



Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA

Eric A. Hanushek
Susanne Link
Ludger Woessmann

CESIFO WORKING PAPER NO. 3648
CATEGORY 5: ECONOMICS OF EDUCATION
NOVEMBER 2011

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA

Abstract

Decentralization of decision-making is among the most intriguing recent school reforms, in part because countries went in opposite directions over the past decade and because prior evidence is inconclusive. We suggest that autonomy may be conducive to student achievement in well-developed systems but detrimental in low-performing systems. We construct a panel dataset from the four waves of international PISA tests spanning 2000-2009, comprising over one million students in 42 countries. Relying on panel estimation with country fixed effects, we identify the effect of school autonomy from within-country changes in the average share of schools with autonomy over key elements of school operations. Our results show that autonomy affects student achievement negatively in developing and low-performing countries, but positively in developed and high-performing countries. These results are unaffected by a wide variety of robustness and specification tests, providing confidence in the need for nuanced application of reform ideas.

JEL-Code: I200, O150, H750, I250.

Keywords: school autonomy, decentralization, developing countries, educational production, international student achievement tests, panel estimation.

Eric A. Hanushek
Hoover Institution
Stanford University
USA – Stanford, CA 94305-6010
hanushek@stanford.edu

Susanne Link
Ifo Institute – Leibniz Institute for Economic
Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
link@ifo.de

Ludger Woessmann
Ifo Institute – Leibniz Institute for Economic Research
at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
woessmann@ifo.de

November 7, 2011

We would like to thank participants at the Workshop on Human Capital and Economic Development at Harvard University, in particular Philippe Aghion, Robert Barro, and Mark Rosenzweig, for helpful comments and discussion. Support from the Asian Development Bank is gratefully acknowledged. Link gratefully acknowledges financial support from the German Science Foundation (DFG) through GRK 801. Woessmann gratefully acknowledges support from the Pact for Research and Innovation of the Leibniz Association.

1. Introduction

Virtually every country in the world accepts the importance of human capital investment as an element of economic development, but this has introduced a set of important policy questions about how best to pursue such investments. Over time, attention has shifted away from simply ensuring access to schooling to an interest in the quality of learning.¹ But this shift has introduced new policy uncertainty since the process of expanding school attainment is better understood than is the process of improving achievement, leaving many countries with limited success after adopting a variety of popular policies. The uncertainty has perhaps been largest in the case of institutional design questions, as the evidence in that area has been thinner and less reliable. This paper focuses on one popular institutional change – altering the degree of local school autonomy in decision-making – and brings a new analytical approach to the analysis of its impact.² By introducing cross-country panel analysis, we can exploit the substantial international variation in policy initiatives focused on autonomy. We find that autonomy does significantly affect the performance of a country’s schools, but the impact is quite heterogeneous across stages of development: The effect of school autonomy in decision-making is positive in developed countries, but in fact turns negative in developing countries.

Local autonomy has been a policy discussed intensively in both developing and developed countries. While many countries have moved toward more decentralization in such areas as the hiring of teachers or the choice of curricular elements, others have actually gone to more centralized decision-making. The opposing movements reflect a fundamental tension. The prime argument favoring decentralization is that local decision-makers have better understanding of the capacity of their schools and the demands that are placed on them by varying student populations. This knowledge in turn permits them to make better resource decisions, to improve the productivity of the schools, and to meet the varying demands of their local constituents. Yet,

¹ Hanushek and Woessmann (2008) show that cognitive skills can have substantial impacts on economic development. At the same time, access and attainment goals dominate many policy discussions. The clearest statement of school attainment goals can be found in discussions of the Education for All Initiative of the World Bank and UNESCO (see the description in http://en.wikipedia.org/wiki/Education_For_All, accessed July 31, 2011) and the Millennium Development Goals of the United Nations (see the description in http://en.wikipedia.org/wiki/Millennium_Development_Goals, accessed July 31, 2011). In both instances, while there is some discussion of quality issues, the main objective has been seen as providing all children with at least a lower secondary education.

² Local autonomy for decision-making is referred to in various ways including decentralized decision-making and site-based or school-based management. Here, we typically use the term local or school autonomy, although we think of it as a synonym for these alternative names.

countervailing arguments, centered on lack of decision-making capacity and conflicting incentives, push in the opposite direction. With local autonomy comes the possibility that individual schools pursue goals other than achievement maximization and a potential threat to maintaining common standards across the nation.

Bruns, Filmer, and Patrinos (2011) provide a thoughtful review of the topic in developing countries, focusing on several evaluations of recent decentralization programs, mostly conducted by the World Bank. Their review highlights how methodological approaches – including the use of random control trials, various instrumental-variable applications, and more standard cross-sectional evaluations – influence the program evaluations.³ Similarly, reviews of the experiences of the U.S. tend to find mixed results (Summers and Johnson (1996)).⁴ Despite mixed evidence, there remains considerable policy support for further local autonomy in decision-making (e.g., Ouchi (2003); World Bank (2004); Governor’s Committee on Educational Excellence (2007)).⁵

From an analytical viewpoint, four significant issues arise. First, the very concept of local decision-making and local autonomy is multifaceted and difficult to measure on a consistent basis. It is possible, for example, for local schools to decide some things – such as teacher hiring or facility upgrades – and not others such as the appropriate outcome standards or the pay of teachers. Conceptually, some decisions are more appropriately made locally – e.g., operational decisions like hiring and budget allocations where local knowledge is needed and standardization is not crucial – than others where standardization may be more desirable – e.g., course offerings and requirements (see Bishop and Woessmann (2004)).

Second, the impact of autonomy may well vary with other elements of the system. For example, local autonomy permits using localized knowledge to improve performance, but it also opens up the possibility for more opportunistic behavior on the part of local school personnel. As a result, the impact on student outcomes may well interact with the level of accountability, because centralized accountability provides a way of monitoring local behavior (Woessmann

³ For studies of autonomy in developing countries, see also, e.g., Galiani, Gertler, and Schargrodsy (2008) and the references therein, Behrman and King (2001), Jimenez and Sawada (1999), and Vegas (1999).

⁴ For further evidence from developed countries, see, e.g., Barankay and Lockwood (2007) and Clark (2009).

⁵ In analyzing governance aspects at the level of tertiary education, Aghion et al. (2010) show that autonomy is positively related to universities’ research output in the U.S. and in Europe and argue for benefits from combining autonomy with accountability.

(2005)).⁶ In a larger sense, the results of autonomy may depend on the performance level (Mourshed, Chijioke, and Barber (2010)) and – as a corollary – on the overall development level of the country and the entire school system.

Third, much of the evidence on autonomy comes from cross-sectional analyses where any effects are not well identified.⁷ Specifically, one must often question whether observable characteristics adequately describe differences in schools that are and are not granted more autonomy in decision-making. For example, if more dynamic schools get greater autonomy or if demanding parents choose autonomous schools, it is difficult to extract the independent effect of local decision-making on student achievement.

Fourth, many aspects of the locus of decision-making are set at the national level. For example, many countries set national educational standards, national assessments and accountability regimes, and various rules about what decisions are permissible at the local level, leaving little to no within-country variation in decision-making authority. Relatedly, any general-equilibrium effects are extremely difficult to disentangle if, for example, the pattern of local decision-making brings a competitive response from schools without local decision-making or if the nature of local decisions alters the supply of teachers or administrators. But dealing with these issues through international comparisons – where institutional variation can be found – brings other identification issues related to variations in culture, governmental institutions, and other things that are difficult to measure.

This paper introduces new international panel data to address each of these issues. We develop a panel of international test results from the Programme for International Student Assessment (PISA), covering 42 countries and four waves that span a time period of ten years.⁸ While we estimate our micro models with over one million student observations to account for

⁶ Such considerations have also entered into the interpretation of mixed results from autonomy in the U.S. (see Hanushek (1994); Loeb and Strunk (2007)). A further U.S. example comes from charter schools, which depend significantly on the regulatory environment they face. Charter schools are publicly financed and regulated schools that are allowed to have considerable autonomy, frequently being stand-alone schools. At least a portion of the variation in the evaluations of charter schools probably reflects interactions with other forces such as degree of parental choice, the quality of information, and constraints on school location. For estimates of the variation in charter outcomes, see CREDO (2009), Hanushek et al. (2007), Booker et al. (2007), and Bifulco and Ladd (2006).

⁷ Note that more recent investigations, particularly in developing countries, have relied on randomized control trials – although these are difficult to implement and a number have not been well executed (Bruns, Filmer, and Patrinos (2011)).

⁸ For a discussion of international assessments along with background material for this analysis, see Hanushek and Woessmann (2011).

family and school inputs at the individual level, the panel character of the analysis is at the country level. The survey information that accompanies the student assessments provides rich detail about individual students and schools along with specific descriptions of the decisions that are and are not permissible at the school level. While a variety of studies have looked at some of these institutional features across countries, all prior work has been purely cross-sectional (see Hanushek and Woessmann (2011)).⁹ We identify autonomy effects by exploiting country-level variation over time after including country (and year) fixed effects that control for systematic, time-invariant cultural and institutional differences at the country level.¹⁰ This framework ensures that our estimates are not affected by within-country selection into autonomy and by unobserved country-specific heterogeneity. Within this framework, we can readily test the heterogeneous effects of autonomy by specific types of decisions; by variations in development levels and educational performance levels; and by the existence of centralized accountability.

Our central finding is that local autonomy has an important impact on student achievement, but this impact varies systematically across countries, depending on the level of economic and educational development. In simplest terms, countries with otherwise strong institutions gain considerably from decentralized decision-making in their schools, while countries that lack such a strong existing structure may actually be hurt by decentralizing decision-making. The negative effect in developing countries emerges most clearly for autonomy in areas relating to academic content, but also appears for autonomy in the areas of personnel and budgets. An extensive series of robustness and specification tests corroborate the central finding.

We use the income level of a country (GDP per capita) as an indicator of overall skills and institutions. Higher-income countries tend to have better societal and economic institutions that promote productivity, societal vision, and smooth social interactions. As such, this indicator is broad and multifaceted, leading us also to investigate more specific and nuanced aspects of institutions. We find indications that the development of the educational system (measured by higher achievement) adds another significant dimension to the success of greater local autonomy. Further, consistent with the underlying motivation for constraining opportunistic behavior, the benefits of greater autonomy are enhanced by accountability through centralized examinations.

⁹ For examples of existing investigations of institutions – and particularly of autonomy – across countries, see Woessmann (2003, 2005), Fuchs and Woessmann (2007), and Woessmann et al. (2009).

¹⁰ An early discussion of the underlying concept can be found in Gustafsson (2006).

At a methodological level, the results show the potential perils of cross-country analyses that cannot control for other institutional and development factors. In our specific analysis, we find different and conflicting results between simple cross-sectional analysis (albeit with extensive controls of measured family and schooling inputs) and our new panel estimators. Further, the heterogeneity of results across different levels of development suggests caution in attempting to generalize from developed-country analyses to developing countries (and vice versa).

The next section discusses the underlying conceptual framework. Section 3 describes the new database and key variation across countries in various kinds of local autonomy. Section 4 develops our empirical model. Section 5 presents our estimation results and extensive robustness and specification tests. Section 6 expands the investigation of interactions to centralized examinations and the performance level of the education system. Section 7 concludes.

2. Conceptual Framework

Substantial research has gone into understanding the determinants of educational achievement. This work, generally under the label of educational production functions, has incorporated a wide range of studies designed to understand how such factors as school resources and family background affect achievement.¹¹ Here we take an expanded view of this approach that highlights the importance of institutions and, in particular, of local autonomy.

A typical formulation of an educational production function has student outcomes (T) as a function of family (F) and schools (S):

$$(1) \quad T = f(F, S)$$

Here, however, we introduce the simple idea that the productivity of any input is directly related to the institutional structure of country c (I_c) that determines the basic environment and rules of schools, how decisions are made, the overall incentives in the system, and so forth:

$$(2) \quad T = I_c \cdot f(F, S)$$

For many analyses of educational production within countries, the institutional structure is constant, and analyses that ignore it provide accurate information about the impacts of resources

¹¹ See Hanushek (2002, 2003) on the general framework and U.S. evidence; see Woessmann (2003) on international evidence.

even if these might not transfer well to institutional structures in other countries. In many ways, I_c is similar to total factor productivity in a macro context where it determines the efficiency with which any given set of inputs is translated into student achievement. In this formulation, we are specifically interested in investigating the decision-making institutions of different countries.

School autonomy or the decentralization of decision-making power can be understood as the delegation of a task by a principal (in this case, the government agency in charge of the school system), who wishes to facilitate the provision of knowledge, to agents, namely the schools (see Woessmann (2005)). In the absence of divergent interests or asymmetric information, agents can be expected to behave in conformity with system objectives. In fact, economic models of school governance often suggest that greater autonomy can lead to increased efficiency of public schools (e.g., Hoxby (1999); Nechyba (2003)), because autonomy offers the possibility of using superior local knowledge, with positive consequences for outcomes. Additionally, by bringing decisions closer to the interested local community, decentralization may improve the monitoring of teachers and schools by parents and local communities (see Galiani, Gertler, and Schargrodsky (2008) and the references therein).

However, when divergent interests and asymmetric information are present in a decision-making area, agents have incentives and perhaps substantial opportunities to act in their own self-interest with little risk that such behavior will be noticed and sanctioned. In this case, autonomy opens the scope for opportunistic behavior, with negative consequences for outcomes (Woessmann (2005)). Agents may use their greater autonomy to further goals other than advancing student achievement. Furthermore, the quality of decision-making may also be inferior at the local level when the technical capabilities of local decision-makers to provide high-quality services are limited and when local communities lack the ability to ensure high-quality services (see Galiani, Gertler, and Schargrodsky (2008)). Consequently, the success of autonomy reforms may depend on the general level of human capital which affects the quality of parental monitoring.¹²

Against the background of these opposing sets of mechanisms of how autonomy affects performance, we argue that the impact of autonomy likely depends on the level of development. In general terms, a country's development level captures such aspects as local capacity, abilities

¹² While we focus on issues of decision-making, there may also be technological differences. Centralization opens the possibility to exploit economies of scale, for example in evaluation, teacher training systems, and the like.

of local decision-makers, governance effectiveness, state capacity, parental human capital, and monitoring abilities of local communities. Also specifically in the education system, systems that already work at a high performance level may have such features as external evaluations and well-trained teachers that facilitate local decision-making by setting and ensuring high educational standards.¹³ In particular, accountability systems may limit the extent to which local decision-makers can act opportunistically without getting caught (Woessmann (2005)).

In sum, there are a number of channels through which a higher level of development, both in the education system and in society more generally, strengthens the positive mechanisms of autonomy and weakens the negative ones. The impact of autonomy on performance thus depends on the level of development, being positive in well-functioning systems but possibly even negative in dysfunctional systems.

Finally, the impact of local autonomy may differ by area of decision-making. While standardization may be important in decisions on academic content, it may not be as important in decisions on process operations and personnel-management (Bishop and Woessmann (2004)). Thus, local decision-making over basic issues of standards such as course offerings or course content might have a negative effect of autonomy when the whole system is dysfunctional. But even in such a system, local decision-making over hiring teachers and budget allocations may not be as negative.

3. International Panel Data

An essential component of our analytical strategy is the construction of a cross-country panel of student achievement data. For this, we can take advantage of the recent expansion of international assessments (cf. Hanushek and Woessmann (2011)).

3.1 Building a PISA Panel Database

Our empirical analysis relies on the Programme for International Student Assessment (PISA), an internationally standardized assessment conducted by the Organisation for Economic Co-operation and Development (OECD). The PISA study, first conducted in 2000, is designed

¹³ For example, in diagnosing what leads to improved performance at different stages of development, Mourshed, Chijioke, and Barber (2010) observe that going from ‘great to excellent’ is such that “the interventions of this stage move the locus of improvement from the center to the schools themselves” (p. 26).

to obtain internationally comparable data on the educational achievement of 15-year-old students in math, science, and reading.

Four distinct assessments have been carried out: in 2000/2002, 2003, 2006, and 2009. In PISA 2000, 32 countries, including 28 OECD countries, participated in the assessment. In 2002, a further 11 non-OECD countries administered the PISA 2000 assessment. By PISA 2009, the latest assessment, the number of participating countries reached 65 countries including a range of emerging economies.

PISA's target population is the 15-year-old students in each country, regardless of the institution and grade they currently attend. The PISA sampling procedure ensures that a representative sample of the target population is tested in each country. Most countries employ a two-stage sampling technique. The first stage draws a random sample of schools in which 15-year-old students are enrolled, where the probability of a school to be selected is proportional to its size as measured by the estimated number of 15-year-old students attending. The second stage randomly samples 35 students of the 15-year-old students in each of these schools, with each 15-year-old student having the same sampling probability.

The performance tests are paper and pencil tests, lasting up to two hours for each student. The PISA tests are constructed to test a range of relevant skills and competencies. Each subject is tested using a broad sample of tasks with differing levels of difficulty to represent a comprehensive indicator of the continuum of students' abilities. The performance in each domain is mapped on a scale with a mean of 500 test-score points and a standard deviation of 100 test-score points across the OECD countries.¹⁴

In addition to the achievement data, PISA also provides a rich array of background information on each student and her school. Students are asked to provide information on personal characteristics and their family background. School principals provide information on the schools' resource endowment and institutional settings. While some questionnaire items, such as the questions on student gender and age, remain the same in each assessment cycle, some information is not available or directly comparable across all PISA waves.

¹⁴ While the reading test has been psychometrically scaled on a uniform scale since 2000, the math test was re-scaled in 2003 (and the science test in 2006) to have again mean 500 and standard deviation 100 across the OECD countries and has a common psychometric scale since then. In our analyses below, year fixed effects take account of this. Furthermore, we show that results are qualitatively the same when restricting the math analysis to the waves since 2003 that have a common psychometric scale.

By merging the four PISA assessment cycles, we are, for the first time, able to construct a panel dataset at the country level. In a first step, we combine students' test scores in math, science, and reading literacy with individual students' characteristics, family background information, and school-level data for each of the four PISA waves. Since the background questionnaires are not fully standardized, in a second step we select a set of core variables that are available in each of the four PISA waves and merge the cross-sectional data of 2000/2002, 2003, 2006, and 2009 into one dataset.

Our sample comprises all countries that participated in at least three of the four PISA waves.¹⁵ Combining the available data, we construct a dataset containing 1,042,995 students in 42 countries. As is evident from Table 1, the panel includes a broad sample of both high-income and lower-income countries. Following the World Bank classification, 25 countries in our sample are classified as high-income countries. But there is also one low-income country, seven lower-middle-income countries, and nine upper-middle-income countries in the sample. Figure 1 depicts the available achievement data for the 42 countries in our sample. The average test performance across all countries in the sample hardly changed between 2000 and 2009 (see also Table 1). But some countries saw substantial increases in average achievement (most notably Brazil, Luxembourg, Chile, Portugal, Mexico, and Germany with increases surpassing one quarter of a standard deviation), while others saw substantial decreases (mostly notably the United Kingdom and Japan with decreases surpassing one quarter of a standard deviation).

To have a complete dataset of all students with performance data, we imputed country-by-wave means for missing values of control variables. Since we consider a large set of explanatory variables and since a portion of these variables is missing for some students, dropping all student observations with missing values would result in substantial sample reduction. To ensure that imputed data are not driving our results, all our regressions include an indicator for each variable with missing data that equals one for imputed values and zero otherwise.

We combine the student and school data with additional country-level data. GDP per capita, measured in current US\$, is provided by the World Bank and OECD national accounts data files. Data on annual expenditure per student in lower secondary education in 2000, 2003, and 2006 are taken from the OECD Education at a Glance indicators (see Organisation for Economic Co-

¹⁵ France had to be excluded from the analysis because it provides no information on the school-level questionnaire. Due to their small size, Liechtenstein and Macao were also dropped.

operation and Development (2010)). Data on the existence of curriculum-based external exit exams are an updated version of the data used by Bishop (2006).

3.2 Measuring School Autonomy

We construct our measures of school autonomy for each country from the background questionnaires of the four PISA studies. In all waves, principals were asked to report the level of responsibility for different types of decisions regarding the management of their school. We make use of six decision-making types: 1. Deciding which courses are offered; 2. Determining course content; 3. Choosing which textbooks are used; 4. Selecting teachers for hire; 5. Establishing teachers' starting salaries; and 6. Deciding on budget allocations within the school.

In 2000 and 2003, principals were asked, "In your school, who has the main responsibility for ...?" For each of the enumerated areas, principals had to tick whether decisions were mainly a responsibility of the school's governing board, the principal, department heads, or teachers as opposed to not being a responsibility of the school. Similarly, in 2006 and 2009, principals were asked who has a considerable responsibility for the enumerated tasks and had to choose whether the regional or national education authority as opposed to the principal or teachers had considerable responsibility.¹⁶ In all four waves, respondents were explicitly allowed to tick as many options as appropriate in each area.

For each area, we begin by constructing a variable indicating full autonomy at the school level, which equals one if a school entity – the principal, the school's board, department heads, or teachers – is the only one to carry responsibility (and zero otherwise). Thus, as soon as responsibility is also carried by external education authorities, we do not classify a school as autonomous. (As part of the robustness checks below, for each area we also construct a variable indicating whether the school has any influence on the decision-making process as opposed to exercising full responsibility.) Then, because our interest is focused on countries' institutional structures, we aggregate across all schools in a country to obtain the share of schools with full autonomy in each of the areas. As will be made explicit in the next section, we do not use the individual school measures of autonomy in the modeling of achievement because of concerns about introducing selection bias and because of the possibility of general-equilibrium effects.

¹⁶ See Table A2 in the appendix for an overview of the answer options and a discussion on their comparability across the PISA waves.

Figure 2 shows exemplary graphs across the four waves of aggregate autonomy for determining courses offered and hiring in each country. While many countries have rather flat profiles of autonomy over time, there are also clear movements that differ between the two autonomy areas. For example, among low-achieving countries, Brazil, Chile, and Mexico have seen strong reductions in course autonomy, but smaller reductions (or even increases) in hiring autonomy. Similarly, among medium-achieving countries, Greece, Portugal, and to a lesser extent Turkey have reduced course autonomy, but this is not the case for hiring autonomy in Portugal and Turkey. At a higher level of achievement, Germany has increased school autonomy, particularly in course offerings, whereas countries such as Great Britain, Australia, Denmark, Ireland, and Sweden have all seen slight decreases in the autonomy measures.

Table 2 presents correlations among the six autonomy areas, both in their 2009 levels and in their difference between 2000 and 2009 (which provides the main source of identification in our analysis). Obviously, the three autonomy areas on decisions that are related to academic content – namely courses offered, course content, and textbooks used – are highly correlated among each other, both in levels and in changes. Also, the two autonomy areas on personnel decisions – hiring teachers and establishing their starting salaries – are strongly related. As a consequence, we combine the three variables of courses offered, course content, and textbooks used into one category of autonomy regarding academic content by using their arithmetic mean. Similarly, the mean of hiring teachers and establishing their starting salaries represents our measure of autonomy in personnel decisions.¹⁷ Since autonomy on budget allocations is not correlated with any of the other autonomy areas (apart from the personnel areas when considered in differences rather than levels), we retain it as a separate third autonomy category.

3.3 Descriptive Statistics

Table 1 presents country-level means of the three autonomy measures, as well as mean PISA math scores, in 2000 and 2009. Throughout the paper, our analysis focuses on mathematical literacy, which is generally viewed as being most readily comparable across countries; however, we also report main results in reading and science. Table A1 in the appendix reports pooled international descriptive statistics for all variables employed in the analysis.

¹⁷ Results are very similar if, rather than using the mean across the autonomy categories, we use the share of schools in a country that have autonomy in two or three of the subcategories of the combined variables. In the appendix, we also report results for the six separate autonomy categories.

Table 1 also shows a country's GDP per capita in 2000, our main measure of initial level of development. Figure 3 plots this measure of initial economic development against initial educational achievement, measured as the PISA math score in 2000. There is a strong relation between the initial levels of economic and educational development, which we will further explore below. Most importantly, the figure visualizes where different countries stand on these measures of initial development, which is informative for our analysis of heterogeneity across initial country situations below.

From Figure 1, we can assess the development of PISA math test scores across waves for all 42 countries. Among the low-performing countries with initial test scores below 400 points, Brazil, Chile, Mexico, and moderately Tunisia managed to increase their test scores over time, whereas Argentina's and Indonesia's achievement is mostly flat. Within the group of medium performers, Greece, Italy, Israel, Portugal, and Turkey show a slightly positive trend, whereas Thailand followed a slight downward trend. Among the countries with initially relatively high scores, only Germany shows a consistent upward trend, whereas Great Britain and Japan, and to a lesser extent Australia, Austria, Denmark, Ireland, New Zealand, and Sweden, show a downward trend. The other countries are mostly flat.

Comparing these achievement trends to the autonomy trends seen in Figure 2, there are many examples where the combined achievement and autonomy trends are consistent with increased autonomy, particularly over academic content, being bad in low-performing but good in high-performing countries. For example, starting at a low level of achievement, the increasing achievement levels of Brazil, Chile, and Mexico are accompanied by reductions in autonomy of their schools in particular over course offerings. Similarly, Greece, Portugal, and Turkey have reduced their course autonomy and slightly increased their achievement. By contrast, Thailand – which had quite flat autonomy – saw mostly flat achievement. Finally, at a higher level of initial achievement, Germany's increased autonomy, particularly over course offerings, goes along with consistent increases in achievement. Great Britain, Australia, Denmark, Ireland, and Sweden all slightly reduced their autonomy, which is mirrored by slightly decreasing achievement.

4. Empirical Model

To test the effect of autonomy on student achievement and its dependence on a country's development level more formally, we make use of the education production function framework

introduced above. The empirical issues can be most easily seen from a simple linear formulation which now introduces a time dimension to the analysis:

$$(3) \quad T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \varepsilon_{cti}$$

where achievement T in country c at time t for student i is a function of a country's institutions I (here autonomy), the inputs from a student's family (F) and from schools (S), and an error term, ε_{cti} . We start our exposition with a linearized and additive version of the model, but our analyses below will test for rich multiplicative interactions of the institutional effect with other input factors. Our interest is estimating $\alpha = \partial T / \partial I$, the impact of local autonomy on achievement holding constant other inputs. For this, we have the panel data from PISA that has individual-level data about T , F , and S and data about institutions I aggregated at the country level.

Our approach to identify the impact of institutions is best seen by expanding the error term:

$$(4) \quad \varepsilon_{cti} = \eta_c + \eta_{ct} + \eta_{cti}$$

where η_c is a time-invariant set of cultural and educational factors for country c (such as awareness of the importance of education, the commitment of families to their children's education, or more generally the state of development of societal and economic institutions); η_{ct} is a time-varying set of aggregate educational factors for country c (such as changes in spending levels or private involvement); and η_{cti} is an individual-specific, time-varying error.

The key to identification of α , the parameter of interest, is that ε_{cti} is orthogonal to the included explanatory factors and, importantly, to the measure of local autonomy. The formulation in equation (4) shows the main elements of our approach. First, at the individual student and school level, there are concerns about selection bias, reflecting unmeasured attributes of schools or students in circumstances with varying local decision-making.¹⁸ If, for example, particularly good students are attracted to schools with more local autonomy, η_{cti} would tend to be correlated with I , leading to bias in the estimation of α . But, by aggregating over all schools in the country and measuring autonomy by the proportion of schools with local autonomy, we eliminate the selection bias from school choice. The aggregation also allows us to capture any general-equilibrium effect whereby, for example, autonomy of one school may elicit competitive responses from schools that do not have autonomy themselves.

¹⁸ These concerns are central to the interpretation of most within-country analyses of decentralization.

Second, with the panel data, we can include country fixed effects, μ_c , which effectively eliminate any stable country-specific factors contained in η_c ,¹⁹

$$(5) \quad T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \mu_c + \mu_t + \varepsilon_{cti}$$

By implication, the estimation of α is based upon variations in autonomy over time, since time-invariant institutional features are absorbed into the country fixed effect. The relevant variation with which we estimate α is within-country changes for our sample of PISA countries.

The most significant remaining issue is whether there are time-varying country factors (η_{ct}) that are correlated with the pattern of local autonomy in the country. The underlying identifying assumption is that there are no educationally important time-varying country factors that are correlated with variation in the institutional input, I . We will partially test this by including several additional time-varying factors of countries' education systems, C_{ct} , in the analysis:

$$(6) \quad T_{cti} = \alpha I_{ct} + \beta_F F_{cti} + \beta_S S_{cti} + \beta_C C_{ct} + \mu_c + \mu_t + \varepsilon_{cti}$$

Other details are also important. In order to obtain the best estimates of α , we attempt to eliminate as much other variation in test scores as possible by estimating the β parameters for family and school effects on a large set of individual measures and by conducting the estimation at the individual student level. Additionally, the limited variation in institutional factors – which occurs at the country level – means that it is hard to simultaneously estimate measures of alternative forms of local decision-making. As a result, most of our analysis sequentially estimates models with combined autonomy measures, although we also report specifications that include several autonomy measures together.

A central component of the analysis is the possibility of significant interactions of institutional factors with other institutions or country-specific elements such as school accountability systems or level of capacity and stage of development. We pursue this parametrically by interacting I , the specific measure of autonomy in each model, with the initial level of development (of the country and/or educational system), D_c :

$$(7) \quad T_{cti} = \alpha_1 I_{ct} + \alpha_2 (I_{ct} \times D_c) + \beta_F F_{cti} + \beta_S S_{cti} + \beta_C C_{ct} + \mu_c + \mu_t + \varepsilon_{cti}$$

¹⁹ The estimation also includes time fixed effects to allow for any common shocks across waves.

In this model, which represents our main specification, the effect of autonomy reforms is allowed to differ depending on the surrounding conditions captured by D_c . We can then test our main conceptual proposition that autonomy is beneficial for student achievement in otherwise well-functioning systems but detrimental in dysfunctional systems.

5. Results

5.1 Main Results

Conventional estimation identifies the effect of autonomy from the cross-sectional variation. For comparison to our identification below, such models are reported in Table 3. A simple pooled cross-section with school autonomy measured at the individual level shows a positive association of the three areas of autonomy with student achievement in math (significant for academic-content and budget autonomy), after controlling for standard measures of family and school background (column 1). There is little indication that this association differs across levels of development, although the positive association of academic-content autonomy seems to increase slightly with a country's development level, measured by the initial GDP per capita in 2000 (column 2). When averaging the autonomy measures at the country level (while keeping all other variables at the individual level) to ignore the within-country variation in autonomy, which may be highly prone to self-selection, the estimates increase substantially (column 3). Again, there is little sign of effect heterogeneity across development levels (column 4).

However, results change dramatically when, consistent with our identification strategy, we focus on within-country changes over time. The cross-sectional association vanishes, with point estimates turning negative, once country fixed effects are added (column 5), where the autonomy effect is now identified from aggregate within-country variation over time. Still, this average effect may hide substantial heterogeneity of the autonomy effect across countries.

Thus, Table 4, which shows our main results, adds an interaction term of autonomy with initial GDP per capita to the panel specification with country fixed effects and with autonomy measured at the country level.²⁰ The results indicate clear evidence of substantial effect heterogeneity for all three areas of autonomy: The autonomy effects become significantly more

²⁰ Table A1 in the appendix shows the coefficients of the control variables in this specification for the academic-content autonomy category.

positive with increasing initial GDP per capita. GDP per capita is centered at \$8,000 (in 2000) in this specification, implying that the main effect reflects the impact of autonomy on student achievement in a country at the upper end of the upper-middle-income category of countries such as Argentina (see Table 1 and Figure 3).

As indicated by the negative main effect, a country near Argentina's level of development that increased its academic-content autonomy over time would expect to see a significant and substantial drop in achievement. In such a country, going from no autonomy to full autonomy over academic content would reduce math achievement by 0.34 standard deviations according to this model. Moreover, the significant positive interaction indicates that the autonomy effect is significantly negative for all low- and middle-income countries in our sample. At the extreme of the poorest country in our sample (Indonesia at \$803 GDP per capita in 2000), the negative effect of academic-content autonomy reaches 0.55 standard deviations (column 3).

By contrast, the effect of academic-content autonomy turns significantly positive in most of the high-income countries. For the richest country in our sample (Luxembourg at \$46,457 GDP per capita in 2000), the positive effect of academic-content autonomy is as large as 0.79 standard deviations (column 4). The level of 2000 GDP per capita at which the autonomy effect switches its sign from negative to positive is \$19,555 (column 2). As is evident from Table A3 in the appendix, this pattern holds separately for all three categories of autonomy – course offerings, course content, and textbooks – contained in the aggregated measure of academic-content autonomy in this table.

As the lower two panels show, the basic pattern of results is quite similar in the other two areas of autonomy – personnel and budget autonomy. The autonomy effect increases significantly with initial GDP per capita, and there is a large and significant positive autonomy effect for rich countries. The only difference from the academic-content autonomy category is that the negative effect in the categories of personnel and budget autonomy is smaller and not statistically distinguishable from zero at the upper end of the upper-middle-income countries. For budget autonomy, the negative autonomy effect does not reach statistical significance for even the poorest country in our sample.

The substantial correlation between the different categories of autonomy limits the extent to which we can distinguish among the three categories, but Table 5 presents models with pairs of two autonomy variables, as well as all three of them, combined. When academic-content

autonomy is included together with the other autonomy categories, only the interaction of academic-content autonomy with initial GDP per capita retains statistical significance. When only personnel and budget autonomy are included, the interaction of initial GDP per capita with personnel autonomy is statistically significant but the interaction with budget autonomy is not. Given the high correlation of academic-content and personnel autonomy (Table 2) and the size of the standard errors, multicollinearity does not allow us to rule out a substantial positive interaction for personnel autonomy. However, given that the correlation of budget autonomy with the other autonomy categories is quite low, these specifications tentatively indicate that budget autonomy has no separate effect once the other autonomy categories are considered.²¹ Therefore, in the remainder of the paper, we focus on the two aggregated measures of autonomy over academic content and over personnel.

5.2 Robustness Tests

Several extended tests confirm the robustness of our main finding. The various modifications for measurement issues and estimation samples leave our basic findings intact.

The first set of robustness tests relates to the measurement of variables. The main results prove quite independent of the specific way in which the interaction with initial GDP per capita is specified. As shown in the first three columns of Table 6, the basic result does not change when initial GDP per capita is not measured linearly, but instead in logs; as a dummy for countries with a GDP per capita higher than \$8,000 (roughly the upper end of the upper-middle-income category of countries in our sample); or as a dummy for countries with higher-than-median GDP per capita in our sample (which is at \$14,000).

Our main model includes measures of school characteristics, but the final columns of Table 6 show that results are robust to alternative treatments of school controls. First, giving autonomy to schools may mean that schools use their autonomy to alter other school characteristics, such as reducing the school size or raising teacher education requirements. Such changes would thus be channels through which school autonomy affects student outcomes. In this perspective, these school measures should not be controlled for in the estimation. As is evident in column 4, leaving the school-level variables out of our basic model does not affect our qualitative results.

²¹ The significant correlation between the *change* in budget and personnel autonomy (panel B of Table 2) suggests that there is still some possibility that multicollinearity is driving the lack of significance.

Second, there may be a concern that other school reforms may have coincided with the autonomy reforms that identify our main result. To capture such other reforms, column 5 includes all school variables measured as country averages, aggregating them to the same level at which the autonomy variables are measured. Despite concerns with statistical power with a large number of country-level variables, the qualitative results for autonomy again remain the same.

Autonomy reforms might also have coincided with expenditure reforms across countries. Because there is no consistent data on expenditure per student for all countries and waves, our basic model does not control for expenditure per student. But for the waves 2000-2006, we have consistent data on annual expenditures per student in lower secondary education for a subset of (mostly OECD) countries. The first column of Table 7 shows that our basic results hold similarly in this subset of country-by-wave observations. Column 2 adds the expenditure variable to this model, and the qualitative results are unaffected. Changes in expenditure per student are actually significantly negatively related to changes in student achievement, which dilutes concerns about the lack of expenditure controls in our basic specification. The coefficient on expenditures may capture forces that push for increased spending but that at the same time lower the efficiency of their use.²²

The other four columns of Table 7 test for robustness in different sub-samples. The PISA math test was scaled to have mean 500 and standard deviation 100 across the OECD countries in 2000 and in 2003 each, and it was designed psychometrically to have a common scale since 2003. Column 3 shows that results are qualitatively unaffected when dropping the 2000 wave and restricting the analysis to the three waves since 2003 in which the tests are psychometrically scaled to be intertemporally comparable.

In order to ensure that the effect is identified only from long-term changes and not driven by short-term oscillations, column 4 restricts the analysis to waves 2000 and 2009. When identified from the nine-year differences in autonomy and test scores, results are even more pronounced than in the four-wave specification.

Our main specification employs an unbalanced sample, as some countries did not participate in all four PISA waves (see Figure 1). Column 5 of Table 7 replicates our analysis for the fully

²² As reviewed in Hanushek and Woessmann (2011), international comparative studies of the impact of expenditures provide mixed results but tend to indicate no consistent relationships between spending and international test scores.

balanced sample of 29 countries with achievement and autonomy data in all four PISA waves. Again, qualitative results are the same. Column 6 restricts the sample to OECD countries, without substantive changes in results.

Additional robustness tests show that results also do not hinge on any specific country being included in the estimation. All results are robust when we drop one country at a time from the estimation sample.²³ In particular, results look very similar when Luxembourg – a slight outlier with the highest GDP per capita (see Figure 3) – is excluded from the sample.

Finally, results are also very similar when we separate the student and country estimations into two steps. In the two-step model, test scores are “cleaned” from impacts of the student- and school-level controls in a first, student-level regression. The residuals of this regression, which capture that part of the test-score variation that cannot be attributed to the controls, are then collapsed to the country-by-wave level. In a second, country-level regression, we use the country-level data to run a “classical” panel fixed-effects model, where the level of observation coincides both with the level of the fixed effects and with the level at which the variables of interest are measured. Results (shown in Table A4 in the appendix) are qualitatively the same as in our preferred one-step specification, and they do not differ depending on whether the model does or does not already include country fixed effects in the estimation of the first step.

5.3 Specification Tests

Our identification derives from country-level variation in autonomy over time and its interaction with initial development levels in a panel model with country fixed effects. To analyze the validity of the specification, we present a set of specification tests that address several possible remaining concerns with the identification and that also indicate possible channels and sources of heterogeneity in the impacts. Given that the tests corroborate our main specification mostly by producing the result of insignificant alternative effects, we simply summarize the findings here. Detailed results are available from the authors upon request.

A first possible concern with identification from panel variation is that variation in autonomy over time may be endogenous to the initial level of student achievement. For example, poor initial achievement might theoretically induce governments to implement decentralization – or centralization – reforms. In order to test for the empirical relevance of this

²³ Detailed results are available from the authors on request.

concern, we estimate several models where the changes in autonomy that identify our results are regressed on initial PISA scores. Thus, we test whether the PISA score in 2000 predicts the change in autonomy from 2000 to 2003 or from 2000 to 2009. We also test whether the PISA level in one cycle predicts the change in autonomy from this to the subsequent cycle in a panel model of the four PISA waves. In all tests, lagged PISA scores do not significantly predict subsequent changes in autonomy, corroborating the identifying assumption of our panel model.

A second possible concern is that the development level may interact not only with autonomy reforms, but also with other education policy measures. In other words, the heterogeneity of impact may not be specific to the dimension of school autonomy, as other policies may also be more effective within a well-functioning surrounding. To investigate this, we included in the regression interactions of initial GDP per capita with country-level measures of several other features of the school system: competition (proxied by the share of privately operated schools), funding sources (share of public funding in the school budget), school size (number of students per school), teacher education (share of certified teachers), and shortage of math teachers. Our results show that none of these variables interacts significantly with initial GDP per capita in determining student achievement, and the autonomy results remain robust when these additional interactions are included in the model.

Third, to investigate whether the heterogeneity of the autonomy effect is specific to the development level and does not capture heterogeneity with respect to other country characteristics, we also estimated specifications that interact autonomy with a number of other country measures. (For interactions specifically with the overall performance of the education system and with accountability, see the next section). Some of these measures may also be interpreted as possible channels through which the level of economic development may matter for the impact of autonomy on student achievement. Specifically, autonomy may interact with the size of a country, as school autonomy may mean different things in small and large countries; with its ethnic homogeneity, as autonomy may work better in homogenous societies; with a country's political regime, corruption level, or governance effectiveness, which may determine restraints on how well autonomy can work; or with a country's culture, which may be more or less complementary to autonomous decision-making. In addition, parental human capital may moderate the quality of local monitoring, their ability to pay for private schooling may affect the incentives of autonomous schools, and autonomous schools may use specific local policies.

Thus, we estimated specifications that interact autonomy with population size; with the Alesina et al. (2003) measure of ethnic fractionalization; with the Polity IV index that measures governing authority on a scale from institutionalized autocracies to consolidated democracy; with the corruption perceptions index of Transparency International; with the Governance Effectiveness Index of the World Bank's Worldwide Governance Indicators project, which aims to capture the perceived quality of public services and of policy formulation and implementation; and with the six Hofstede dimensions of national culture, in particular the measures of individualism versus collectivism (integration into groups) and of power distance (acceptance of power inequality). We also interacted autonomy with average measures of parents' human capital available in the PISA dataset (white collar occupations and books at home), with the share of private funding in the school budget, and with such school aspects as the share of certified teachers, shortages of math teachers, school size, and share of private schools.

In models that enter these interactions separately and do not include the interaction of autonomy with initial GDP per capita, there is an indication that autonomy interacts positively with democracy, government effectiveness, individualism, the share of privately operated schools, and the share of certified teachers, and negatively with population size, corruption, and acceptance of power inequality. However, in all these cases, the significance of the interaction vanishes once the interaction of academic-content autonomy with initial GDP per capita is also entered, and the latter retains statistical significance throughout.²⁴ Thus, while the interaction with the development level clearly entails dimensions of democracy, governance effectiveness, cultural values, and effective school environments, the overall measure of economic development in terms of GDP per capita dominates these other separate interactions. Variations in these other measures that are not correlated with the standard measure of economic development do not interact significantly with the autonomy effect.

Fourth, we test whether the autonomy effect is heterogeneous for students with different individual social backgrounds. Such heterogeneity may reflect another channel of the autonomy effect, as decentralization may work better with sophisticated parents (Galiani, Gertler, and Schargrodsky (2008)). It also provides evidence on the effect of autonomy on inequality, as

²⁴ Results for personnel and budget autonomy are similar, but sometimes less strong. While the negative interaction of autonomy with ethnic fractionalization is insignificant in the separate model, it turns marginally significant in the model that also includes the interaction of autonomy with initial GDP per capita (which is fully robust), indicating that autonomy may work better in ethnically more homogeneous countries.

differential impacts by social background would narrow or widen the performance gap between well-off and disadvantaged families. To test this, we add interaction terms between autonomy and family background measures as well as the triple interaction between autonomy, initial GDP per capita, and the family measures to our basic specification. Our measures of individual family background include parental white-collar occupation, parental university education, books at home, and immigration background. For all four measures, neither the interaction with the autonomy variable nor the triple interaction is statistically significant, and point estimates suggest different directions of effects. Consequently, autonomy reforms do not seem to affect children from different background differently and thus do not seem to magnify or lessen inequality, either in developed or in developing countries.²⁵

5.4 Further Results

While the results so far relate to math achievement, which is most readily tested comparably across countries, PISA also tested students in reading and science. As shown in column 1 of Table 8, results are qualitatively the same in reading. This is particularly interesting because reading scores have been psychometrically scaled to be comparable over all four PISA waves. Results on academic-content autonomy are also found for science achievement, where results on personnel autonomy are less pronounced and lose statistical significance (column 2).

In our analysis so far, we have defined autonomy as a school entity having the sole responsibility for a task. Alternatively, one can consider cases where a school entity has considerable responsibility, but an authority beyond the school has considerable responsibility as well – something that one might term “joint decision-making.” Conceptually, one might expect that both the negative and the positive aspects of autonomy discussed in our conceptual framework might be somewhat limited when an external authority has a joint say on a matter. To test this, we use as an alternative autonomy measure the share of schools in a country that have considerable responsibility on a task but where an external authority may also have a say.

²⁵ We also estimated a specification that adds an interaction of autonomy with the initial Gini coefficient of income inequality, provided by the World Bank. While the interaction of autonomy with the initial per-capita GDP level remains qualitatively unaffected, there is also some indication that academic-content autonomy is more beneficial in more equal societies. However, this pattern is not confirmed by distributional measures of family background taken from the PISA dataset that directly relate to the parents of the tested students.

Column 3 of Table 8 shows that results are considerably weaker for this “joint authority” measure of school autonomy than for the measure of “full” school autonomy used throughout this paper. Both negative and positive effects of autonomy are reduced when external education authorities may also have a say in decision-making. Thus, the main effects of autonomy derive from independent decision-making at the school level.

6. Adding Accountability and Educational Development

The prior analysis presumes that a country’s income level can sufficiently characterize the set of institutional features that are complementary to local autonomy in schools – including, for example, experience with general economic structures, the importance of the rule of law as seen in economic operations, generally functioning governmental institutions, and the like. It has the potential disadvantages of ignoring specific educational institutions and the overall development of the educational sector. For these reasons, we present exploratory estimates of more education-specific features of a country that might provide a more refined look at autonomy.

As described in our conceptual principal-agent framework, the effect of autonomy may not only depend on the level of development, but also on the extent to which a school system directly monitors results through accountability systems. Existing cross-sectional research has found significant interactions of school-level autonomy with country-level existence of the accountability measure of central exit exams across countries (see Woessmann (2005); Hanushek and Woessmann (2011)). Thus, the first column of Table 9 adds an interaction term between autonomy and central exit exams to our basic model. There is a sizeable positive interaction between (time-variant) school autonomy and the (time-invariant) measure of central exit exams, statistically significant in the case of academic-content autonomy. The effect of introducing autonomy is more positive in countries that hold the system accountable by central exit exams. At the same time, our main effect of an interaction between autonomy and level of development is unaffected by including the autonomy-exam interaction. As is evident in column 2, there is no significant triple interaction between autonomy, exams, and initial GDP per capita, suggesting that the impact of the development level on the autonomy effect does not depend on whether there are central exams in the school system, and vice versa.

We have consistently measured the initial level of development by overall economic development (GDP per capita). An alternative way of measuring development is to look at the

achievement level of the education system, which we measure by the initial average PISA score in 2000. As shown in Table 10, the effect of school autonomy indeed increases significantly with the initial achievement level. The negative autonomy effect in poorly performing systems is again larger for academic-content autonomy than for personnel autonomy. For a country at the relatively low initial achievement level of 400 PISA points, equivalent to one standard deviation below the OECD mean, going from no to full school autonomy reduces student achievement by 0.63 standard deviations in academic-content autonomy and by 0.33 standard deviations in personnel autonomy. The coefficient estimates imply that the autonomy effect turns from negative to positive at a performance level of 485 and 449 PISA points, respectively, for academic-content and for personnel autonomy. At the level of the highest-performing country (Hong Kong with a test score of 560.5), the positive effect of academic-content autonomy is as large as 0.56 standard deviations, and 0.72 standard deviations for personnel autonomy.

Column 2 of Table 10 jointly enters the interactions of autonomy with the initial PISA score and with initial GDP per capita. Both retain statistical significance for interactions with academic-content autonomy, while limited statistical power has the two interaction terms shy of statistical significance for personnel autonomy. Initial educational achievement and initial GDP per capita may thus capture two separable dimensions of the performance level of a country that have relevance for how school autonomy affects student outcomes.²⁶

For robustness, the final two columns use alternative forms of measuring initial achievement. In column 3, qualitative results are similar when the initial achievement level is not measured linearly but as a dummy for countries scoring higher than 400 PISA points (one standard deviation below the OECD mean). Similarly, results hold when measuring initial achievement by a dummy for countries scoring higher than the OECD mean of 500 PISA points (column 4). Results are also very similar for a dummy for countries above the sample median of 480 PISA scores (not shown).

We find both of these extensions – accountability and development of the educational system per se – to be highly suggestive of a more nuanced view of autonomy. At the same time, the limitations of our cross-country approach that come from relatively small effective samples of countries and from imperfect measurement of specific institutions lead us to be cautious in the

²⁶ Results are robust to dropping the former Communist countries, which – as seen in Figure 3 – are noteworthy outliers in the plot of initial GDP per capita against initial achievement.

interpretation. We think there are conceptual reasons that lend credence to these results, particularly about accountability, but there are many details about the form and consequences of accountability that are ignored.²⁷

7. Conclusions

Decentralization of decision-making has been hotly debated in many countries of the world, and prior research has left considerable uncertainty about the expected impact of giving more autonomy to schools. In the face of this uncertainty, many countries have changed the locus of decision-making within their countries over the past decade – and interestingly some have decentralized while others have centralized. We exploit this cross-country variation to investigate the impact of local autonomy on student achievement. We identify the effect of school autonomy from within-country changes in the share of autonomous schools over time in a panel analysis with country (and time) fixed effects.

Our central finding is that autonomy reforms improve student achievement in developed countries, but undermine it in developing countries. At low levels of economic development, increased autonomy actually hurts student outcomes, in particular in decision-making areas related to academic content. By contrast, in high-income countries, increased autonomy over academic content, personnel, and budgets exerts positive effects on student achievement. In general, the autonomy effects are most pronounced in decision-making on academic content, with some additional relevance for personnel autonomy and, less so, for budgetary autonomy.

Empirically, the main result proves highly consistent across a series of robustness and specification checks. Among others, the autonomy effects show up in various forms of measuring initial GDP per capita, alternative specifications of the control model, and different sub-samples in terms of included waves and countries. The basic finding of heterogeneity of the impact of autonomy by development level shows up in students' performance in math, in reading, and in science. It is much more pronounced for full school-level autonomy than for joint authority between schools and external authorities.

In terms of the model specification, we confirm that policy decisions about the introduction of autonomy reforms are not related to previous achievement levels, corroborating the panel

²⁷ To illustrate the details on accountability, see the alternative estimates of its impact on student achievement in the U.S. (Figlio and Loeb (2011)).

identification. In addition, there are no significant interactions of the development level with other education policy measures, suggesting that the specific institutional effect and its heterogeneity are particular to autonomy reforms. Also, the significant interaction of autonomy with the level of economic development prevails when interactions of autonomy with measures of democracy, governance effectiveness, cultural values, and effective school environments are additionally taken into account, and the latter interactions are not significantly related to student outcomes once the interaction with economic development is held constant. Finally, there is no indication that autonomy differentially affects students with well-off and disadvantaged backgrounds. This suggests that autonomy reforms do not affect inequality between students with different social backgrounds in either developed or developing countries.

There is an indication that local decision-making works better when there is also external accountability that limits any opportunistic behavior of schools. Further, having generally well-functioning schools, indicated by initial performance levels, appears complementary with autonomy. Nonetheless, these specific issues require further research and confirmation.

From an analytical perspective, the innovation in this work is the development of panel data that permit cross-country analysis. Within this framework, we can exploit the pattern of policy changes within countries to obtain cleaner estimates of the institutional differences.

Does school autonomy make sense everywhere? Our results indicate that the impact of school autonomy on student achievement is highly heterogeneous, varying by the level of development of a country. This overall result may have broader implications for the generalizability of findings across countries and education systems. It suggests that lessons from educational policies in developed countries may not translate directly into advice for developing countries, and vice versa.

References

- Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell, and André Sapir. 2010. "The governance and performance of universities: Evidence from Europe and the US." *Economic Policy* 25, no. 61: 7-59.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8, no. 2: 155-194.
- Barankay, Iwan, and Ben Lockwood. 2007. "Decentralization and the productive efficiency of government: Evidence from Swiss cantons." *Journal of Public Economics* 91, no. 5-6: 1197-1218.
- Behrman, Jere R., and Elizabeth M. King. 2001. "Household schooling behaviors and decentralization." *Economics of Education Review* 20, no. 4: 321-341.
- Bifulco, Robert, and Helen F. Ladd. 2006. "The impacts of charter schools on student achievement: Evidence from North Carolina." *Education Finance and Policy* 1, no. 1: 50-90.
- Bishop, John H. 2006. "Drinking from the fountain of knowledge: Student incentive to study and learn - Externalities, information problems, and peer pressure." In *Handbook of the Economics of Education, Vol. 2*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 909-944.
- Bishop, John H., and Ludger Woessmann. 2004. "Institutional effects in a simple model of educational production." *Education Economics* 12, no. 1: 17-38.
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2007. "The impact of charter school attendance on student performance." *Journal of Public Economics* 91, no. 5-6: 849-876.
- Bruns, Barbara, Deon Filmer, and Harry A. Patrinos. 2011. *Making schools work: New evidence on accountability reforms*. Washington, D.C.: The World Bank.
- Clark, Damon. 2009. "The performance and competitive effects of school autonomy." *Journal of Political Economy* 117, no. 4: 745-783.
- CREDO. 2009. *Multiple choice: Charter school performance in 16 states*. Stanford, CA: Center for Research on Education Outcomes, Stanford University.
- Figlio, David, and Susanna Loeb. 2011. "School accountability." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 383-421.
- Fuchs, Thomas, and Ludger Woessmann. 2007. "What accounts for international differences in student performance? A re-examination using PISA data." *Empirical Economics* 32, no. 2-3: 433-462.

- Galiani, Sebastian, Paul Gertler, and Ernesto Schargrotsky. 2008. "School decentralization: Helping the good get better, but leaving the poor behind." *Journal of Public Economics* 92, no. 10-11: 2106-2120.
- Governor's Committee on Educational Excellence. 2007. *Students first: Renewing hope for California's future*. Sacramento, CA: Governor's Committee on Educational Excellence.
- Gustafsson, Jan-Eric. 2006. "Understanding causal influences on educational achievement through analysis of within-country differences over time." Invited paper presented at the 2nd IEA Research Conference, Washington, November 8-11 http://www.iea.nl/fileadmin/user_upload/IRC2006/Brookings_Institution_Program/Gustafsson.pdf.
- Hanushek, Eric A. 1994. *Making schools work: Improving performance and controlling costs*. Washington, D.C.: The Brookings Institution.
- Hanushek, Eric A. 2002. "Publicly provided education." In *Handbook of Public Economics, Vol. 4*, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: North Holland: 2045-2141.
- Hanushek, Eric A. 2003. "The failure of input-based schooling policies." *Economic Journal* 113, no. 485: F64-F98.
- Hanushek, Eric A., John F. Kain, Steve G. Rivkin, and Gregory F. Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of Public Economics* 91, no. 5-6: 823-848.
- Hanushek, Eric A., and Ludger Woessmann. 2008. "The role of cognitive skills in economic development." *Journal of Economic Literature* 46, no. 3: 607-668.
- Hanushek, Eric A., and Ludger Woessmann. 2011. "The economics of international differences in educational achievement." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 89-200.
- Hoxby, Caroline M. 1999. "The productivity of schools and other local public goods producers." *Journal of Public Economics* 74, no. 1: 1-30.
- Jimenez, Emmanuel, and Yasuyuki Sawada. 1999. "Do community-managed schools work? An evaluation of El Salvador's EDUCO program." *World Bank Economic Review* 13, no. 3: 415-441.
- Loeb, Susanna, and Katharine Strunk. 2007. "Accountability and local control: Response to incentives with and without authority over resource generation and allocation." *Education Finance and Policy* 2, no. 1: 10-39.
- Mourshed, Mona, Chinezi Chijioke, and Michael Barber. 2010. *How the world's most improved school systems keep getting better*. McKinsey and Company.
- Nechyba, Thomas J. 2003. "Centralization, fiscal federalism and private school attendance." *International Economic Review* 44, no. 1: 179-204.

- Organisation for Economic Co-operation and Development. 2010. *Education at a glance 2010: OECD indicators*. Paris: Organisation for Economic Co-operation and Development.
- Ouchi, William G. 2003. *Making schools work: A revolutionary plan to get your children the education they need*. New York, NY: Simon & Schuster.
- Summers, Anita A., and Amy W. Johnson. 1996. "The effects of school-based management plans." In *Improving America's schools: The role of incentives*, edited by Eric A. Hanushek and Dale W. Jorgenson. Washington, DC: National Academy Press: 75-96.
- Vegas, Emiliana. 1999. "School choice, student performance, and teacher and school characteristics: The Chilean case." World Bank Policy Research Working Paper No. 2833. Washington: World Bank.
- Woessmann, Ludger. 2003. "Schooling resources, educational institutions, and student performance: The international evidence." *Oxford Bulletin of Economics and Statistics* 65, no. 2: 117-170.
- Woessmann, Ludger. 2005. "The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA." *Education Economics* 13, no. 2: 143-169.
- Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, and Martin R. West. 2009. *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.
- World Bank. 2004. *World Development Report 2004: Making services work for poor people*. Washington, D.C.: The World Bank.

Table A1: Descriptive statistics and complete model of basic specification

	Descriptive statistics		Basic model	
	Mean	Std. dev.	Coeff.	Std. err.
Academic-content autonomy	0.663	0.259	-34.018***	(12.211)
Academic-content autonomy x Initial GDP p.c.	5.760	8.512	2.944***	(0.590)
Student and family characteristics				
Female	0.502		-13.028***	(0.917)
Age (years)	15.750	0.108	13.449***	(1.335)
<i>Immigration background</i>				
Native student	0.914			
First generation student	0.043		-20.976***	(4.690)
Non-native student	0.043		-12.607**	(5.124)
Other language than test language or national dialect spoken at home	0.092		-9.181**	(3.692)
<i>Parents' education</i>				
None	0.020			
Primary	0.079		10.697***	(2.115)
Lower secondary	0.107		11.724***	(2.610)
Upper secondary I	0.096		20.863***	(3.381)
Upper secondary II	0.277		25.784***	(2.866)
University	0.421		32.766***	(3.019)
<i>Parents' occupation</i>				
Blue collar low skilled	0.116			
Blue collar high skilled	0.152		6.013***	(1.184)
White collar low skilled	0.229		14.502***	(1.155)
White collar high skilled	0.502		35.714***	(1.953)
<i>Books at home</i>				
0-10 books	0.139			
11-100 books	0.464		29.430***	(2.339)
101-500 books	0.311		63.003***	(2.650)
More than 500 books	0.086		74.589***	(3.329)
School characteristics				
Number of students	782	327	0.016***	(0.003)
Privately operated	0.191		6.438	(4.481)
Share of government funding	0.841	0.154	-18.628***	(5.153)
Share of fully certified teachers at school	0.783	0.257	15.669***	(3.786)
Shortage of math teachers	0.178		6.984***	(1.449)
<i>School's community location</i>				
Village or rural area (<3,000)	0.110			
Town (3,000-15,000)	0.212		4.816**	(2.220)
Large town (15,000-100,000)	0.326		8.097***	(2.563)
City (100,000-1,000,000)	0.222		11.182***	(3.016)
Large city (>1,000,000)	0.129		12.191***	(3.633)
GDP per capita (1,000 \$)	23,199	18,785	0.416*	(0.245)
Country fixed effects			Yes	
Year fixed effects			Yes	
Student observations	1,042,995		1,042,995	
Country observations	42		42	
Country-by-wave observations	155		155	
R ²			0.385	

Notes: Descriptive statistics: Mean: international mean (weighted by sampling probabilities). Std. dev.: international standard deviation (only for continuous variables). Basic model: Full results of the specification reported in the top panel of Table 4. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability. Regression includes imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table A2: Questionnaire item on autonomy across PISA waves

Wave	Question	Answer options
2000	In your school, who has the main responsibility for ... (Please tick as many boxes as appropriate in each row)	Not a school responsibility Appointed or elected board Principal Department head Teachers
2003	In your school, who has the main responsibility for ... (Please tick as many boxes as appropriate in each row)	Not a main responsibility of the school School's governing board Principal Department head Teacher(s)
2006	Regarding your school, who has a considerable responsibility for the following tasks? (Please tick as many boxes as appropriate in each row)	Principals or teachers School governing board Regional or local education authority National education authority
2009	Regarding your school, who has a considerable responsibility for the following tasks? (Please tick as many boxes as appropriate in each row)	Principals Teachers School governing board Regional or local education authority National education authority

Notes: For each decision-making task, we constructed a variable indicating full autonomy at the school level if the principal, the school's board, department heads, or teachers carry sole responsibility. Consequently, if the task is not a school responsibility (2000 and 2003 data) or the responsibility is also carried at regional/local or national education authorities (2006 and 2009 data), we do not classify a school as autonomous. Figure 2 does not indicate consistent changes across waves in the measure of autonomy over countries or tasks, indicating that changes in response options are unlikely to substantially affect our estimates. Furthermore, in our models, time fixed effects capture consistent changes between waves.

Table A3: Disaggregation of basic model: Results for separate autonomy categories

	Estimation result		Details on autonomy effect at different levels of GDP per capita		
	Main effect (at initial GDP p.c. of \$8000)	Interaction with initial GDP per capita	GDP p.c. at which autonomy effect switches sign	Effect in country with minimum GDP p.c.	Effect in country with maximum GDP p.c.
	(1)	(2)	(3)	(4)	(5)
School autonomy over courses	-21.973 ^{***} (8.022)	2.380 ^{***} (0.481)	17,232	-39.102 ^{***} (10.308)	69.567 ^{***} (15.723)
R^2		0.385		0.385	0.385
School autonomy over content	-27.604 ^{**} (10.283)	2.303 ^{***} (0.513)	19,986	-44.174 ^{***} (12.826)	60.952 ^{***} (15.951)
R^2		0.385		0.385	0.385
School autonomy over textbooks	-23.635 ^{***} (8.207)	2.762 ^{***} (0.779)	16,557	-43.514 ^{***} (11.409)	82.604 ^{***} (28.213)
R^2		0.385		0.385	0.385
School autonomy over hiring	-31.763 [*] (16.369)	3.266 ^{***} (1.100)	17,725	-55.262 ^{**} (22.920)	93.826 ^{***} (31.892)
R^2		0.384		0.384	0.384
School autonomy over salaries	-2.115 (12.467)	2.476 ^{**} (1.156)	8,854	-19.932 (16.211)	93.103 ^{**} (43.917)
R^2		0.384		0.384	0.384
School autonomy over budget allocations	-5.935 (9.369)	1.914 ^{**} (0.730)	11,101	-19.707 (12.970)	67.665 ^{***} (24.346)
R^2		0.384		0.384	0.384

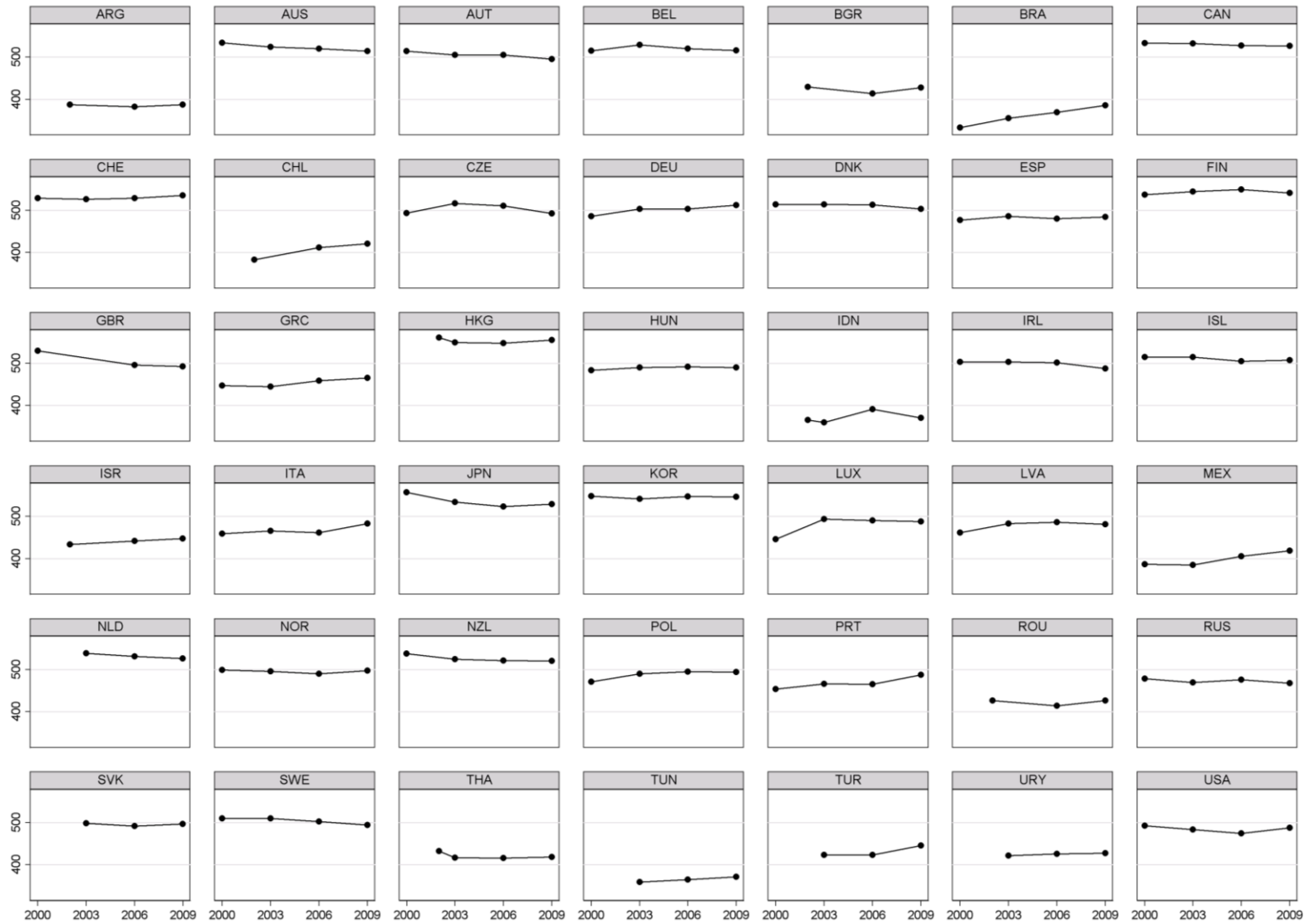
Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table A4: Alternative estimation of the impact of autonomy: Country-level estimation of two-step model

	First step does not include country fixed effects (1)	First step includes country fixed effects (2)
Academic-content autonomy	-30.247** (12.757)	-26.378** (10.691)
Academic-content autonomy x Initial GDP per capita	3.025*** (0.817)	2.892*** (0.701)
R^2	0.869	0.186
Personnel autonomy	-8.219 (1.494)	-14.322 (15.116)
Personnel autonomy x Initial GDP per capita	3.172** (17.284)	3.348*** (1.257)
R^2	0.856	0.099
Budget autonomy	-8.700 (12.141)	-7.480 (10.773)
Budget autonomy x Initial GDP per capita	1.679 (1.319)	0.945 (1.010)
R^2	0.853	0.051

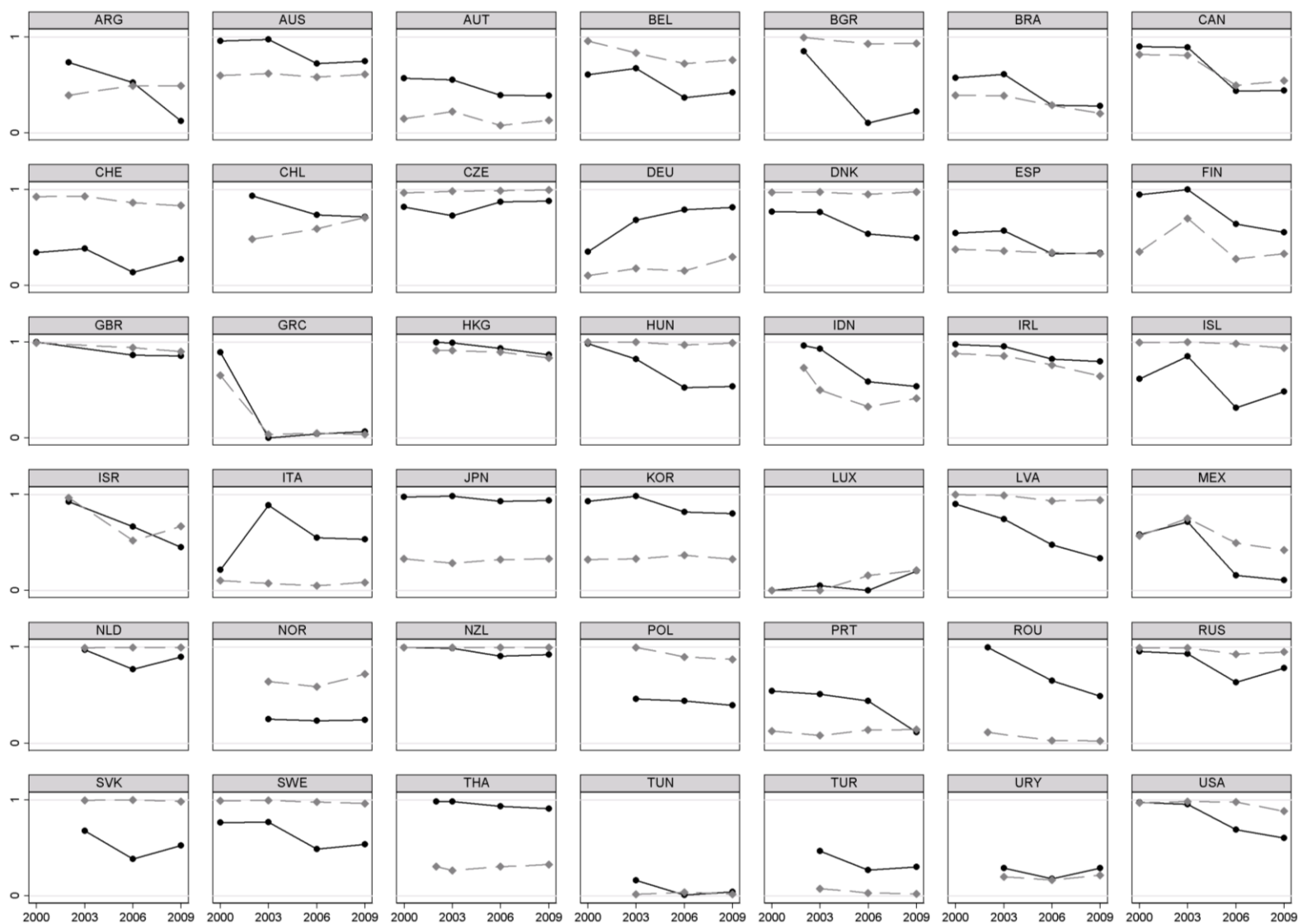
Notes: Each panel-by-column presents results of a separate regression. Reported coefficients stem from a country-level least squares regression with country and year fixed effects, controlling for GDP per capita. Sample: country-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 155 country-by-wave observations covering a total of 42 countries. Dependent variable: Country-level aggregation of the residuals of a first-step estimation at the student level that regresses the PISA math test score on student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home, school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school, and imputation dummies (and, in column 2, country's GDP per capita, country fixed effects, and year fixed effects). Initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Robust standard errors in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Figure 1: Performance on the PISA math tests, 2000-2009



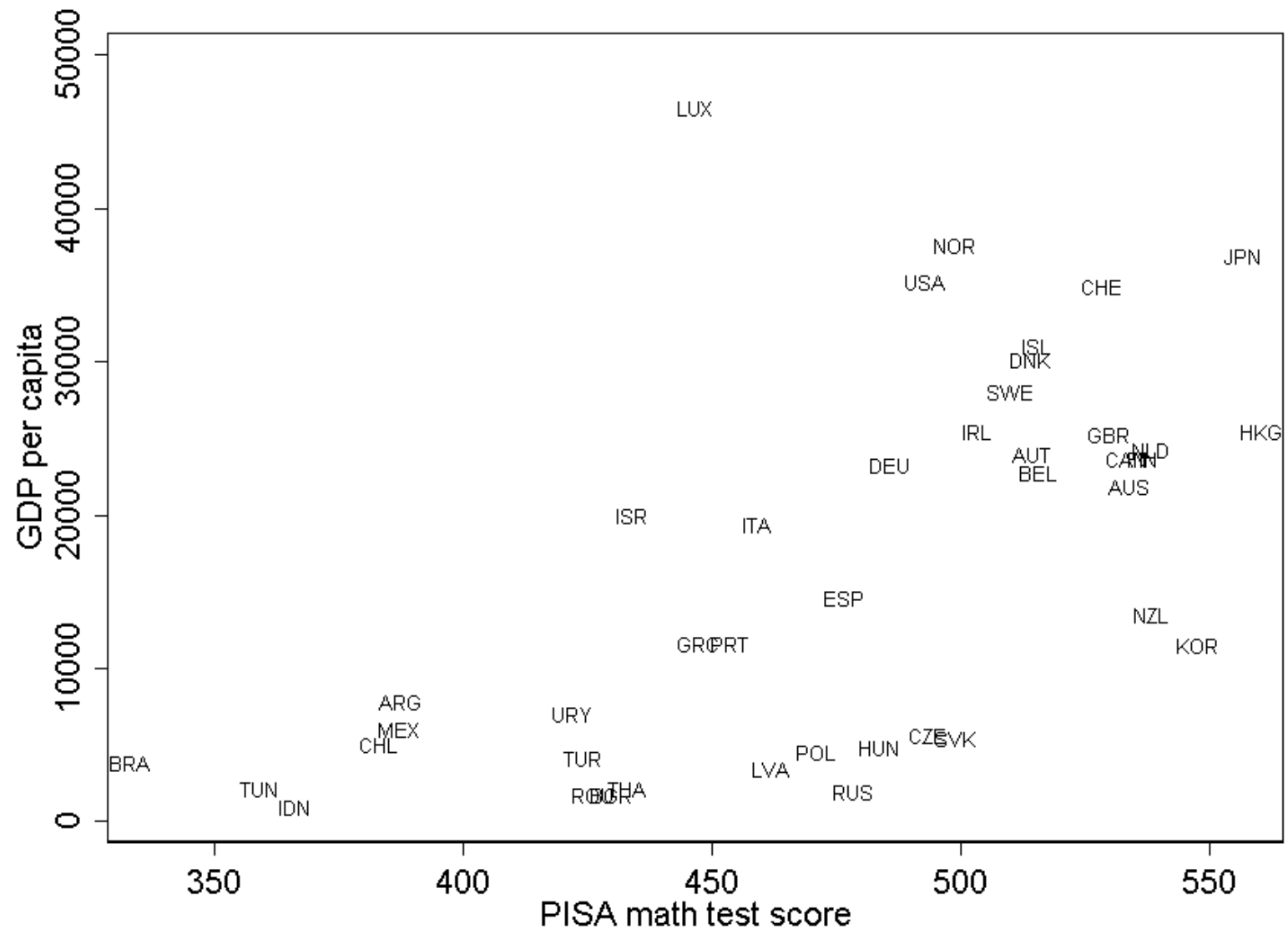
Notes: Country mean performance in the PISA math test. Own depiction based on PISA tests conducted in 2000/2002, 2003, 2006, and 2009.

Figure 2: School autonomy over courses and over hiring, 2000-2009



Notes: Straight black lines: autonomy in deciding which courses are offered. Dashed gray lines: autonomy in selecting teachers for hire. Own depiction based on school background questionnaires in the PISA tests conducted in 2000/2002, 2003, 2006, and 2009.

Figure 3: Development level and PISA performance, 2000



Notes: Test scores for Argentina, Bulgaria, Chile, Hong Kong, Indonesia, Israel, Romania, and Thailand refer to 2002. Test scores for Slovak Republic, Tunisia, Turkey, and Uruguay refer to 2003.

Table 1: Descriptive statistics by country

	GDP per capita	PISA math test score		Academic-content autonomy		Personnel autonomy		Budget autonomy	
	2000 (1)	2000 (2)	2009 (3)	2000 (4)	2009 (5)	2000 (6)	2009 (7)	2000 (8)	2009 (9)
Low-income countries^a									
Indonesia ^b	803	366.1	371.1	.915	.695	.689	.403	.973	.809
Lower-middle-income countries^a									
Brazil	3,701	332.8	386.0	.824	.516	.245	.156	.748	.348
Bulgaria ^b	1,600	429.6	427.9	.720	.410	.572	.796	.693	.923
Romania ^b	1,650	426.1	426.4	.737	.607	.113	.018	.996	.632
Russia	1,775	478.3	467.9	.958	.574	.704	.658	.701	.538
Thailand ^b	1,968	432.7	418.6	.961	.900	.284	.316	.896	.916
Tunisia ^c	2,033	358.9	371.5	.100	.028	.150	.016	.978	.810
Turkey ^c	4,010	423.8	445.7	.598	.218	.065	.017	.683	.772
Upper-middle-income countries^a									
Argentina ^b	7,693	387.4	387.6	.823	.407	.212	.266	.470	.738
Chile ^b	4,877	382.9	420.7	.900	.395	.394	.635	.651	.789
Czech Republic	5,521	493.3	492.6	.878	.864	.834	.883	.990	.746
Hungary	4,689	483.3	490.0	.983	.681	.705	.744	.922	.943
Latvia	3,302	461.7	481.5	.884	.413	.625	.523	.889	.826
Mexico	5,934	386.8	418.5	.661	.317	.414	.305	.772	.783
Poland ^d	4,454	470.7	494.2	.821	.750	.607	.484	.903	.264
Slovak Republic ^c	5,326	498.6	496.7	.754	.521	.798	.686	.955	.698
Uruguay ^c	6,914	421.8	427.2	.392	.216	.198	.192	.504	.577

(continued on next page)

Table 1 (continued)

	GDP per capita	PISA math test score		Academic-content autonomy		Personnel autonomy		Budget autonomy	
	2000 (1)	2000 (2)	2009 (3)	2000 (4)	2009 (5)	2000 (6)	2009 (7)	2000 (8)	2009 (9)
High-income countries^a									
Australia	21,768	533.7	514.6	.933	.708	.389	.373	.996	.934
Austria	23,865	514.2	495.3	.700	.569	.076	.069	.925	.845
Belgium	22,665	515.2	515.7	.726	.561	.512	.381	.992	.672
Canada	23,559	533.0	526.3	.759	.323	.577	.315	.986	.763
Denmark	29,992	513.7	503.2	.888	.684	.551	.596	.978	.981
Finland	23,514	536.4	540.4	.954	.616	.181	.211	.987	.925
Germany	23,114	485.5	512.1	.552	.644	.060	.176	.955	.975
Greece	11,500	447.3	465.4	.902	.047	.689	.027	.946	.858
Hong Kong ^b	25,374	560.5	554.7	.991	.870	.586	.519	.978	.911
Iceland	30,951	515.0	507.4	.797	.675	.517	.506	.871	.774
Israel ^b	19,836	433.6	447.4	.910	.506	.740	.379	.949	.658
Ireland	25,380	503.0	487.3	.781	.687	.461	.327	.878	.898
Italy	19,269	458.8	483.3	.716	.702	.057	.060	.570	.832
Japan	36,789	556.8	529.2	.988	.919	.328	.324	.912	.902
Korea	11,346	547.6	545.9	.973	.887	.234	.207	.947	.883
Luxembourg	46,457	446.1	488.2	.000	.139	.025	.171	1	.809
Netherlands ^c	24,179	538.1	525.9	.978	.922	.594	.896	.994	1
New Zealand	13,336	537.9	519.9	.957	.901	.586	.547	1	.992
Norway ^d	37,472	498.7	497.5	.571	.505	.325	.403	.982	.884
Portugal	11,443	453.4	487.3	.581	.382	.068	.107	.949	.932
Spain	14,421	476.4	483.7	.800	.538	.234	.185	.981	.959
Sweden	27,879	509.7	493.9	.879	.727	.804	.768	.993	.929
Switzerland	34,787	528.3	535.0	.381	.298	.526	.476	.868	.848
United Kingdom	25,089	529.7	492.5	.978	.871	.854	.719	.999	.945
United States	35,080	492.6	487.4	.912	.552	.867	.716	.986	.858
Country average	16,317	477.3	477.7	.780	.571	.445	.394	.902	.811

Notes: PISA data: Country means, based on non-imputed data for each variable, weighted by sampling probabilities. – = not available. ^a Country classification according to World Bank classification in 2002. ^b PISA data refer to 2002 instead of 2000. ^c PISA data refer to 2003 instead of 2000. ^d Autonomy data refer to 2002 instead of 2000.

Table 2: Country-level correlation matrix of autonomy measures

(A) 2009 levels	Courses	Content	Textbooks	Hiring	Salaries	Budget	Academic-content	Personnel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School autonomy over courses	1							
School autonomy over content	0.739 ^{***}	1						
School autonomy over textbooks	0.511 ^{***}	0.598 ^{***}	1					
School autonomy over hiring	0.385 ^{**}	0.384 ^{**}	0.366 ^{**}	1				
School autonomy over salaries	0.417 ^{***}	0.398 ^{***}	0.209	0.576 ^{***}	1			
School autonomy over budget allocations	0.274 [*]	0.060	0.228	0.089	0.186	1		
Academic-content autonomy	0.865 ^{***}	0.905 ^{***}	0.817 ^{***}	0.438 ^{***}	0.395 ^{***}	0.215	1	
Personnel autonomy	0.445 ^{***}	0.436 ^{***}	0.340 ^{**}	0.933 ^{***}	0.832 ^{***}	0.143	0.472 ^{***}	1

(B) 2000-2009 differences	Courses	Content	Textbooks	Hiring	Salaries	Budget	Academic-content	Personnel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
School autonomy over courses	1							
School autonomy over content	0.461 ^{***}	1						
School autonomy over textbooks	0.560 ^{***}	0.547 ^{***}	1					
School autonomy over hiring	0.559 ^{***}	0.342 ^{**}	0.688 ^{***}	1				
School autonomy over salaries	0.316 [*]	0.295 [*]	0.730 ^{***}	0.749 ^{***}	1			
School autonomy over budget allocations	0.066	-0.150	0.199	0.403 ^{**}	0.427 ^{**}	1		
Academic-content autonomy	0.846 ^{***}	0.813 ^{***}	0.811 ^{***}	0.626 ^{***}	0.503 ^{***}	0.030	1	
Personnel autonomy	0.454 ^{***}	0.338 ^{**}	0.760 ^{***}	0.921 ^{***}	0.948 ^{***}	0.445 ^{***}	0.597 ^{***}	1

Notes: Correlation coefficient of country-level autonomy measures across 42 countries. Data for Argentina, Bulgaria, Chile, Hong Kong, Indonesia, Israel, Romania, and Thailand refer to 2002 instead of 2000. Data for Slovak Republic, Tunisia, Turkey, and Uruguay refer to 2003 instead of 2000. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 3: Conventional cross-sectional estimation of the effect of school autonomy on student achievement

<i>Autonomy measured at level: Country fixed effects:</i>	School		Country		Country
		No	No		Yes
	(1)	(2)	(3)	(4)	(5)
Academic-content autonomy	20.713*** (6.181)	13.539* (7.455)	47.201*** (11.257)	37.114** (14.076)	-20.556 (12.627)
Academic-content autonomy x Initial GDP p.c.		0.771* (0.455)		0.908 (0.616)	
R^2	0.312	0.315	0.319	0.321	0.384
Personnel autonomy	9.640 (7.015)	10.479 (7.586)	24.701* (13.492)	24.913* (13.313)	-0.180 (11.708)
Personnel autonomy x Initial GDP per capita		-0.103 (0.535)		-0.024 (1.055)	
R^2	0.310	0.310	0.312	0.312	0.384
Budget autonomy	7.549* (4.248)	5.411 (4.694)	32.987 (25.976)	31.239 (25.228)	-7.163 (10.162)
Budget autonomy x Initial GDP per capita		0.493 (0.336)		1.127* (0.631)	
R^2	0.310	0.310	0.311	0.313	0.384

Notes: Each column-by-panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability. In columns 2 and 4, initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 4: Panel fixed-effects results on the effect of school autonomy on student achievement by development level

	Estimation result	Details on autonomy effect at different levels of GDP per capita		
	(1)	GDP p.c. at which autonomy effect switches sign (2)	Effect in country with minimum GDP p.c. (3)	Effect in country with maximum GDP p.c. (4)
Academic-content autonomy	-34.018*** (12.211)	19,555	-55.205*** (14.471)	79.209*** (20.912)
Academic-content autonomy x Initial GDP p.c.	2.944*** (0.590)			
R^2	0.385			
Personnel autonomy	-17.968 (14.071)	13,413	-41.854** (19.201)	109.687*** (37.857)
Personnel autonomy x Initial GDP per capita	3.319*** (1.106)			
R^2	0.384			
Budget autonomy	-6.347 (9.363)	11,449	-19.576 (13.282)	64.351** (26.939)
Budget autonomy x Initial GDP per capita	1.838** (0.796)			
R^2	0.384			

Notes: Each panel presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the main estimation, initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Complete model of the first specification displayed in Table A1. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 5: Robustness: Impact of including several autonomy measures together in the same estimation

	(1)	(2)	(3)	(4)
Academic-content autonomy	-42.013*** (14.248)	-41.012*** (14.120)	-33.732*** (12.054)	
Academic-content autonomy x Initial GDP p.c.	2.658*** (0.674)	2.736*** (0.676)	2.888*** (0.616)	
Personnel autonomy	17.830 (16.250)	13.897 (14.449)		-14.998 (13.534)
Personnel autonomy x Initial GDP per capita	0.212 (1.293)	0.333 (1.256)		2.868** (1.184)
Budget autonomy	-5.370 (8.858)		-1.918 (8.106)	-2.525 (9.205)
Budget autonomy x Initial GDP per capita	0.240 (0.898)		0.151 (0.832)	1.049 (0.923)
R^2	0.385	0.385	0.385	0.385

Notes: Each column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 6: Robustness: Different forms of measuring initial GDP per capita and different school controls

<i>Measure of initial GDP per capita:</i>	log GDP p.c.	Dummy for GDP p.c. above \$8,000	Dummy for GDP p.c. above median (\$14,000)	GDP per capita	
				No school controls	School controls measured as country means
<i>School controls:</i>	School controls measured at school level				
	(1)	(2)	(3)	(4)	(5)
Academic-content autonomy	-74.059 ^{***} (21.937)	-58.521 ^{***} (14.222)	-37.826 ^{**} (17.193)	-29.920 ^{***} (10.443)	-27.136 ^{**} (10.123)
Academic-content autonomy x Initial GDP p.c.	24.071 ^{***} (7.466)	60.362 ^{***} (13.013)	60.865 ^{***} (12.093)	2.646 ^{***} (0.539)	1.981 ^{***} (0.625)
R^2	0.385	0.385	0.385	0.373	0.356
Personnel autonomy	-55.008 [*] (29.136)	-41.921 [*] (24.595)	-21.154 (13.863)	-15.813 (15.393)	-15.993 (14.529)
Personnel autonomy x Initial GDP per capita	23.004 ^{**} (10.172)	62.247 ^{**} (27.308)	64.163 ^{***} (20.026)	2.750 ^{***} (0.968)	1.990 (1.300)
R^2	0.384	0.384	0.384	0.372	0.356

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In columns 4 and 5, initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 7: Robustness: Including expenditure per student and different sub-samples of waves and countries

	<i>Sample:</i>	Sample with expenditure data		Waves 2003, 2006, and 2009	Waves 2000 and 2009	Balanced panel	OECD countries
		(1)	(2)	(3)	(4)	(5)	(6)
Academic-content autonomy		-31.849 (22.327)	-24.753 (17.526)	-32.263*** (10.661)	-54.262** (22.019)	-36.980** (14.97)	-28.218** (13.324)
Academic-content autonomy x Initial GDP p.c.		2.976** (1.106)	2.645*** (0.913)	1.948*** (0.495)	4.050*** (1.132)	2.958*** (0.702)	2.529*** (0.760)
Expenditure per student (in 1000\$)			-11.375** (4.826)				
R^2		0.362	0.363	0.389	0.382	0.362	0.308
Personnel autonomy		-52.044*** (14.381)	-39.557** (15.577)	-28.282 (20.462)	-7.312 (18.926)	-45.458*** (15.033)	-21.601 (13.879)
Personnel autonomy x Initial GDP per capita		2.973** (1.002)	1.977* (0.977)	3.006** (1.204)	3.441* (1.914)	4.442*** (1.267)	3.060** (1.498)
Expenditure per student (in 1000\$)			-11.867* (5.932)				
R^2		0.361	0.362	0.389	0.379	0.361	0.308
Students		392,862	392,862	931,831	435,502	846,221	835,478
Countries		29	29	42	36	29	31
Countries-by-waves		69	69	120	72	116	116

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in the sample indicated on top of each column. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 8: Further results: Other subjects and joint authority

<i>Subject:</i> <i>Measurement of autonomy:</i>	Reading	Science	Math
	Full autonomy	Full autonomy	Joint authority
	(1)	(2)	(3)
Academic-content autonomy	-12.938 (8.928)	-28.529** (11.484)	-26.070* (13.152)
Academic-content autonomy x Initial GDP per capita	2.094*** (0.557)	1.115** (0.505)	1.185* (0.627)
R^2	0.351	0.337	0.384
Personnel autonomy	-6.929 (14.018)	-12.430 (10.810)	0.709 (13.838)
Personnel autonomy x Initial GDP per capita	3.098*** (1.066)	0.550 (0.853)	1.335 (0.921)
R^2	0.351	0.336	0.384
Students	1,125,794	1,042,791	1,042,995
Countries	42	42	42
Countries-by-waves	154	155	155

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA test score in respective subject. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 9: Extended model: Including central exit exams

	(1)	(2)
Academic-content autonomy	-48.511** (19.363)	-48.645** (19.921)
Academic-content autonomy x Central exit exams (CEE)	32.750** (14.374)	32.931* (16.382)
Academic-content autonomy x Initial GDP per capita	3.141*** (0.563)	3.168*** (0.938)
Academic-content autonomy x CEE x Initial GDP per capita		-0.042 (1.161)
R^2	0.380	0.380
Personnel autonomy	-28.555* (14.574)	-19.300 (17.994)
Personnel autonomy x Central exit exams (CEE)	18.310 (21.815)	5.755 (27.312)
Personnel autonomy x Initial GDP per capita	3.446*** (1.057)	0.897 (2.149)
Personnel autonomy x CEE x Initial GDP per capita		3.493 (2.545)
R^2	0.379	0.379

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. Initial GDP per capita is centered at \$8000 (measured in \$1000), so that the main effect shows the effect of autonomy on test scores in a country with a GDP per capita in 2000 of \$8000. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,031,652 students, 41 countries, 152 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.

Table 10: Alternative measure of development level: Initial level of student achievement

<i>Measure of initial achievement:</i>	<i>Average PISA score</i>		<i>Dummy for average PISA score above</i>	
	(1)	(2)	400 points (3)	500 points (4)
Academic-content autonomy	-63.257*** (14.544)	-60.480*** (13.773)	-85.567*** (19.203)	-32.530** (14.818)
Academic-content autonomy x Initial achievement	0.744*** (0.076)	0.601*** (0.089)	74.590*** (16.739)	73.258*** (12.443)
Academic-content autonomy x Initial GDP per capita		1.193** (0.535)		
R^2	0.386	0.386	0.385	0.385
Personnel autonomy	-32.691* (17.660)	-29.342 (18.356)	-51.538 (31.462)	-14.953 (12.657)
Personnel autonomy x Initial achievement	0.654*** (0.216)	0.372 (0.292)	63.266* (32.166)	82.534*** (24.584)
Personnel autonomy x Initial GDP per capita		1.991 (1.406)		
R^2	0.384	0.384	0.384	0.384

Notes: Each panel-by-column presents results of a separate regression. Dependent variable: PISA math test score. Least squares regression weighted by students' sampling probability, including country (and year) fixed effects. School autonomy measured as country average. In the first two columns, the initial average PISA score is centered at 400 (one standard deviation below the OECD mean), so that the main effect shows the effect of autonomy on test scores in a country that in 2000 performed at a level one standard deviation below the OECD mean. Sample: student-level observations in PISA waves 2000, 2003, 2006, and 2009. Sample size in each specification: 1,042,995 students, 42 countries, 155 country-by-wave observations. Control variables include: student gender, age, parental occupation, parental education, books at home, immigration status, language spoken at home; school location, school size, share of fully certified teachers at school, shortage of math teachers, private vs. public school management, share of government funding at school; country's GDP per capita, country fixed effects, year fixed effects; and imputation dummies. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent.