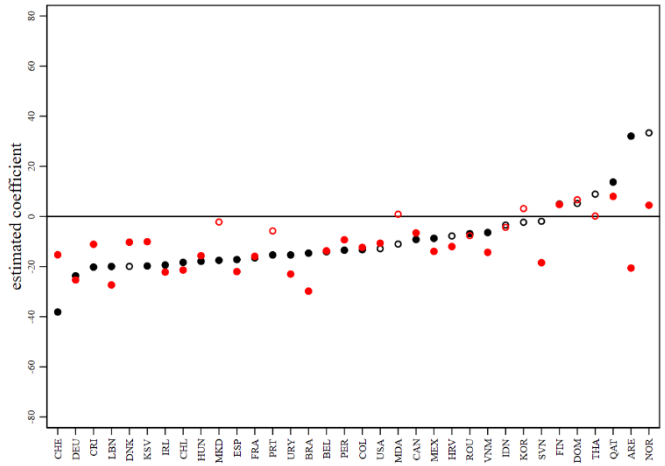
(d) Admission autonomy



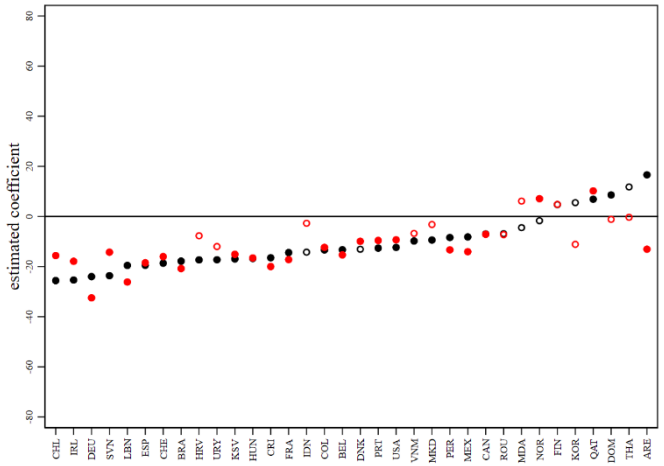
Notes: Graphs display estimates of the effects of autonomy on PISA scores without (black) and with (red) control variables. Filled circles are statistically significant at 95% or above. Student's variables included as controls are: gender, age, migration status, and mother's education, School controls are: the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree.

Figure 3: Estimated gender gap in PISA scores for schools with and without autonomy

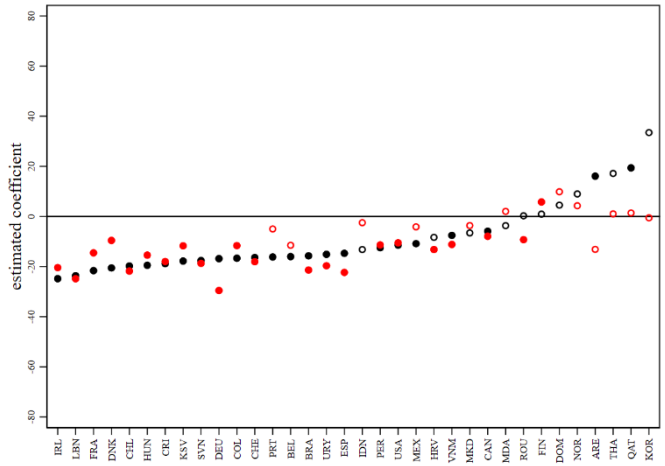
(a) Hiring autonomy



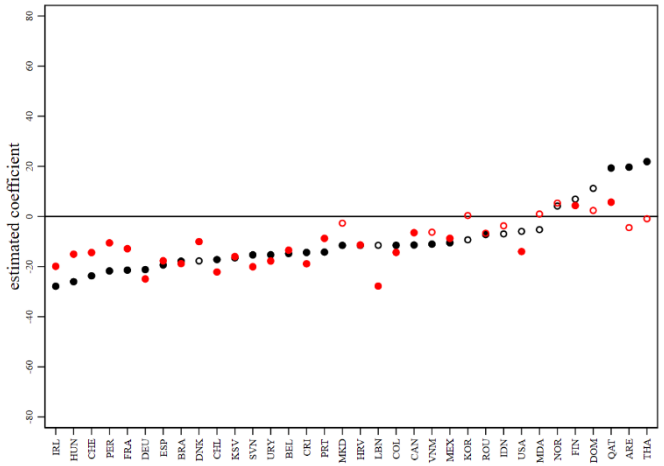
(b) Budget autonomy



(c) Content autonomy



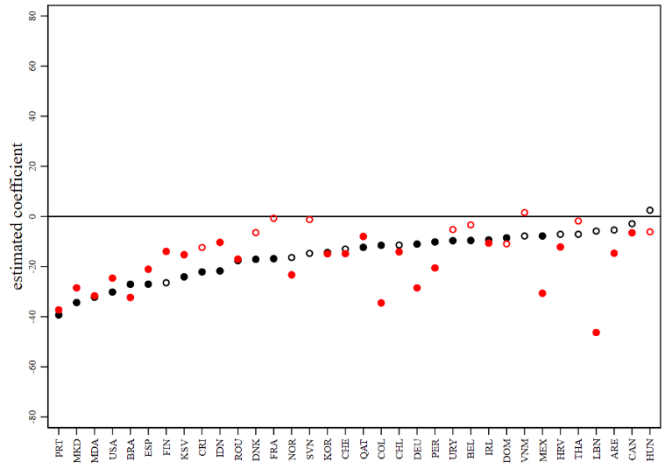
(d) Admission autonomy



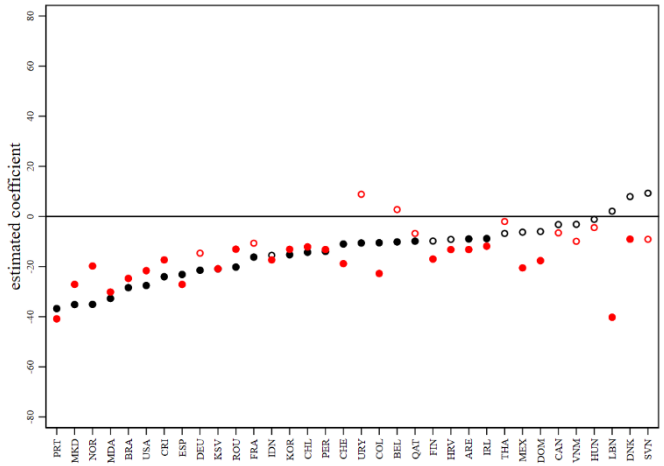
Notes: Graphs display estimates of the gender gap schools without (black) and with (red) autonomy. Gender gap is the difference between average female scores and average male scores, *ceteris paribus*. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Figure 4: Estimated wealth gap in PISA scores for schools with and without autonomy

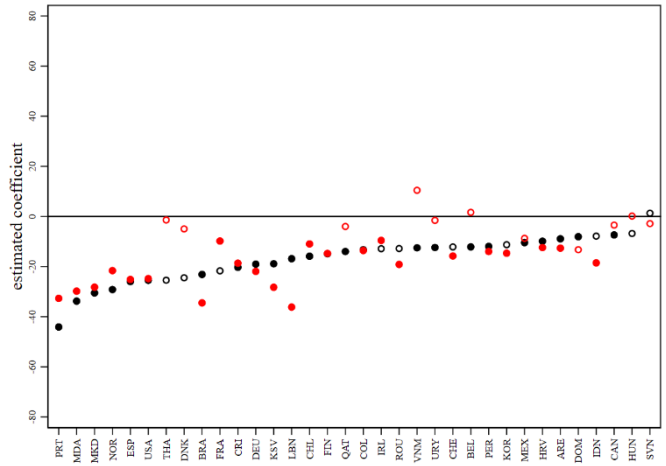
(a) Hiring autonomy



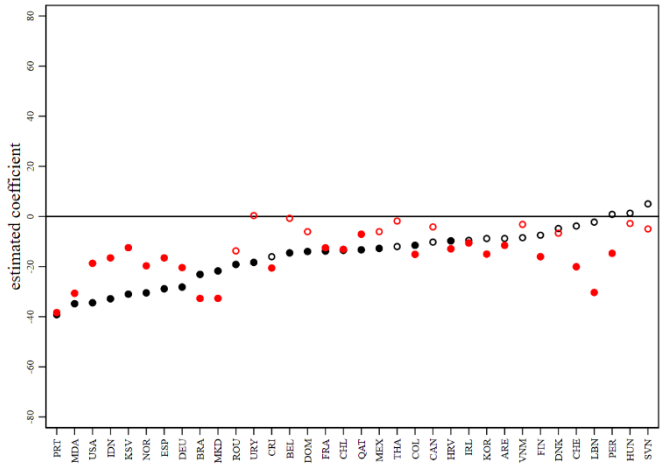
(b) Budget autonomy



(c) Content autonomy



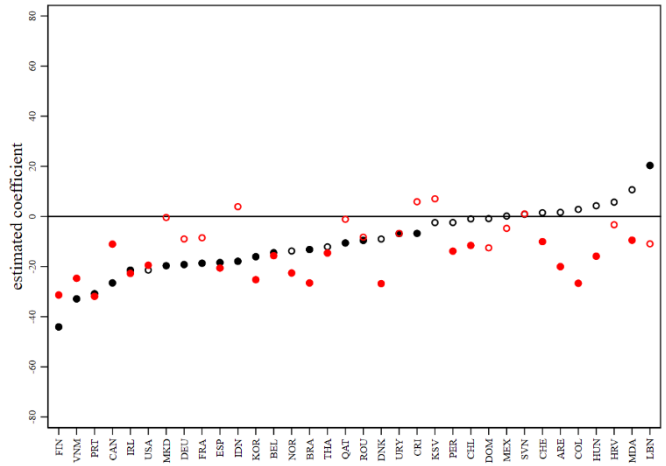
(d) Admission autonomy



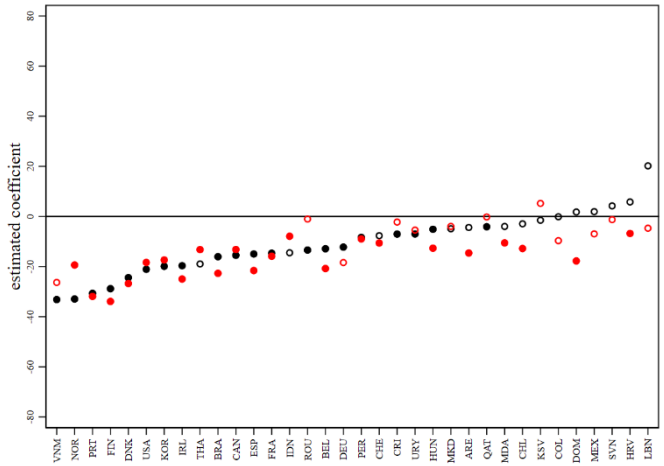
Notes: Graphs display estimates of wealth gaps for schools without (black) and with (red) autonomy variables. Wealth gap is difference in scores between the average student in lowest PISA wealth quintile and the average student in the top 4 wealth quintiles, *ceteris paribus*. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Figure 5: Estimated parent education gap in PISA scores for schools with and without autonomy

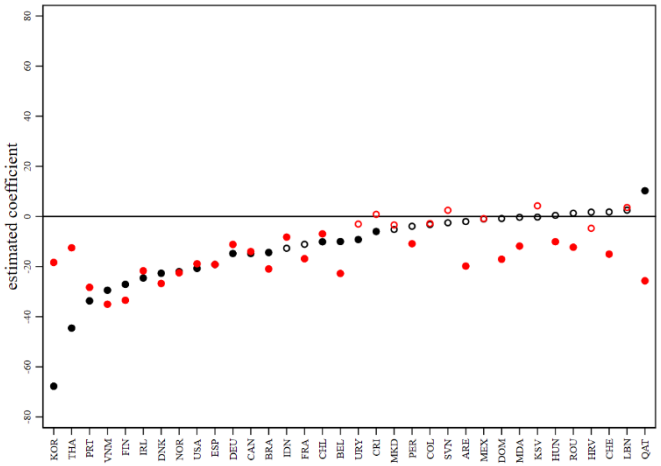
(a) Hiring autonomy



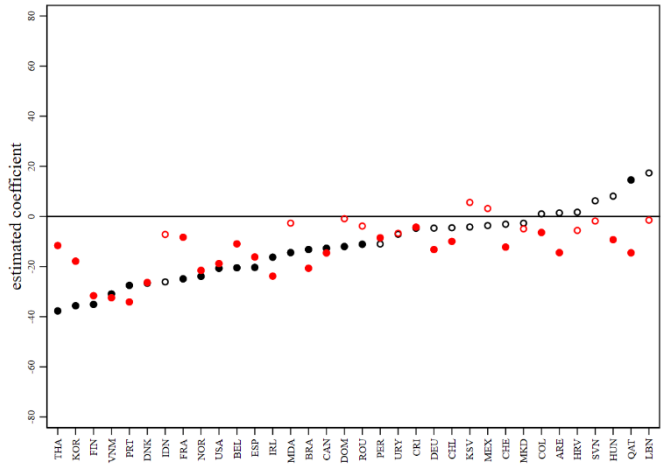
(b) Budget autonomy



(c) Content autonomy



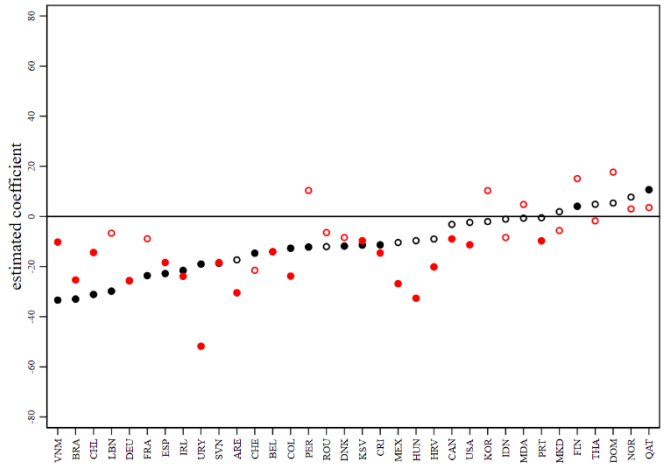
(d) Admission autonomy



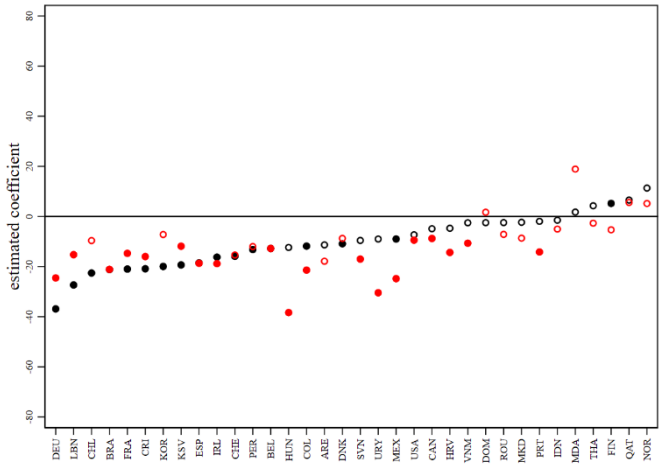
Notes: Graphs display estimates of parent education gaps for schools without (black) and with (red) autonomy. Parent education gap is the difference between the average students with a mother who did not attend college and the average students whose mother did attend college, *ceteris paribus*. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Figure 6: Estimated gender gap in PISA scores for autonomous schools with and without accountability

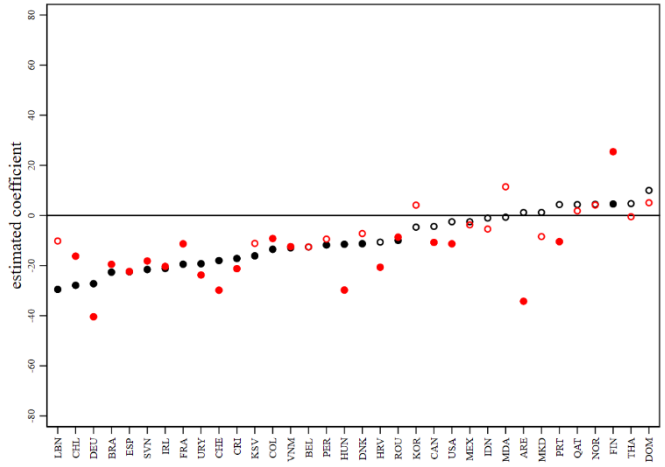
(a) Hiring autonomy



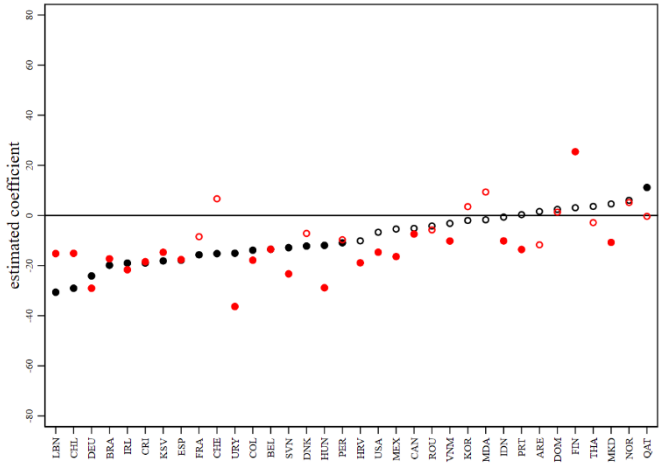
(b) Budget autonomy



(c) Content autonomy



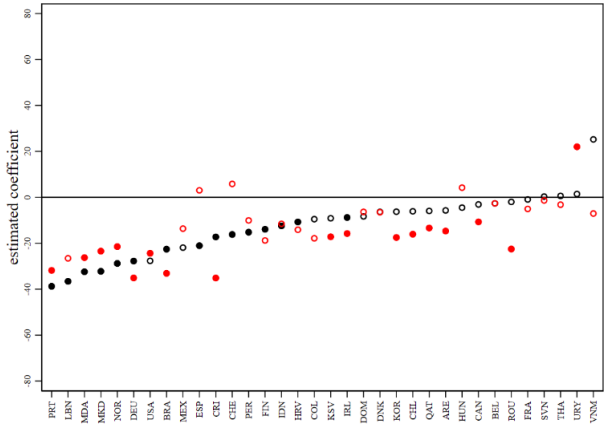
(d) Admission autonomy



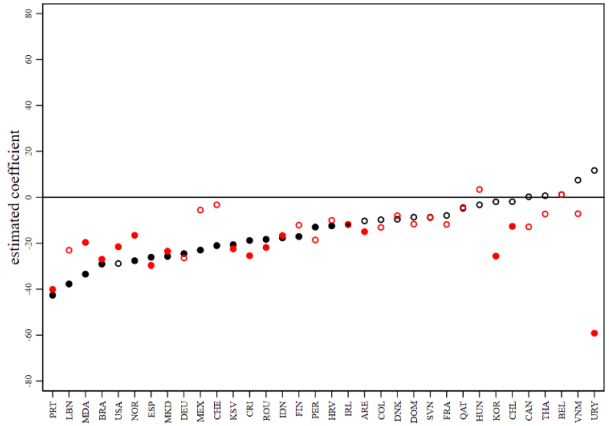
Notes: Graphs display estimates of gender gaps for autonomous schools without (black) and with (red) accountability. Gender gap is the difference between average female scores and average male scores, *ceteris paribus*. Accountability is defined by a requirement to make student achievement data publicly available. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Figure 7: Estimated wealth gap in PISA scores for autonomous schools with and without accountability

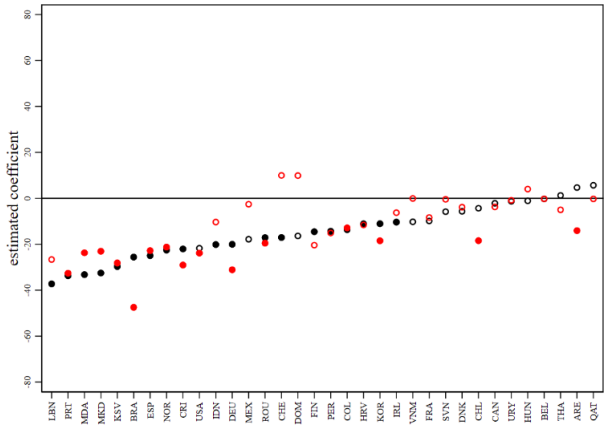
(a) Hiring autonomy



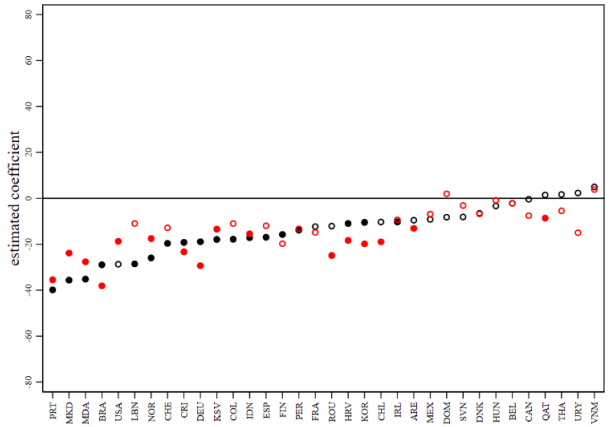
(b) Budget autonomy



(c) Content autonomy



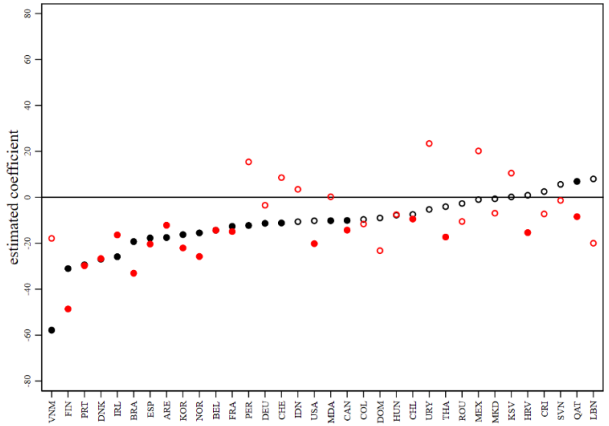
(d) Admission autonomy



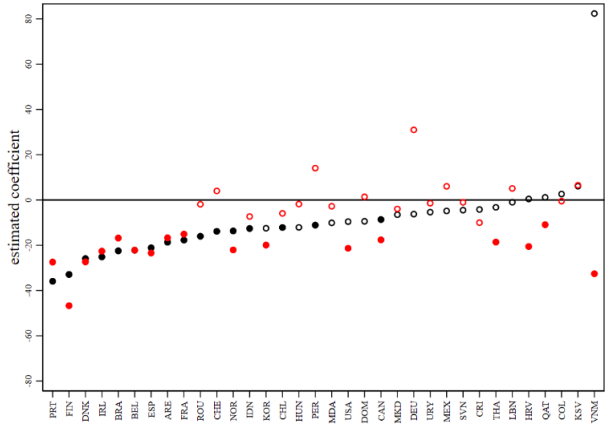
Notes: Graphs display estimates of wealth gaps on test scores for autonomous schools without (black) and with (red) accountability. Wealth gap is the difference in scores between the average student in the lowest PISA wealth quintile and the average student in the top 4 wealth quintiles, ceteris paribus. Accountability is defined by a requirement to make student achievement data publicly available. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Figure 8: Estimated parent education gap in PISA scores for autonomous schools with and without accountability

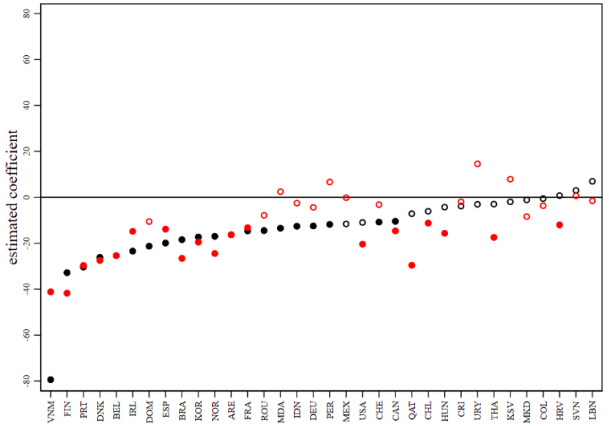
(a) Hiring autonomy



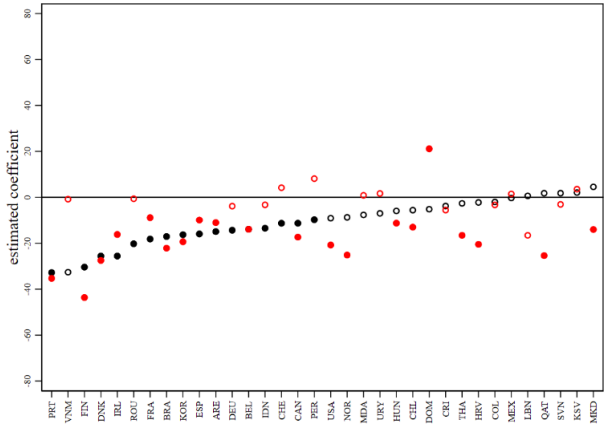
(b) Budget autonomy



(c) Content autonomy



(d) Admission autonomy



Notes: Graphs display estimates of parent education gaps for autonomous schools without (black) and with (red) accountability. Parent education gap is the difference in scores between the average student whose mother did not attend college and the average student whose mother attended college, *ceteris paribus*. Accountability is defined by a requirement to make student achievement data publicly available. Filled circles are statistically significant at 95% or above. All estimates include school controls (the size of the school, whether it is public or private, rural location, the proportion of students living in households at the bottom quintile of the wealth distribution, and the proportion of students whose mother has not attained a college degree).

Table 1: Rates and Correlations for Four Autonomy Indicators

Autonomy	Mean	Correlation Coefficient		
		Hiring	Budget	Course Content
All four types	0.24			
Hiring	0.56			
Budget	0.48	0.33		
Course content	0.70	0.26	0.26	
Admissions	0.58	0.26	0.23	0.25

Notes: Author calculations for PISA school surveys. Sample includes 7,524 schools and 34 countries.

Table 2. Descriptive Statistics for Student and School Characteristics

Variable	Min	Max	Mean	SD
Student variables (n=204,944)				
PISA Math Score	197.74	674.01	441.03	76.44
Age	15.25	16.31	15.79	0.12
Wealth index	-5.86	2.72	-0.54	1.10
Female			0.50	
Non-native			0.11	
Mother did not attend college			0.61	
School variables (n=7,524)				
School size	1	17,805	787.52	779.51
Public school			0.81	
Rural			0.66	
Share of student in wealth quintile 1			0.24	
Share of students whose mother did not attend college			0.59	
Required to publish achievement data			0.40	

Notes: Author calculations from PISA student and school surveys. Samples includes 34 PISA countries with variation in school autonomy.

Table 3: Pooled Regression Estimates of the Effects of Autonomy on PISA Math Scores

Autonomy Indicators	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Hiring	31.46*				-1.20				34.49*	0.79	-3.89	-2.99
	(3.14)				(2.23)				(3.26)	(2.51)	(2.67)	(2.36)
Budget		0.26				4.85*			-7.35*	5.07*	6.90*	4.62*
		(2.89)				(2.07)			(3.11)	(2.52)	(2.49)	(2.25)
Content			8.05*				4.22		6.66	6.90*	5.08	3.44
			(3.18)				(2.46)		(3.45)	(2.77)	(2.62)	(2.60)
Admissions				-7.99*				1.84	-15.29*	-1.93	-0.36	0.97
				(3.13)				(2.08)	(3.24)	(2.61)	(2.52)	(2.19)
Constant	423.77*	440.17*	435.21*	445.71*	438.20*	435.99*	434.15*	436.17*	432.17*	461.52*	528.64*	433.73*
	(1.69)	(1.80)	(2.17)	(2.36)	(27.25)	(27.06)	(27.06)	(26.97)	(2.60)	(29.22)	(29.39)	(27.23)
Individual controls					√	√	√	√		√	√	√
School controls					√	√	√	√			√	√
Country FE					√	√	√	√				√
Observations	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944
R2	0.03	0.00	0.00	0.00	0.41	0.41	0.41	0.41	0.03	0.20	0.23	0.41

Notes: Results based on pooled OLS regressions of students in 34 countries. Standard errors are in parentheses. Individual controls include gender, age, PISA wealth index, nativity, and if the mother attended college. School controls include rural location, publicly funded, enrollment, share of students at the bottom wealth quintile, and share of students whose mothers did not attend college. Standard errors were clustered at the school level, and balanced repeated replication with Fay's adjustment was used as variance method. * $p<0.05$

Table 4. Tabulated Results of Country-Level Estimates of Autonomy Effects with and without Control Variables

(a) Hiring Autonomy

estimated without controls	Positive	Negative	Null	Row Total
	estimated with controls			
Positive	**2**	2	10	14
Negative	0	0	2	2
Null	0	2	16	18
Column Total	2	4	28	34

(b) Budget Autonomy

estimated without controls	Positive	Negative	Null	Row Total
	estimated with controls			
Positive	**1**	0	9	10
Negative	0	0	0	0
Null	1	2	21	24
Column Total	2	2	30	34

(c) Content Autonomy

estimated without controls	Positive	Negative	Null	Row Total
	estimated with controls			
Positive	**1**	0	7	8
Negative	0	**3**	0	3
Null	1	1	21	23
Column Total	2	4	28	34

(d) Admissions Autonomy

estimated without controls	Positive	Negative	Null	Row Total
	estimated with controls			
Positive	**3**	0	6	9
Negative	0	**1**	2	3
Null	1	1	20	22
Column Total	4	2	28	34

Notes: Table counts the number of PISA countries with significant positive coefficients, significant negative, or non-significant (null) coefficients for school autonomy indicators (95% confidence level). Counts in rows are from equation 1, estimating autonomy effects without controls. Counts in columns are from equation 2, estimating autonomy with controls for individual characteristics (gender, age, PISA wealth index, nativity, and if the mother attended college) and school characteristics (rural location, publicly funded, enrollment, share of students at the bottom wealth quintile, and share of students whose mothers did not attend college). Numbers in bold show countries where statistically significant autonomy effects are robust to the includes of covariates.

Table 5: Pooled Regression Estimates for the Effects of Autonomy and Subgroup Membership

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Subgroup	Female				Bottom Wealth Quintile				Mother did not Attend College			
Autonomy Type	Hiring	Budget	Content	Admit	Hiring	Budget	Content	Admit	Hiring	Budget	Content	Admit
Subgroup member	-8.55*	-11.46*	-11.36*	-10.64*	-16.43*	-16.53*	-14.80*	-19.17*	-13.56*	-13.19*	-10.78*	-14.10*
	(0.88)	(0.96)	(1.12)	(1.35)	(1.98)	(1.47)	(1.89)	(1.92)	(1.53)	(1.43)	(2.04)	(1.87)
Autonomy	3.10	2.74	1.03	-0.48	2.23	4.82*	3.47	-0.13	4.23	7.34*	7.18*	1.44
	(2.70)	(2.66)	(2.82)	(2.44)	(2.53)	(2.42)	(2.75)	(2.39)	(3.28)	(2.87)	(3.35)	(3.03)
Subgroup x Autonomy	-1.29	4.28*	3.38*	2.10	-0.75	-0.43	-3.24	3.43	-2.44	-3.68	-6.65*	-1.37
	(1.36)	(1.61)	(1.46)	(1.68)	(2.61)	(2.61)	(2.90)	(2.86)	(2.62)	(2.37)	(2.87)	(2.74)
Constant	495.83*	496.28*	497.54*	498.20*	491.57*	490.45*	491.11*	493.17*	490.26*	489.17*	488.54*	492.19*
	(4.84)	(4.21)	(4.41)	(4.42)	(4.71)	(4.03)	(4.23)	(4.23)	(5.15)	(4.16)	(4.63)	(4.52)
Observations	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944	204,944
R2	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37

Notes: Results based on pooled OLS regressions of student-level data for 34 countries in PISA 2015. Standard errors are in parentheses. All regressions include school controls and country fixed effects. School controls include an indicator for schools located in rural areas, publicly funded, a variable that measures its size, the share of students at the bottom quintile of wealth distribution and the share of students having mothers without a college degree. Standard errors were clustered at the school level, and balanced repeated replication with Fay's adjustment was used as variance method. Statistical significance: *p<0.05

Table 6. Tabulated Results of Country-Level Estimates of Student Performance Gaps for Schools with and without Autonomy

Panel A: Gender Gap
Gap if Male > Female

(a) Hiring Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	18	3	21
	no gap	5	8	13
	total	23	11	

(b) Budget Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	19	4	23
	no gap	2	9	11
	total	21	13	

(c) Content Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	19	3	22
	no gap	2	10	12
	total	21	13	

(d) Admissions Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	19	3	22
	no gap	3	9	12
	total	22	12	

Panel B: Wealth Gap
Gap if Quintiles 2-4 > Quintile 1

(a) Hiring Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	16	6	22
	no gap	8	4	12
	total	24	10	

(b) Budget Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	17	5	22
	no gap	7	5	12
	total	24	10	

(c) Content Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	16	7	23
	no gap	7	4	11
	total	23	11	

(d) Admissions Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	15	5	20
	no gap	8	6	14
	total	23	11	

Panel C: Parent Education Gap
Gap if Mother attended College > Mother had no College

(a) Hiring Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	9	8	17
	no gap	12	5	17
	total	21	13	

(b) Budget Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	14	6	20
	no gap	8	6	14
	total	22	12	

(c) Content Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	15	2	17
	no gap	10	7	17
	total	25	9	

(d) Admissions Autonomy

		schools with autonomy: gap	schools with autonomy: no gap	schools with autonomy: total
schools without autonomy	sig gap	14	3	17
	no gap	9	8	17
	total	23	11	

Notes: Table displays the number of countries with significant PISA performance gaps for schools without autonomy (rows) and schools with autonomy (columns). Gaps are determined based on estimation of B2 and B3 in equation 2. Countries are counted as having a gender gap if boys significantly outperform girls, an income gap if the students in the top 4 quintiles significantly outperform those is the bottom quintile, and parent education gap if those with college-education mothers significantly outperform those without college-educated mothers.

Table 7: Pooled Regression Estimates for the Effects of Accountability and Subgroup Membership at Autonomous Schools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Subgroup	Female				Bottom Wealth Quintile				Mother did not Attend College			
Autonomy Type	Hiring	Budget	Content	Admit	Hiring	Budget	Content	Admit	Hiring	Budget	Content	Admit
Accountability	1.30	7.58*	6.60*	6.26*	0.45	5.77	5.27*	3.74	2.37	4.82	4.33	4.72
	(3.52)	(3.34)	(2.80)	(2.95)	(3.55)	(3.37)	(2.66)	(2.90)	(4.05)	(4.67)	(3.59)	(3.70)
Subgroup member	-9.33*	-6.32*	-7.96*	-6.95*	-17.63*	-17.01*	-19.00*	-16.32*	-14.72*	-16.96*	-17.91*	-14.74*
	(1.74)	(1.62)	(1.43)	(1.30)	(2.40)	(3.76)	(3.16)	(2.84)	(2.04)	(2.59)	(2.04)	(1.97)
Subgroup x accountability	-0.87	-2.08	-0.40	-3.50	0.17	2.24	3.23	1.61	-2.35	2.36	2.98	-0.37
	(2.30)	(2.31)	(2.17)	(2.12)	(3.24)	(4.51)	(3.81)	(3.68)	(3.07)	(4.15)	(3.41)	(3.32)
Constant	508.62*	502.95*	513.73*	501.24*	506.14*	500.62*	512.28*	498.59*	503.26*	500.36*	511.16*	497.53*
	(5.27)	(5.92)	(5.42)	(5.40)	(5.29)	(5.82)	(5.39)	(5.31)	(5.27)	(5.92)	(5.66)	(5.62)
Observations	116,770	96,761	117,488	139,255	116,770	96,761	117,488	139,255	116,770	96,761	117,488	139,255
R2	0.32	0.37	0.38	0.38	0.33	0.38	0.38	0.38	0.33	0.38	0.38	0.38

Notes: Results based on pooled OLS regressions of student-level data for 34 countries in PISA 2015. Standard errors are in parentheses. Sample is restricted to schools that report having autonomy. All regressions include school controls and country fixed effects. School controls include an indicator for schools located in rural areas, publicly funded, a variable that measures its size, the share of students at the bottom quintile of wealth distribution and the share of students having mothers without a college degree. Standard errors were clustered at the school level, and balanced repeated replication with Fay's adjustment was used as variance method. Statistical significance: * $p<0.05$

Table 8. Tabulated Results of Country-Level Estimates of Student Performance Gaps for Autonomous Schools with and without Accountability

Panel A: Gender Gap
Gap if Male > Female

(a) Hiring Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	12	5	17
	no gap	7	10	17
	total	19	15	

(b) Budget Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	11	5	16
	no gap	8	10	18
	total	19	15	

(c) Content Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	13	4	17
	no gap	6	11	17
	total	19	15	

(d) Admissions Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	13	4	17
	no gap	8	9	17
	total	21	13	

Panel B: Wealth Gap
Gap if Quintiles 2-4 > Quintile 1

(a) Hiring Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	8	7	15
	no gap	8	11	19
	total	16	18	

(b) Budget Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	10	8	18
	no gap	5	11	16
	total	15	19	

(c) Content Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	12	7	19
	no gap	3	12	15
	total	15	19	

(d) Admissions Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	11	7	18
	no gap	5	11	16
	total			

Panel C: Parent Education Gap
Gap if Mother attended College > Mother had no College

(a) Hiring Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	12	5	17
	no gap	5	12	17
	total	17	17	

(b) Budget Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	11	5	16
	no gap	6	12	18
	total	17	17	

(c) Content Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	13	7	20
	no gap	6	8	14
	total	19	15	

(d) Admissions Autonomy

		schools with accountability		
		gap	no gap	total
schools without acctability	sig gap	11	5	16
	no gap	8	10	18
	total	19	15	

Notes: Table displays the number of countries with significant PISA performance gaps for autonomous school without accountability (rows) and autonomous schools with accountability (columns). Gaps are determined based on estimation of B2 and B3 in equation 4. Countries are counted as having a gender gap if boys significantly outperform girls, an income gap if the students in the top 4 quintiles significantly outperform those is the bottom quintile, and parent education gap if those with college-education mothers significantly outperform those without college-educated mothers.

Appendix

Table A1. List of Countries included in Analytic Sample

Country	Number of Students in PISA 2015	World Bank Income Classification
East Asia and Pacific		
Indonesia	5,581	Lower middle income
Korea	5,395	High income
Thailand	7,230	Upper middle income
Vietnam	5,599	Lower middle income
Europe and Central Asia		
Belgium	2,536	High income
Croatia	5,500	High income
Denmark	5,130	High income
Finland	5,464	High income
France	4,970	High income
Germany	3,723	High income
Hungary	4,914	High income
Ireland	4,671	High income
Kosovo	4,518	Upper middle income
Moldova	4,644	Upper middle income
North Macedonia	4,072	Upper middle income
Norway	4,012	High income
Portugal	6,547	High income
Romania	4,633	Upper middle income
Slovenia	5,580	High income
Spain	5,765	High income
Switzerland	4,920	High income
Latin America and the Caribbean		
Brazil	13,216	Upper middle income
Chile	5,618	High income
Colombia	8,770	Upper middle income
Costa Rica	6,184	Upper middle income
Dominican Republic	3,560	Upper middle income
Mexico	6,707	Upper middle income
Peru	6,255	Upper middle income
Uruguay	5,674	High income
Middle East and North Africa		
Lebanon	2,872	Upper middle income
Qatar	10,328	High income
United Arab Emirates	9,924	High income
North America		
Canada	15,457	High income
United States	4,975	High income
Total	204,944	