Please cite this paper as:

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

OECD Education Working Papers
No. 13

School Accountability, Autonomy, Choice, and the Level of Student Achievement

INTERNATIONAL EVIDENCE FROM PISA 2003

Ludger Wöbmann*, Elke Lüdemann, Gabriela Schütz, Martin R. West

OECDPUBLISHING

*Ifo Institute for Economic Research,University of Munich, Germany

Unclassified **EDU/WKP(2007)8**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **21-Dec-2007**

English text only

DIRECTORATE FOR EDUCATION

EDU/WKP(2007)8
Unclassified

School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003

By Ludger Wößmann, Elke Lüdemann, Gabriela Schütz, Martin R. West

(Education Working Paper No. 13)

Accountability, autonomy, and choice feature prominently in recent school reforms in many countries. This report provide new evidence on the relationship between student performance and these system level features. It finds that different facets of accountability, autonomy, and choice are strongly and positively associated with the level of student performance, and that the combined achievement differences amount to more than one and a half PISA grade-level equivalents. This report was prepared under the Education Policy Committee's activity on Parental Choice, School Autonomy and System Accountability. Financial support for this work was provided by the OECD and by additional voluntary contributions from Belgium (Flemish Community), New Zealand, Norway, Sweden, and Switzerland.

Gregory Wurzburg, Education and Training Policy Division, OECD Directorate for Education
Tel: +33 1 45 24 92 95; E-mail: gregory.wurzburg@oecd.org

JT03238353

English text only

School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003

Ludger Wößmann, Elke Lüdemann, Gabriela Schütz, Martin R. West[1]

9 September 2007

[1] The authors are from Ifo Institute for Economic Research at the University of Munich; Department Human Capital and Innovation; Poschingerstr. 5; 81679 Munich, Germany; Phone: (+49) 89 / 9224-1692; E-mail: woessmann@ifo.de; Internet: www.cesifo.de/woessmann. Martin R. West is affiliated with the Education Department, Brown University, Providence, RI.

OECD DIRECTORATE FOR EDUCATION

OECD EDUCATION WORKING PAPERS SERIES

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Directorate for Education. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

Comment on the series is welcome, and should be sent to either edu.contact@oecd.org or the Directorate for Education, 2, rue André Pascal, 75775 Paris CEDEX 16, France.

ABSTRACT

Accountability, autonomy, and choice play a leading role in recent school reforms in many countries. This report provides new evidence on whether students perform better in school systems that have such institutional measures in place. We implement an internationally comparative approach within a rigorous micro-econometric framework that accounts for the influences of a large set of student, family, school, and country characteristics. The student-level data used in the analysis comes from the PISA 2003 international student achievement test that encompasses up to 265,000 students from 37 countries.

Our results reveal that different facets of accountability, autonomy, and choice are strongly associated with the level of student achievement across countries. With respect to accountability, students perform better where policies are in place that aim at students (external exit exams), teachers (monitoring of lessons), and schools (assessment-based comparisons). The combined achievement differences amount to more than one and a half PISA grade-level equivalents.

Students in schools with hiring autonomy perform better on average, while they perform worse in schools with autonomy in formulating their budget. School autonomy over the budget, salaries, and course contents appears to be more beneficial when external exit exams hold schools accountable for their decisions.

Students perform better in countries with more choice and competition as measured by the share of privately managed schools, the share of total school funding from government sources, and the equality of government funding between public and private schools. Cross-country differences in private school operation account for up to two PISA grade-level equivalents. The performance advantage of privately operated schools within countries is stronger where schools face external accountability measures and are autonomous. In urban areas, indicators of choice among public schools are also associated with superior outcomes.

Several aspects of accountability, autonomy, and choice are also associated with superior non-cognitive outcomes such as student morale and commitment, non-disruptive behaviour, disciplinary climate, and tardiness. We find no evidence that these policies have led schools to focus on raising student achievement at the expense of non-cognitive skills.

RÉSUMÉ

La responsabilité, l'autonomie et le choix sont au cœur des réformes récentes des systèmes scolaires de nombreux pays. Ce rapport apporte de nouveaux éléments sur la question de savoir si les élèves réussissent mieux dans des systèmes scolaires qui ont adopté ces mesures. Une analyse comparative à l'échelle internationale est menée dans un cadre micro-économétrique rigoureux qui tient compte des incidences d'un large éventail de paramètres liés à l'élève, au milieu familial, à l'établissement et au pays. Les données relatives aux élèves utilisées pour l'analyse sont tirées du test international de niveau des élèves de l'enquête PISA 2003, qui porte sur 265 000 élèves de 37 pays.

Les résultats montrent que différentes facettes de la responsabilité, de l'autonomie et du choix sont étroitement associées au degré de réussite des élèves dans l'ensemble des pays. S'agissant de la responsabilité, les élèves réussissent mieux lorsqu'il existe des mesures concernant les élèves (examens de sortie externes), les enseignants (suivi des leçons) et les établissements scolaires (comparaisons fondées sur des évaluations). Les écarts de niveau combinés vont jusqu'à l'équivalent PISA de plus d'une année et demi d'études.

Les élèves inscrits dans des établissements ayant la possibilité de recruter librement leurs enseignants réussissent mieux en moyenne, alors qu'ils réussissent moins bien dans les établissements libres d'établir leur budget. L'autonomie des établissements scolaires en matière de budget, de salaires et de contenu des programmes semble plus bénéfique lorsqu'ont été mis en place des examens de sortie externes qui rendent les établissements comptables de leurs décisions.

Les élèves ont de meilleurs résultats dans les pays où le niveau de choix et de concurrence est plus élevé, mesuré d'après le pourcentage d'établissements privés, la part totale des financements publics et l'égalité des financements publics entre les écoles publiques et privées. Les différences de fonctionnement des établissements privés selon les pays sont à l'origine d'écarts de niveau pouvant atteindre l'équivalent PISA de deux années d'études. Les bonnes performances des écoles privées au sein des pays sont plus nettes lorsque les établissements sont soumis à des mesures externes de responsabilité et sont autonomes. Dans les zones urbaines, les indicateurs de choix au niveau des établissements publics sont également associés à de meilleurs résultats.

Plusieurs aspects de la responsabilité, de l'autonomie et du choix sont par ailleurs liés à de meilleurs résultats non cognitifs dans des domaines tels que moral et participation des élèves, comportement non perturbateur, discipline et retards. Nous n'avons pas trouvé d'éléments montrant que ces mesures avaient poussé les établissements scolaires à privilégier l'amélioration des résultats des élèves au détriment de leurs compétences non cognitives.

TABLE OF CONTENTS

List of Tables and Figures

1. INTRODUCTION

Governments around the globe have for decades worked to improve their national school systems in order to provide the best education possible to their students. More often than not this has meant spending more on public education in the hope that additional resources would translate into better student outcomes. However, it is increasingly clear that more spending on its own does not guarantee more learning; in most cases, it does not seem to have any significant effect on student achievement within existing school systems (e.g., Hanushek 2002; Wößmann 2002, 2007a). As a consequence, policymakers in many countries have begun to focus more on reforming the institutional structure of their school systems.

1.1 Evaluating Recent Incentive-Based Reform Movements

In this most recent wave of school reforms, three institutional reform strategies have played a leading role: accountability, autonomy, and choice. Policymakers in many countries have implemented or are considering reforms along one of these three dimensions; others suggest that any one of them will not be effective if the others are not already in place. A notable example of the introduction of far-reaching accountability systems is the 2001 No Child Left Behind legislation in the United States, which requires each state to establish standards for student achievement, test students annually to see whether those standards have been met, and impose sanctions on low-performing schools. Several countries with traditionally centralized school systems are considering the decentralization of decision-making authority in certain domains to schools, a policy that has been implemented on a pilot basis in two German states. Still other countries have expanded parental choice among schools, as when Sweden in the 1990s introduced both free parental choice of schools and a voucher system that placed privately operated schools on equal footing with public schools in terms of access to public funding. The introduction of the "quasi-market" in education in the United Kingdom since 1988 has included aspects of all three strategies: the publication of external exam results, devolution of control over resource allocation to the school level, and increased parental choice within districts with public funding following students to the schools of their choice.

Proponents of greater accountability, autonomy, and choice contend that these reforms will improve student outcomes by heightening incentives for various actors to perform at high levels. Accountability systems combine clear standards, external monitoring of results, and corresponding rewards and sanctions based on performance indicators. By providing better information on student outcomes, proponents argue, such systems directly and indirectly reward students, teachers, and school leaders for their efforts. Decentralizing decision-making to the schools, advocates suggest, substitutes the creativity and knowledge of local decision-makers for the inertia and rigidity of centralized bureaucracies. Supporters of school choice contend that giving parents free choice among schools and enabling private providers of education to receive government funding unleashes competitive forces that will drive school improvement.

Institutional reform strategies are not without controversy, however. Part of the opposition comes from individuals and organizations working within the current school system who may fear the loss of accustomed benefits. But others contend that accountability, autonomy, and choice will not improve outcomes and could even have adverse effects, especially if they are poorly implemented. For example, critics of accountability note that isolating the impact of teachers or schools on student outcomes is complex and that many valuable schooling outcomes are difficult to measure reliably. They warn that high-

stakes testing systems can narrow curricula, stifle creativity, and undermine student engagement. Critics of autonomy argue that it often overburdens school leaders and creates opportunities for the misallocation of resources at the local level. Most of all, critics assert that choice and competition in schooling will hurt the most disadvantaged, thereby weakening social cohesion. The best schools in a choice-based system will take only the top students, they argue, leaving behind those who are most in need of assistance.

So what is the evidence? Do accountability, autonomy, and choice raise or lower the level of student achievement? This report provides new evidence on whether or not students perform better in school systems that have various forms of accountability, autonomy, and choice policies in place relative to systems that do not. We also place a particular focus on how these three factors interact to determine student outcomes. While this report focuses on the level of student achievement, Schütz, West, and Wößmann (2007) focus on how school accountability, autonomy, and choice affect the equity of student achievement.

1.2 The Internationally Comparative Approach

The approach that we take in this report is to compare the achievement of students in countries exposed to accountability, autonomy, and choice to students in countries not exposed to them within a rigorous micro-econometric framework. This internationally comparative approach has great potential to shed light on the effects of institutional variation on student outcomes. Its chief advantage stems from the ability to exploit the substantial variation in national education policies across countries. This international variation can be used to estimate whether various forms of accountability, autonomy, and choice are associated with higher or lower performance. By contrast, there is typically much less variation in institutional structures within countries. In most cases, the extent to which students and schools are subject to accountability systems, school leaders have autonomy over basic functions, and parents have choice among schools is similar for all schools in a country, leaving no way to examine their consequences.

Moreover, even where within-country variation exists, for example in the case of public and private schools operating within the same system, comparisons of student achievement are often subject to severe selection problems. Students who choose to attend a private school may differ along both observable and unobservable dimensions from students taught in neighborhood public schools. While it is possible to control for differences in student, family background, and school characteristics when estimating the effects of institutional structures, thereby comparing students who are observationally equivalent, such estimates may still suffer from selection on unobserved characteristics. By aggregating the institutional variables to the country level, we circumvent the selection problem – in effect measuring the effect of, for example, the share of students in a country attending private schools on student achievement in the country as a whole. Cross-country evidence therefore cannot be biased by standard issues of selection at the individual level.

In addition, the presence of private schools may influence the behaviour of nearby public schools with which they compete for students. As a result, simple comparisons of private and public schools may miss an important part of the effects of greater private involvement in education. Again, aggregated measures of the institutional feature can solve the problem: By comparing the average performance of systems with larger and smaller shares of private schools, the cross-country approach captures any systemic effect of competition from private schools.

1.3 The PISA 2003 Micro Database

We implement this internationally comparative approach using the student-level database from the 2003 Programme for International Student Assessment (PISA) study (cf. OECD 2004 for details). The PISA study provides comparable information on students' mathematics, science, and reading literacy for

41 countries, including all 30 member countries of the Organisation for Economic Co-operation and Development (OECD).[2] Because the PISA 2003 test focused on mathematics, with less detailed testing in science and reading, the current report focuses mostly on mathematics.[3] In contrast to previous international studies following a curriculum-based testing approach, the questions in the PISA literacy domains aim to test how well students are prepared to meet the real-life challenges of modern societies.

PISA tested representative samples of 15-year-old students in each participating country. Most countries implemented a two-stage sampling design, drawing a stratified random sample of schools in a first stage and then randomly testing 35 students in each school in a second stage. The student sample sizes in the different OECD countries range from 3,350 students in 129 schools in Iceland to 29,983 students in 1,124 schools in Mexico, yielding an international dataset of more than 200,000 OECD-country students. Using item response theory, PISA mapped performance in the three subjects on a scale with an international mean of 500 and a standard deviation of 100 test-score points across the OECD countries. As a benchmark to which to compare the magnitude of effects reported below, note that the simple test-score difference between the two grades with the largest share of 15-year-olds (9^{th} grade and 10^{th} grade) is 22.1 test-score points in mathematics (25.7 in science, 23.6 in reading). This "grade-level equivalent" gives a rough idea of how much students learn on average during one school year.

In addition to students' educational achievement in the three subjects, the PISA 2003 database also contains a host of background information on the participating students and schools. Separate background questionnaires completed by students and by school principals provide detailed information on students' demographic characteristics, their family backgrounds, and their home environments, as well as school characteristics such as location and resource endowments. In addition, the school background questionnaires contain information on aspects of accountability, autonomy, and choice that serve as key institutional measures in our analyses. Details on these indicators are discussed in the relevant chapters below. General details on the database used in this report, including the construction of a workable student-level micro database and descriptive statistics of the international data and selected national measures, are provided in Appendix A.

1.4 Why It Matters

Understanding the sources of international variation in student achievement levels is an important project, all the more so because recent research shows that international differences in student achievement are a key driver of differences in long-run economic growth rates (cf. Hanushek and Kimko 2000; Wößmann 2002; Hanushek and Wößmann 2007a, 2007b). Economic theory suggests that strong education systems will increase the long-run rate of economic growth because education is an investment in human capital that increases labor productivity and because it is a leading input for innovation and technical progress which in turn influence growth rates (e.g., Barro and Sala-i-Martin 2004).

Hanushek and Wößmann (2007a) combine data from 36 international student achievement tests administered on 12 occasions between 1964 and 2003 to develop an aggregate measure of the average educational achievement of a country. Entering this measure of educational achievement in standard cross-country growth regressions that control for the initial level of per-capita Gross Domestic Product (GDP) and years of education, educational achievement turns out to be a powerful predictor of economic growth in 1960-2000, both in the full sample of 50 countries with available data on achievement and growth and in

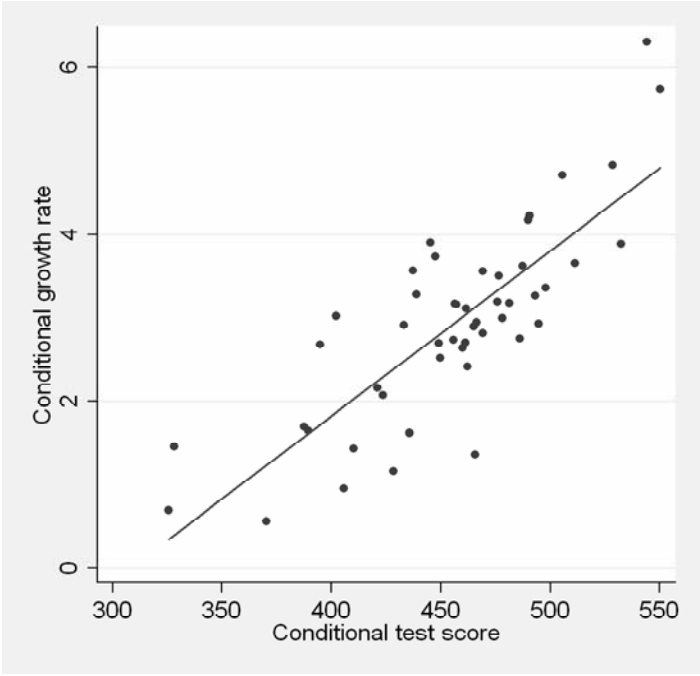
[2] We excluded France from the analyses in this report because the PISA 2003 database does not include school-level information for any of its schools. Outside the OECD, Liechtenstein, Macao, and Serbia/Montenegro had to be discarded from the analysis because of lack of internationally comparable information on key country-level variables.

[3] Another "minor domain" of PISA 2003 was problem-solving skills.

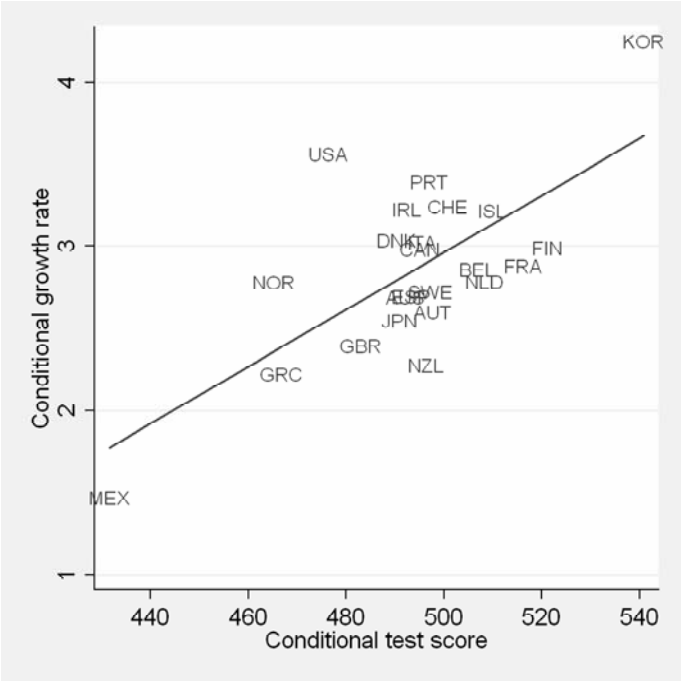
the sub-sample of OECD countries. After accounting for the impact of test-score achievement, the quantity of schooling as measured by years of education no longer has a significant effect on growth.

Figure 1: Student achievement and long-run economic growth

(a) 50-country sample



(b) OECD-country sample



Figures are based on a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960-2000 on the initial level of real GDP per capita in 1960, average years of schooling in 1960, and average test scores on several international student achievement tests. Source: Based on Hanushek and Wößmann (2007).

These results are depicted in Figure 1, which shows that countries with better educational achievement had substantially higher growth rates. This effect is robust to the inclusion of additional control variables, different sub-samples of countries, different specifications of the test-score measure, using only early test scores to predict later growth in 1980-2000, and a variety of other checks. An extension of this analysis indicates that skill levels for the population as a whole and for the top end of the achievement distribution in each country have independent positive effects on growth. The size of the relationship suggests that in the very long run, the average annual growth rate would increase by about 1.2 percentage points for a one standard deviation improvement in test scores. An educational reform that improved test scores by half a standard deviation over a 20-year span would increase real GDP by 36 percent over a 75 year horizon; the initial effects are more limited, of course, because it takes a long time before students who have attended the reformed school system have replaced the total labor force.

In highlighting the importance of policies affecting student achievement, this new cross-country evidence only adds to the compelling evidence that better test scores are associated with better economic outcomes at the individual level. Several recent studies suggest that a one standard deviation increase in mathematics performance at the end of high school translates into about 12 percent higher annual earnings (e.g., Mulligan 1999; Murnane, Willett, Duhaldeborde, and Tyler 2000; Lazear 2003). Higher test scores are also associated with a lower probability of being unemployed (e.g., Bishop 1992; OECD 2000; McIntosh and Vignoles 2001).

1.5 Structure of the Report

Given the crucial importance of educational achievement for economic outcomes, this report estimates how different facets of the institutional structures of accountability, autonomy, and choice are related to student achievement across countries. After presenting a basic model relying on summary indicators of the three institutional features in the next chapter, Chapters 3-5 examine in greater detail various facets of each of the three institutional dimensions. These chapters also discuss the theoretical background for each dimension and existing international evidence on its effects on student achievement. In addition, they explore possible interactions between the three dimensions, so as to see whether, for example, autonomy is more beneficial for student outcomes in the presence of a strong accountability system. While these chapters focus on the set of cognitive skills measured in the PISA test, in particular on student achievement in mathematics, Chapter 6 offers a complementary analysis of how accountability, autonomy, and choice relate to available indicators of non-cognitive skills. The final chapter concludes with the main lessons to be drawn from the detailed results presented in the preceding chapters.

Before going into the details of each of the three institutional features separately, Chapter 2 presents a basic model that provides a snapshot of some of the main effects of accountability, autonomy, and choice. It starts by providing some general background on how market-oriented reforms can create incentives that may affect student outcomes, as well as on the general structure of how the empirical models of this report are set up. (Details of the econometric modeling are relegated to Appendix B.) It then presents the main results of the basic model based on the PISA 2003 micro database, as a general background for the detailed analyses that follow, and demonstrates the robustness of the basic findings to a host of changes in the basic specification in terms of included controls and specific samples.

Chapter 3 presents detailed analyses of the effects of various facets of accountability on student achievement. The PISA 2003 database provides rich data on different aspects of accountability policies, including measures aimed at students through external exit exams and through schools' use of assessments to decide about students' promotion; measures aimed at teachers by monitoring their classes either through the principal or through external inspectors; and measures aimed at schools by using assessments to compare a school's performance to district or national performance. These analyses rely on measures of

accountability at both the country level and the school level. The chapter also analyzes the interdependence between the effects of external exit exams and standardized testing within schools.

Chapter 4 presents evidence on the effects of school autonomy in different areas of decision-making. These include autonomy in formulating school budgets, in staffing decisions and hiring teachers, in establishing teachers' salaries, and in determining course content. Detailed results probe the dependence of the effects of autonomy in these areas on the presence of accountability systems that hold schools responsible for their decisions.

Chapter 5 analyses the effects of choice and competition on student achievement. It looks both at the availability of privately operated schools, including aspects of the increased competition generated by the availability of government funding for privately operated schools, and at the extent to which parents can choose between different public schools. It also addresses how the effects of choice interact with the extent to which accountability systems provide information on relative school performance and with the extent to which schools have autonomy to respond to market forces.

Chapter 6 supplements the previous analyses, which are based on the cognitive skills tested in the PISA test, with measures of non-cognitive skills. We derive such measures from items in the PISA student and school background questionnaires reporting on students' morale and engagement, on disruptive behaviour, on disciplinary climate, and on tardiness. The models presented in this chapter estimate how issues of accountability, autonomy, and choice are related to these non-cognitive skills of students.

Chapter 7 concludes with an overview of central findings and a discussion of how policy reforms can be informed by the evidence in this report.

1.6 Summary of Main Results

The main empirical results of this report are as follows:

- Different facets of school accountability, autonomy, and choice are strongly associated with the level of student achievement across countries in PISA 2003.

- Students perform better in schools and countries where various forms of **accountability** policies are in place (Chapter 3). This is true for accountability measures aimed primarily at students, such as external exit exams and the use of assessments for decisions on student promotion and retention; for accountability measures aimed at teachers, such as internal and external monitoring of teacher lessons; and for accountability measures aimed at schools, such as assessments used to compare them to district or national performance. Together, these accountability effects sum to a combined effect of the equivalent of more than one and a half grade-level equivalents on the PISA test.

- On average, students in schools that have **autonomy** in hiring decisions outperform students in schools without staffing autonomy (Chapter 4). By contrast, performance in schools with autonomy in formulating their budget is worse on average. Yet these average effects mask important differences in the effects of autonomy between systems with and without accountability policies: School autonomy over the budget, over salaries, and over course contents is more beneficial when measures of school accountability, especially external exit exams, hold schools accountable for their decisions. An exception is hiring autonomy, the effect of which is smaller in external-exam systems.

- Students in countries with more **choice** and competition in the form of larger shares of privately managed schools, larger shares of government funding, and more equalized government funding between public and private schools perform better (Chapter 5). Cross-country differences in private school operation can account for up to two PISA grade-level equivalents. The positive effect of privately operated schools is stronger when they are held accountable by external inspections of teachers and assessment-based comparisons to national performance, as well as when schools in the system have autonomy to respond to the private competition. By contrast, proxies for choice among public schools, such as the share of students in a country who do not attend their school because it is the local school and who report that they attend their school because it is better than alternatives, are not associated with higher student achievement on average. However, within urban areas where there are schools to choose from, reduced local attendance and increased choice of better schools are associated with superior outcomes.

- Our basic model of differences in accountability, autonomy, and choice, together with the student and school control variables, can account for more than 80 percent of the between-country variation in average student achievement across OECD countries (Chapter 2). The institutional effects prove highly robust to a long list of alternative specifications and robustness checks.

- The higher cognitive achievement of students in schools that are exposed to accountability, autonomy, and choice does not come at the cost of lower **non-cognitive skills** (Chapter 6). On the contrary, several aspects of accountability, autonomy, and choice are associated with superior outcomes in terms of student morale and commitment, levels of disruptive behaviour, the overall disciplinary climate in schools, and tardiness.

2. A BASIC MODEL

This chapter presents a basic model that provides an overview of the main effects of accountability, autonomy, and choice that will be probed in detail in the following chapters. After briefly examining how these institutional structures might be expected to alter incentives within school systems and affect student outcomes, we provide a short description of the empirical model used for the econometric estimation. We then report the main results of the basic model and demonstrate their robustness.

2.1 Background: Incentives Created by Market-Oriented Reforms

All over the world, nations tend to finance and manage the great majority of their schools publicly. Unfortunately, the dominance of the public sector in education often limits incentives to improve student achievement while controlling costs. In the private business sector, market competition tends to encourage firms to operate efficiently so as to generate profits. Inefficiency leads to higher costs and higher prices, which allows competitors to lure away customers. By contrast, a lack of competition and choice in most state-run school systems often creates obstacles to leaving bad schools, thereby constraining the ability of parents to ensure high-quality education. Centralized bureaucracies often allow little flexibility at the school level, limiting schools' ability to respond to parental demands. And information on what students and schools actually achieve is often unavailable, hindering parents' ability to make informed choices.

The rationale of the recent wave of market-oriented reforms in the school system is to change this. They aim to enhance choice on the demand side, to endow suppliers with more autonomy, and to provide parents with more information about student outcomes. The main consequence of these changes in the institutional framework of the system is that they alter the incentives that actors face. The institutions of the school system are the set of rules and regulations that determine rewards and penalties for those involved in the schooling process. Economic theory suggests that people respond to these incentives: If the actors in the education process are rewarded (extrinsically or intrinsically) for producing better student achievement, and if they are penalized for not producing high achievement, they will change their behaviour in a way that improves achievement. While the relative lack of accountability, autonomy, and choice in the compulsory education sector as currently constituted tends to dull incentives to improve quality and restrain costs (cf. Hanushek with others 1994), market-oriented models may create incentives that ultimately lead to better student learning.

Attempts to provide parents with additional choice and to allow non-governmental providers to enter the education marketplace clearly represent market-oriented reforms. And enabling the producer side – the schools – to exercise at least some autonomy is obviously essential for them to compete. However, in decision-making areas where local units have little knowledge leads compared to central units and where local decision-makers have incentives to act opportunistically, furthering their own goals rather than the educational goals of the school system or of parents, school autonomy may also lead to adverse effects.

It may be less obvious why accountability is also a key ingredient of market-oriented reforms. However, one of the major contributions of economic theory in the second half of the 20th century was to show how markets do not work properly if information is absent (e.g., Stiglitz 2002). In the same way, the education market needs sufficient information on performance to ensure that educational choices are made so that incentives are indeed geared towards better student learning. One rationale for accountability

systems is to provide this information. In addition, they can help to inform the political market, enabling voters to make better choices with respect to education policy.

In sum, institutional reforms that ensure informed choice between autonomous schools may be expected to improve student achievement because they create incentives for everyone involved to provide the best learning environment for students (see Bishop and Wößmann 2004 for a general model of institutional effects in education).

2.2 Empirical Model: Cross-Country Student-Level Multiple Regressions

In order to estimate the effects of accountability, autonomy, and choice empirically, we rely primarily on institutional variation across countries. Of course, student achievement depends on a lot of other factors inside and outside of school systems, which must be taken into account if we want to isolate the effects of institutions. For example, if children whose parents are both working are more likely to attend private schools than children whose parents are unemployed, and if parental work status has a direct influence on the students' achievement, then the estimated effect of private schooling would capture the effect of parental work status as long as the effect of the latter was not controlled for. Similarly, the estimated effect of school autonomy would be biased if more autonomous schools were also better equipped with material resources and if the effects of these resource differences were not accounted for.

We therefore estimate so-called "education production functions" (cf., e.g., Hanushek 1994) that control for differences in various student, family, school, and country characteristics that may influence student achievement. To do this as rigorously and efficiently as possible, we perform the cross-country regressions at the student level, which allows for possible intervening effects to be accounted for at the level of each individual student. Thus, our empirical model has three important features: It uses cross-country variation, it is performed at the level of individual students, and it estimates the effects of many variables simultaneously.

Our international education production functions combine individual student-level data on educational achievement with extensive background information mostly taken from student and school background questionnaires in order to express student achievement on the PISA test as a function, *f*, of several determining factors:

$$\textit{Student achievement} = f(\textit{student characteristics, family background, school resources, country characteristics, accountability, autonomy, choice}) \quad (1a)$$

More formally, the achievement test score T_{isc} of student i in school s in country c is regressed on several sets of potential influences:

$$T_{isc} = B_{isc}\alpha + R_{sc}\beta + I_{sc}\gamma + \varepsilon_{isc} \quad (1b)$$

In this specification, B is a vector of student background data including student characteristics, family background, and country characteristics. It consists of 32 variables, including such indicators as the student's gender and age, attendance of institutions of pre-primary education, immigration status, family status, parental occupation and work status, and the per-capita GDP of the country. R is a vector of data on schools' resource endowments and location, comprising 10 variables such as class size, availability of materials, instruction time, teacher education, city size, and average expenditure per student in the country. (See Table C.1 in Appendix C for a complete list of the control variables included in all the models presented in this report.) I is the vector of institutional characteristics of interest in this report, combining several different measures of school accountability, autonomy, and choice.

The parameter vectors α, β and γ are estimated by least-squares micro-econometric regressions at the level of individual students i, with a sample size of more than 200,000 students. The estimation of such micro-econometric models encompasses additional technical details, such as the weighting of student observations with their sampling probabilities, proper statistical inferences in light of the hierarchical structure of the data which adds higher-level components to the error term ε of the model, and the treatment of missing values in the background questionnaires. To be able to use a complete dataset of all students with data on achievement and at least some background characteristics, we imputed missing values using advanced micro-econometric techniques as described in Appendix B.3. To account for this in the estimations, all our models include a complete set of indicators identifying observations with imputed values for each variable. All these technical details on the econometric modeling are discussed in Appendix B at the end of the report.[4]

Our aim in this chapter is to provide an overall summary of the main results for accountability, autonomy, and choice. Therefore, we restrict the modeling of the institutional features to a very simple specification. First, we use only country-level measures of accountability, autonomy, and choice. The main reason for this, as discussed in Section 1.2, is to evade problems of within-country selectivity and to capture potential systemic effects. The downside of using only country-level institutional measures is that the degrees of freedom at the country level are very limited. Specifically, with 29 countries included in the OECD sample and GDP per capita and educational expenditure per student included as country-level controls, there are only 26 degrees of freedom left at the country level for the analysis of institutional effects. Therefore, the second feature of the basic model is that we use only one or two summary indicators of each of the three institutional features.

Note that in all our models, the institutions of accountability, autonomy, and choice are jointly entered in the empirical models, so that possible effects of the other institutions are taken into account in the estimations. Even in the more detailed analyses in subsequent chapters, the measures of the other two institutions included in the basic models are included as control variables when probing the details of the effects in each specific institutional dimension.

2.3 Results

Results of the basic model are reported in Table 1. The model is estimated both for mathematics achievement and for science achievement, and both for the sample of OECD countries and for the extended sample of all countries participating in PISA 2003. Note that all models control for the 42 variables described above measuring student and family background and schooling resources; detailed results on the control variables included in the model are reported in Table C.1 in Appendix C.

[4] See also Wößmann (2003a, 2003b) and Fuchs and Wößmann (2007) for methodological details of the econometric techniques.

Table 1: The basic model

Subject:	Mathematics		Science	
Country sample:	OECD	Extended	OECD	Extended
	(1)	(2)	(3)	(4)
External exit exams	13.724*	11.155*	15.745**	13.824**
	(7.496)	*(6.192)*	*(6.992)*	*(5.205)*
Autonomy in formulating budget	-25.056**	-28.596**	-17.723	-17.655*
	(10.661)	*(10.728)*	*(11.515)*	*(10.377)*
Autonomy in staffing decisions	29.310*	34.974**	21.216	23.177*
	(14.685)	*(13.710)*	*(14.733)*	*(13.051)*
Private operation	61.563***	61.405***	38.985***	42.757***
	(10.419)	*(10.317)*	*(8.517)*	*(8.747)*
Government funding	75.437***	80.114***	58.538**	54.644***
	(20.901)	*(17.352)*	*(21.958)*	*(16.757)*
Observations (students)	219,794	265,878	118,809	143,528
Clustering units (countries)	29	37	29	37
R^2	0.386	0.461	0.348	0.389

Dependent variable: PISA 2003 international test score. Least-squares regressions weighted by students' sampling probability. All five institutional variables are measured at the country level. Controls include: 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. The extended country sample specifications include an OECD dummy. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

The measure of accountability included in the basic model is whether a country has external exit exams at the end of secondary school. Such "curriculum-based external exit examination systems" can be defined by six characteristics (cf. Bishop 1997): 1) They produce signals of student achievement that have real consequences for the student. 2) They define achievement relative to an external standard, not relative to other students in the classroom or the school. 3) They are organized by discipline and keyed to the content of specific course sequences. 4) They signal multiple levels of achievement in the subject, not only a pass-fail signal. 5) They cover almost all secondary school students. 6) They assess a major portion of what students studying a subject are expected to know.

As reported in column (1) of Table 1, students in countries that have external exit exams in mathematics perform 13.7 test-score points better on the PISA mathematics test than students in countries without external exit exams. Compared to the "grade-level equivalent" of 22.1 test-score points, this is more than half of what students on average learn during a whole school year. Likewise, students in countries with external exit exams perform 15.7 test-score points better in science.[5] In both subjects, the positive effect is also present and statistically significant in the extended sample of 37 countries that includes non-OECD members participating in the PISA 2003 study.

[5] The number of student observations in science is only 54 percent of that in mathematics because not all students received test questions in science in PISA 2003. Within each school, students were randomly assigned to test booklets with different questions, all of which contained mathematics questions, but only part of which contained science questions. In our science regressions, we use only those students who were given science questions. The PISA 2003 database reports plausible values for science achievement also for students who did not respond to any science item (cf. OECD 2005a, pp. 206-211). Regression results are similar when including these students in the science analyses.

While these results point towards the achievement-enhancing potential of accountability systems, there are a lot of different ways to implement accountability. External exit exams are mostly aimed at providing incentives for the students, although they may also create indirect accountability pressures for teachers and schools. Other accountability devices, such as external inspection of teachers' lessons and comparison of schools' performance to the national average, may be aimed at teachers and schools. Chapter 3 goes into much greater detail on the possible effects of these different forms of accountability policies.

We include two measures of autonomy in the basic model: the share of schools in a country having main responsibility over formulating the school budget, and the share of schools exerting a direct influence on decision-making about staffing. The effects of the two kinds of autonomy point into opposite directions: While autonomy in formulating the budget is negatively associated with student achievement, autonomy in staffing decisions is positively associated with student achievement. It seems that on average, schools that can formulate their own budget use this in ways that hurt student achievement. By contrast, schools that can decide about staffing issues use this autonomy to advance student achievement.

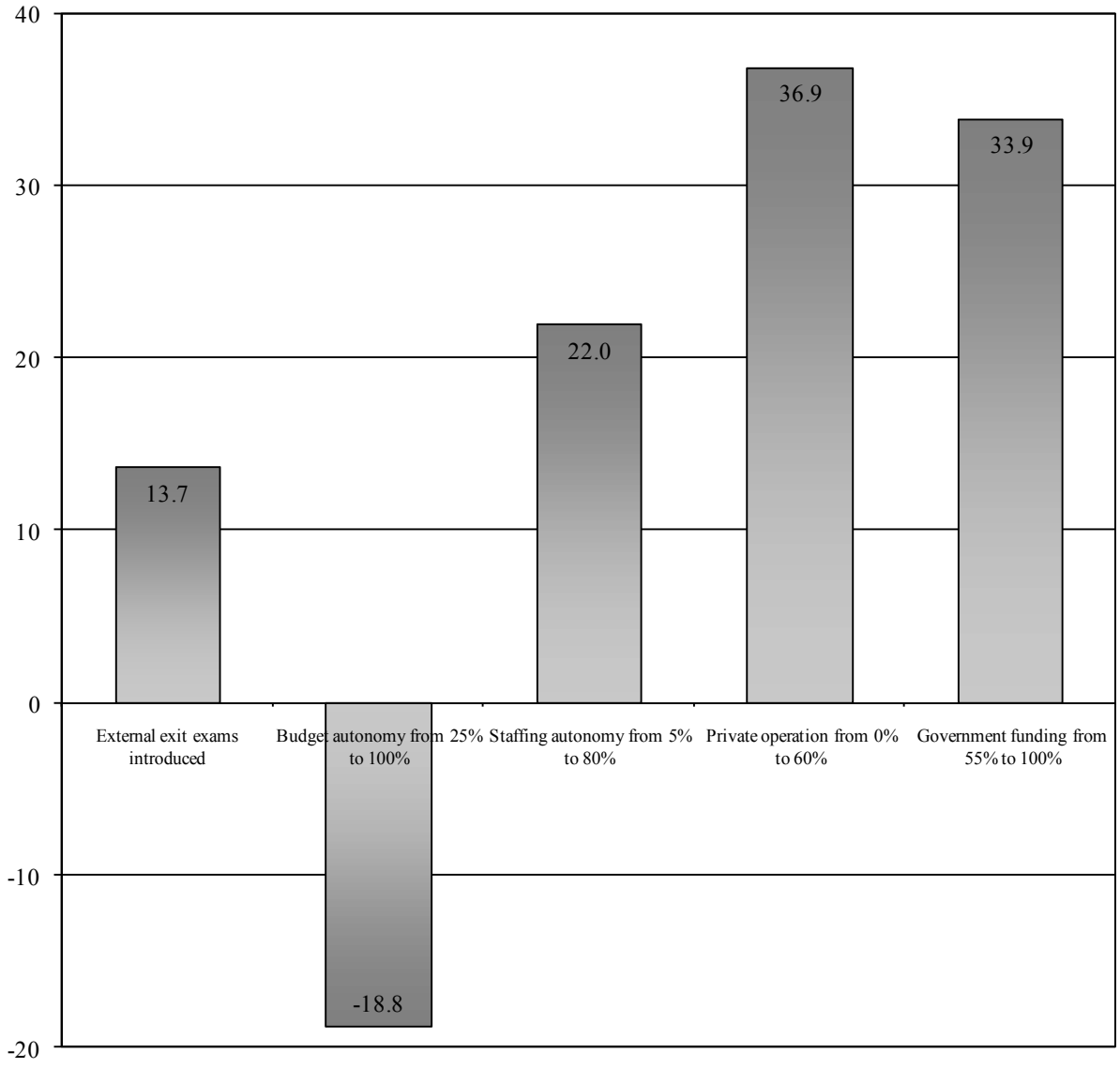
This basic pattern of results suggests that the effects of school autonomy may be complex and depend on the specific decision-making area, as well as on the complementary institutional framework. The diverse effects of school autonomy are probed in much greater detail in Chapter 4.

The main measure of choice included in the basic model is the share of privately operated schools in a country. As is evident from Table 1, private school operation is strongly and significantly positively associated with student achievement. The effect is huge: Going from a system without any private school operation to a system where half the schools are privately operated increases the achievement level by substantially more than the equivalent of one year's average learning in mathematics (three quarters of a grade-level equivalent in science). The extent to which this effect stems from better performance of privately operated schools themselves and from better performance of public schools that are exposed to the competition from private schools is analyzed in Chapter 5.

While in the operation of schools, private involvement is associated with better performance, the association is reversed in the financing of schools: A larger average share of government (as opposed to private) funding of schools is associated with better student achievement. Larger government funding, in particular when it is available to privately operated schools, may create choice for a larger share of the population and thus increase competition. The merits of this hypothesis are also probed in Chapter 5, as are the effects of proxies for the extent of choice among public schools.

Figure 2 depicts what the estimated effects of the basic model mean in terms of performance differences between countries with the highest and lowest values of each of the institutional measures. Disregarding countries with extreme values on the institutional measures, the figure takes the country at the first decile and the country at the ninth decile of the international distribution of each of the institutional measures and depicts the achievement difference between the two countries as predicted by the basic model. For example, the country at the first decile – with only 10 percent of countries (two countries) below it – in terms of the share of schools that have autonomy in staffing decisions, the Czech Republic, has only about 5 percent schools with staffing autonomy. The country at the ninth decile – with only 10 percent of countries above it – is Switzerland, with roughly 80 percent autonomous schools. Our basic model suggests that the effect of staffing autonomy can account for 22 test-score points of the difference in the average PISA test score between these two countries. Measured this way, the share of privately operated schools can account for the largest achievement difference among OECD countries: Going from no privately operated schools (the first decile) to 60 percent privately operated schools (the ninth decile) is associated with 37 additional PISA test-score points.

Figure 2: Estimated achievement difference between countries with different institutions



Estimated size of the effect of each institution when comparing the country at the first and ninth decile (percentages rounded) of the distribution of each institution across countries. Coefficient estimates based on column (1) of Table 1.

These results suggest that differences in accountability, autonomy, and choice can explain large differences in student achievement across countries. Together with the student and school control variables, our model can account for 39 percent of the total student-level variation in student achievement in mathematics across the OECD countries (46 percent in the extended sample, 34 in science in the OECD sample, and 37 in science in the extended sample). This is a substantial explanatory power, given the importance of unobservable student differences in ability that exist within each country. The basic model accounts for as much as 82 percent of the between-country variation in average achievement in both mathematics and science that exists between OECD countries (87 percent in mathematics and 85 percent in science in the extended sample).

Given the similarity of the results in mathematics and science, and given that mathematics was the focus of PISA 2003, the remainder of this report focuses on mathematics achievement.[6] The detailed analyses also focus on the sample of OECD countries, because the results are very similar between the OECD sample and the extended sample and because confidence in international comparability may be greater among the group of advanced and relatively homogeneous countries.

2.4 Robustness of the Basic Model

To test whether the results on the effects of accountability, autonomy, and choice found in the basic model are sensitive to the specific model specification, we perform several robustness checks in terms of the specific sample of students and countries, the set of included control variables, and the specification used to account for data imputations. All the results on the three institutional effects prove very robust to variations of the basic model. We report only the main robustness checks here; results of the specifications discussed are reported in Tables C.2 and C.3 in Appendix C.

Table 1 already showed that results are robust to using the sample of OECD countries or the extended sample of countries. Among the OECD countries, two countries – Mexico and Turkey – stand out by having an average socio-economic status that is a full standard deviation below the OECD average (as measured by the PISA index of Economic, Social and Cultural Status, ESCS; they also have by far the lowest GDP per capita). As specification (1) of Table C.2 reveals, the qualitative results are unaffected by excluding these two countries from the OECD sample.

To test whether results are sensitive to unusual grades, specification (2) of Table C.2 excludes students in grades 6 and 12 from the sample. Results hardly change, as might be expected given the fact that these two grades encompass less than one percent of the students each. Specification (3) restricts the sample further by looking only at the two adjacent grades within each country that have the largest share of 15-year-olds in the respective country (note that these grades differ across countries). Again, the results do not change materially.

The basic model includes controls for the grade level in which the student is taught. However, it might be argued that a student's grade is to some extent endogenous to his or her performance, particularly in systems where grade repetition is common. Therefore, specification (4) of Table C.2 does not include controls for a student's grade level. Our results are qualitatively unaffected by this change. Likewise, our results on the effects of the institutional variables hold when the indicators for grade repetition and for school entry age are dropped from the model.

The first two specifications of Table C.3 probe additional changes to the control model. Specification (1) introduces measures of an additional institutional feature of the school system, the number of years that 15-year-old students are tracked into different school types and the number of tracks. None of the two measures of ability-based tracking is significantly related to the achievement level of a country. Specification (2) adds an indicator variable for Europe to the model, to see whether the results are sensitive to excluding any variation between the 22 European and the remaining 7 countries; they are not. Additional regional indicators for the two Asian countries or the five Eastern European countries are not statistically significant.

---

[6] Qualitative results in reading are similar to results in mathematics and science, albeit usually at a somewhat lower level of statistical significance. The effect of private operation is equally strong and robust. A measure of external exit exams in reading literacy that would be comparable across countries is not available; using the mean of external exit exams in mathematics and science as a proxy, the effect reaches statistical significance at the 15 percent level. In the extended country sample, both autonomy effects and the effect of government funding are statistically significant at standard levels.

The final two columns of Table C.3 report results of two alternative specifications used to account for data imputations (cf. Appendix B.3 for details on data imputation and its implications for the model specification). Specification (3) omits the imputation indicator controls from the model, in effect assuming that observations are missing conditionally at random. Specification (4) uses a simpler and more standard method of imputation, where a simple constant is imputed for each missing value of each variable and imputation indicators for each variable are added to the model. Neither alternative for dealing with data imputations yields substantially different results (the coefficient on external exit exams is statistically significant at the 11% and 15% level, respectively).

In sum, the results of the basic model prove extremely robust to changes in the sample of countries, the sample of grades, country-level controls, and imputation methods.

3. ACCOUNTABILITY

This chapter presents more detailed results on the effects of different facets of accountability on student achievement, including accountability measures aimed at students, teachers, and schools. We first provide theoretical background on the effects of accountability based on a principal-agent model of the schooling process and survey the existing international evidence. Then we present new results on the effects of different accountability policies using the PISA 2003 database.

3.1 Theory: Providing Information to Overcome Principal-Agent Problems

Accountability refers to all devices that attach consequences to measured educational achievement. Accountability systems generally consist of three components: achievement standards, measurement of student achievement, and consequences for measured achievement. These consequences may be positive (rewards) or negative (sanctions), and they may be implicit (e.g. the respect of peers) or explicit (e.g. cash bonuses). Furthermore, their target may be any stakeholder in the education process, including students, teachers, and schools.

From a theoretical viewpoint, the provision of schooling can be understood as a network of principal-agent relationships in which a principal (e.g. the parent) commissions an agent (e.g. the head of a school) to perform a service (the education of the child) on her behalf. Principal-agent theory identifies decentralized information and divergent interests as the fundamental sources of difficulties in principal-agent relationships: "Delegation of a task to an agent who has different objectives than the principal who delegates this task is problematic when information about the agent is imperfect" (Laffont and Martimort 2002, p. 2). If the agent's interests diverge from those of the principal, and if the information on the agent's real performance is asymmetric (available only to the agent), then the agent may pursue his own interests instead of those of the principal; the principal will remain unaware of this behaviour and thus unable to sanction it. Unfortunately, such principal-agent problems are pervasive in school systems.

As a consequence, theoretical models of educational production predict that setting clear performance standards and providing performance information can tilt incentives in favor of superior student achievement (cf., e.g., Costrell 1994; Betts 1998). For example, if schools use performance assessments to make decisions about students' retention or promotion, students may have greater incentives to learn and perform well.

Another accountability device that aims primarily at students is external exams, where a decision-making authority external to the school has exclusive responsibility for or gives final approval of the content of examinations.[7] External exams help resolve the problem of incomplete monitoring of agents' behaviour by supplying information about the performance of individual students relative to the national (or regional) student population. This information is unavailable in the absence of external exams, when grades assigned by classroom teachers provide the only information on student performance. In the latter setting, a mark earned in one class may not reflect the same level of achievement as a mark earned in another class. By signaling the achievement of students relative to an external standard, the information

[7] For a detailed model of the effects of external exams see Bishop and Wößmann (2004), on which our discussion draws.

provided by external-exam systems makes students' performance comparable to the performance of students in other classes and schools. As students receive marks relative to the national average, their educational achievement is made observable and transparent, facilitating the monitoring of the performance of students, teachers, and schools. This profoundly alters the incentive structure in the school system compared to school-based or teacher-based examinations.

The influence of external exams on student achievement may run through three basic channels: increased external rewards for learning, decreased peer pressure against learning, and enhanced monitoring of teachers and schools. Most of all, external exams change the students' incentive structure relative to local exams. By creating comparability to an external standard, external exams improve the signaling of academic performance to advanced educational institutions and potential employers. These institutions will thus place greater weight on educational achievement when making admissions and hiring decisions. As a result, their decisions become less sensitive to other factors such as family connections, racial and religious stereotypes, the chemistry of a brief job interview, performance relative to the class mean, or aptitude tests which lean more to measuring innate ability than to measuring overall educational achievement. The increased rewards for learning heighten students' learning efforts.

A second channel through which external exams may impact on student achievement is through their impact on peer behaviour. Assigning grades relative to the class average gives students an incentive to lower average class achievement so that they will receive the same grades at less effort. The cooperative solution for students to maximize their joint welfare is for everyone not to study hard. Students therefore have an incentive to apply peer pressure on other students in the class not to be too studious and to distract teachers from teaching a high standard. With external exams, in contrast, the peer incentives to denigrate studiousness dissipate because inferior class work leads only to lower marks.

A third potential channel of positive impact of external exams on student achievement runs through the monitoring of teachers and schools. With external exams, for example, it becomes evident whether the bad performance of an individual student is an exception within a class or whether the whole class taught by a teacher is doing badly relative to the country mean. Therefore, parents (and students) have the information they need to initiate action because they can observe whether the teacher (and/or the student) is accountable for the bad performance. If, by contrast, students receive marks relative to the class mean only, the performance of the class relative to the country mean is unobservable and parents have no information on which to intervene. External exams thus reduce the leeway of teachers to act opportunistically and increase the incentives to use resources more effectively. The same argument can be made for the monitoring of schools as a whole. Through external exams, agents are made accountable to their principals: parents can assess the performance of their children, of the teachers, and of the schools; heads of schools can assess the performance of their teachers; and the government and administration can assess the performance of different schools. Similarly, politicians may become more accountable to the electorate for their decisions. Thus, external exams not only induce accountability for students, but ultimately also for teachers, schools, and possibly the political system.

The accountability introduced by external exams can help to create a set of incentives that encourages school personnel to behave in ways that do not necessarily further their own interests, but rather the interest of best student learning. For instance, without the right incentives, teachers may avoid using the most promising teaching techniques, preferring to use the techniques they find most convenient. If a country assesses the performance of students with external exams and uses this information to monitor teachers, teachers may put aside their other interests and focus mainly on raising student achievement.

In terms of teaching and learning, a pivotal difference in the incentive mechanism of external exams relative to teacher-set exams is that neither teachers nor students know beforehand which specific questions are going to be asked. Teachers therefore cannot "get away" with skipping whole content areas in the

classroom. They are instead forced to teach the whole subject areas as prescribed in the standards and cannot effectively scale down the standards. Furthermore, if well implemented, the possibility of teacher cheating – for example by discussing the specific questions of the exam beforehand or by telling students that certain content areas will not be covered in the exam – is eliminated. In sum, because of incomplete contracts and monitoring in the school system, external testing of achievement can lead to better-informed choices and make students and educational providers accountable for what they learn and teach.

Many countries also have other accountability devices in addition to external exams that aim directly at teachers. In particular, it may have profound consequences for teachers' behaviour whether they can expect that the principal of their school or other senior staff comes to observe their lessons regularly or whether this is not the case. Even more, observations of classes by inspectors or others external to the school can be an even more binding monitoring device for teachers. Because teachers can expect explicit or implicit consequences for the quality of their teaching, external inspection of teacher practices may create incentives for better teaching and thus ultimately lead to better student achievement.

More recently, there has been increasing discussion of accountability measures aimed at entire schools. Performance assessments can be used to compare each school to regional or national performance. Countries like England and France publish national league tables of schools based on their students' achievement on central exams. This creates incentives for better performance at the school level. Accountability systems currently in place in some regions of the United States set monetary rewards or sanctions for entire schools in response to their performance. While the external-exam systems discussed above usually work indirectly through implicit consequences that rely on the behaviour of different educational stakeholders, these school-based accountability systems create explicit monetary consequences for schools. Both implicit and explicit consequences for their actions can orientate the efforts of school leaders towards better student achievement.

In sum, because of the general lack of performance information in most school systems, accountability measures that make students, teachers, and schools more responsible for their actions can lead to improved student achievement.

3.2 Existing Evidence

The existing evidence on the effects of accountability using the internationally comparative approach has focused almost exclusively on one specific accountability device, namely external exit exams at the end of secondary school. Evidence from several previous international student achievement tests shows that students perform substantially and statistically significantly better in countries that have external exit-exam systems than in countries without external exit-exam systems. This has been found on the 1991 International Assessment of Educational Progress (IAEP) math, science, and geography tests (Bishop 1997), the 1991 International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy study (Bishop 1999), the 1995 Third International Mathematics and Science Study (TIMSS; cf. Bishop 1997; Wößmann 2001, 2003a), the 1999 TIMSS-Repeat study (Wößmann 2003b), and the PISA 2000 reading, math, and science tests (Bishop 2006; Fuchs and Wößmann 2007). Taken as a whole, the existing cross-country evidence suggests that the effect of external exit exams on student achievement may well be larger than a whole grade-level equivalent.

Similarly, cross-regional studies in countries where some regions have external exit exams while others do not find the same result. Positive effects of external exit exams have been shown for Canadian provinces (Bishop 1997, 1999), U.S. states (e.g., Bishop, Moriarty, and Mane 2000), and German states (Jürges, Schneider, and Büchel 2005; Wößmann 2007c). Wößmann (2007c) even shows that the estimated size of the effect of external exit exams does not differ significantly between the sample of German states and the sample of OECD countries.

Evidence from PISA 2000 also suggests that students perform better where teachers monitor student progress by regular standardized tests and exams (Fuchs and Wößmann 2007). Similar evidence has been found in primary school, as well, using data from the Progress in International Reading Literacy Study (PIRLS; cf. Fuchs and Wößmann 2005).

Evidence from the United States shows positive effects on students' learning achievement for explicit school-focused accountability systems (Carnoy and Loeb 2003; Hanushek and Raymond 2004; Jacob 2005). However, there is also evidence that school-focused accountability systems can lead to strategic responses on part of teachers and schools, for example by increasing placements of low-performing students in special-education programs which are outside the accountability system or by preemptively retaining students (Jacob 2005). High-stakes testing may also introduce incentives for cheating (Jacob and Levitt 2003). A lot seems to depend on the specific implementation of accountability systems that focus on schools.

3.3 New Results

The basic model presented in Table 1 confirmed that the positive effect of external exit exams on student achievement is also evident in PISA 2003. Students in countries with curriculum-based external exit exam systems outperform students elsewhere by more than what the average student learns in half a year, even after controlling for the effects of student characteristics, family background, school location and resources, countries' GDP per capita and expenditure per student, staffing and budgeting autonomy, and the share of private school operation and government funding.

As discussed above, while external exit exams directly affect the incentives that students face to improve their achievement, their effects on teachers and schools are more indirect. Fortunately, the PISA 2003 database also includes measures of accountability policies that focus directly on each of these three stakeholders. The school background questionnaires collected information on student assessments, teacher monitoring, and school accountability. A detailed description of the questions underlying the data on measures of school accountability, autonomy, and choice can be found in Appendix A.3. Table 2 reports the results on the relationship between each of these measures and student achievement, both within and between countries.

Table 2: Accountability

Level at which accountability is measured:	Country	School		
	(1)	(2)	(3)	(4)
External exit exams	24.506**	16.195**	n.i.	15.096**
	(10.059)	*(7.083)*[a]		*(7.305)*[a]
Assessments used to make decisions about students' retention/promotion	27.150**	10.047***	4.463***	10.068***
	(12.766)	*(1.646)*	*(1.744)*	*(1.639)*
Assessments used to group students	-29.596*	-6.069***	-3.877***	-5.913***
	(15.334)	*(1.283)*	*(1.222)*	*(1.276)*
Monitoring of teacher lessons by principal	14.025	5.334***	3.031**	5.150***
	(8.371)	*(1.287)*	*(1.335)*	*(1.282)*
Monitoring of teacher lessons by external inspectors		3.171**	2.349*	3.110**
		(1.444)	*(1.415)*	*(1.445)*
Assessments used to compare school to district or national performance		2.283*	5.300***	2.195*
		(1.241)	*(1.200)*	*(1.232)*
Standardized tests used at least monthly				-14.933***
				(3.357)
External exit exams x Standardized tests used at least monthly				16.825***
				(5.714)
Country fixed effects	no	no	yes	no
Students	219,794	219,794	219,794	219,794
Schools	8,245	8,245	8,245	8,245
Countries	29	29	29	29
R^2	0.391	0.389	0.414	0.390

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. Controls include: autonomy in formulating budget, autonomy in staffing decisions, private operation, government funding, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering in parentheses (column 1: clustering at country level; columns 2 to 4: clustering at school level). n.i. = not identified. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent. [a] Clustering of standard errors at country level.

Column (1) adds three additional country-level measures of accountability to the basic model. The first is the percentage of schools using assessments to make decisions about students' retention or promotion, another accountability device aimed squarely at students. The results reveal that students perform significantly better in countries with larger shares of schools using this accountability measure. That is, after controlling for all other factors, students in countries where hardly any school uses assessments for promotion and retention, such as Denmark and Iceland, perform more than one grade-level equivalent worse than students in countries where nearly all schools use assessments for promotion and retention, which include Belgium, Canada, Finland, Greece, the Netherlands, and Spain (cf. Table A.2 in Appendix A for country-level descriptive statistics on the institutional measures).

The PISA study also asked principals to report on whether they use assessments to group students for instructional purposes. Aggregating this variable to the country-level provides a rough measure of the extent of tracking that goes on within schools. The results indicate that students in countries with a larger share of schools using assessments to group students perform substantially worse than students in countries where fewer schools do so. This negative effect of tracking within schools is consistent with previous international evidence indicating that the tracking of students between schools adversely affects student outcomes (cf. Hanushek and Wößmann 2006). This suggests that how schools use student assessments is

important: Using assessments for promotion decisions seems to incentivize higher achievement, while using them to group students creates an environment in which students are less likely to succeed academically.

The PISA 2003 background questionnaires also provide information on the monitoring of teachers. Principals report whether they or other senior staff have, during the last year, observed lessons to monitor the practice of mathematics teachers at their school. Hardly any principals in Greece, Ireland, and Portugal responded affirmatively, while almost all schools perform this kind of monitoring in the Czech and Slovak Republics, Poland, New Zealand, and the United States. The results show that students in countries with more monitoring of teacher lessons by principals perform better (the effect reaches statistical significance at the 10.5% level). Thus, it appears that accountability policies aimed at teachers can also have positive effects on student achievement.

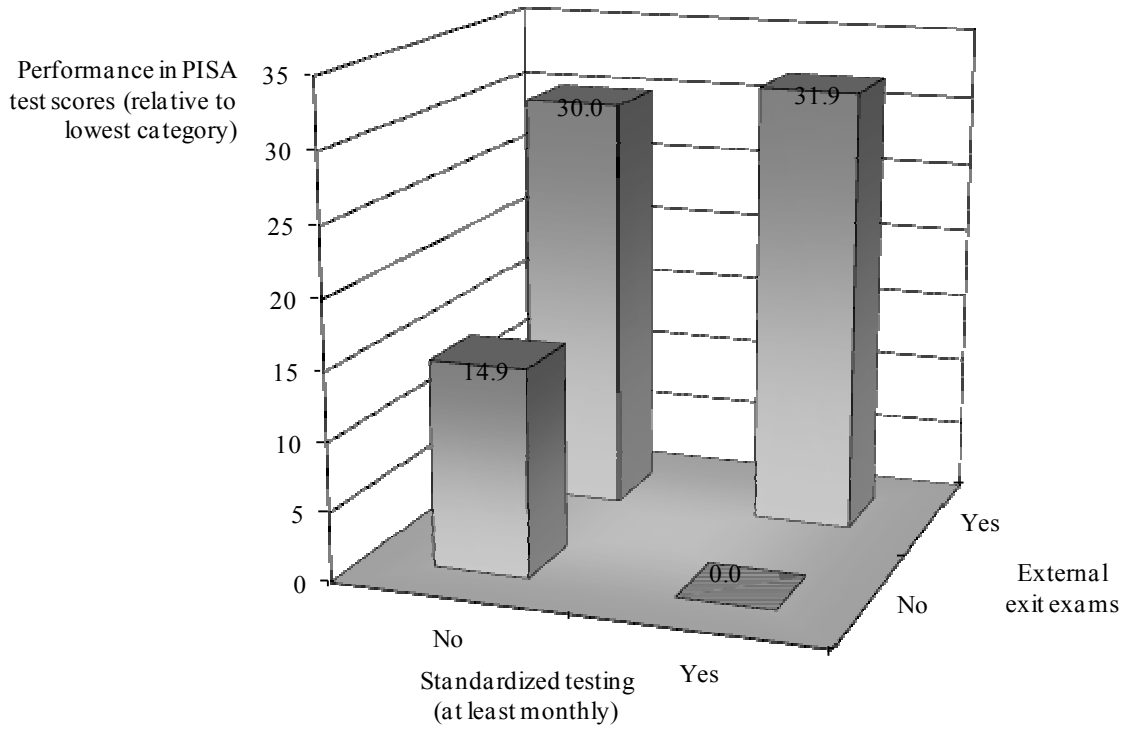
With a sample of only 29 countries, the degrees of freedom available for the country-level analysis are severely limited. The specifications presented in columns (2)-(4) of Table 2 therefore use school-level measures of accountability to test more detailed hypotheses about the effects of accountability policies. Although some caution is warranted in interpreting the results, bias from self-selection and systemic effects is likely to be more limited when analyzing the effects of accountability policies than when comparing private and public schools within the same country.

As the results reported in column (2) reveal, the effects of the various accountability policies are also significant when measured at the school level. In addition, this specification adds variables measuring two further accountability policies (which did not enter significantly in the country-level specification). The first is an alternative measure of teacher monitoring, namely whether inspectors or other persons external to the school have observed classes during the last year to monitor the practice of mathematics teachers at the school, a policy that is quite common in Korea, Switzerland, and the United Kingdom. The results show that teacher monitoring by external inspectors has positive effects on student achievement even after taking into account whether the teachers are also monitored by principals.

The second additional measure is an accountability device aimed not at students or teachers, but at entire schools. More specifically, principals report whether assessments of student achievement are used in their school to compare the school to district or national performance. This is very common in Hungary, New Zealand, the United Kingdom, and the United States, but rare in Austria, Belgium, Denmark, and Greece. The results show that students perform better when their schools use assessments to compare themselves to district or national performance.

Figure 3 depicts the effects of the six different measures of accountability. Note that each of the reported effects is conditional on the presence or absence of the other accountability measures examined. That is, in order to calculate the full effect of having all of the accountability devices relative to none of them, their effects have to be combined. The effects of the five accountability measures that increase student achievement, when measured at the school level, sum to a combined effect of more than 37 PISA test-score points, or the equivalent of more than one and a half grade-level equivalents.

Figure 3: Accountability



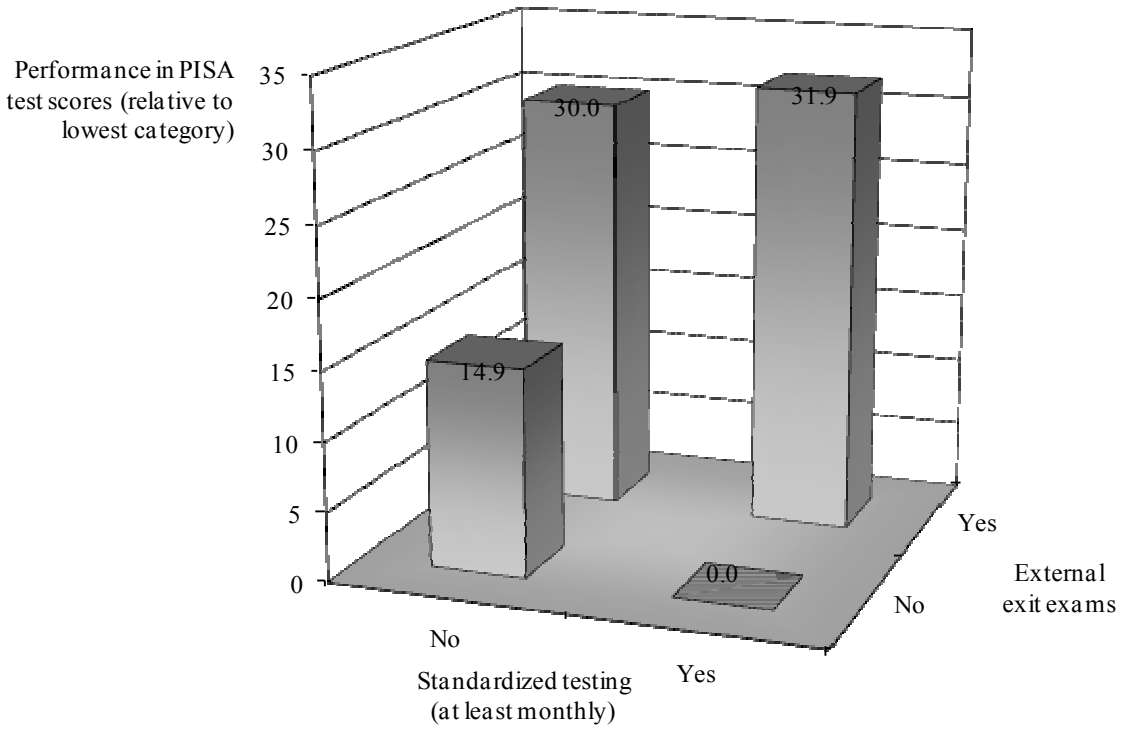
Estimated effect of each institution on PISA test scores. Source: Column (2) of Table 2.

Column (3) of Table 2 adds country fixed effects to the previous model, which becomes possible once the accountability devices are measured at the school level. This specification disregards any variation across countries and compares only schools with and without accountability within each country. The results show that the findings displayed in Figure 3 are very robust to this specification.

Finally, column (4) adds another accountability measure aims at students, namely whether students in the school are assessed using standardized tests at least monthly. Because previous evidence shows that the effect of regular standardized testing can vary between systems with and without external exit exams, the variable is interacted with the country-level measure of external exams.[8] The results, which are depicted in Figure 4, show that regular standardized testing negatively affects student achievement in countries that do not have external exit exams. But its effect differs significantly depending on presence or absence of an external exit exam system: When external exit exams are in place, the negative effect vanishes and turns into an (insignificant) positive effect. This suggests that standardized testing is beneficial only if external exit exams clearly specify the educational goals and standards of the school system, while it can backfire and lead to weaker student achievement without clear external standards.

[8] The external exit exam variable is centered in this specification, so that its coefficient reports the average effect of external exit exams.

Figure 4: External exit exams and standardized testing



Source: Column (4) of Table 2.

In sum, there is ample evidence that accountability is associated with better student achievement. This is true for accountability measures aimed primarily at students, such as external exit exams and the use of assessments for decisions on student promotion and retention; for accountability measures aimed at teachers, such as internal and external monitoring of teacher lessons; and for accountability measures aimed at schools, such as assessments used to compare them to district or national performance. By contrast, if assessments are not used to provide incentives for better performance, but for example to group students, this ability tracking of students even seems to have negative effects. The rich data available in PISA 2003 on different facets of accountability shows that student testing, internal and external teacher inspection, and school accountability can all work towards improving student achievement.

4. AUTONOMY

This chapter provides theoretical background, previous evidence, and new results on the effects of different forms of school autonomy on student achievement. We focus in particular on whether certain forms of autonomy are more or less beneficial when accountability measures are in place.

4.1 Theory: Local Knowledge and Opportunism With and Without Monitoring

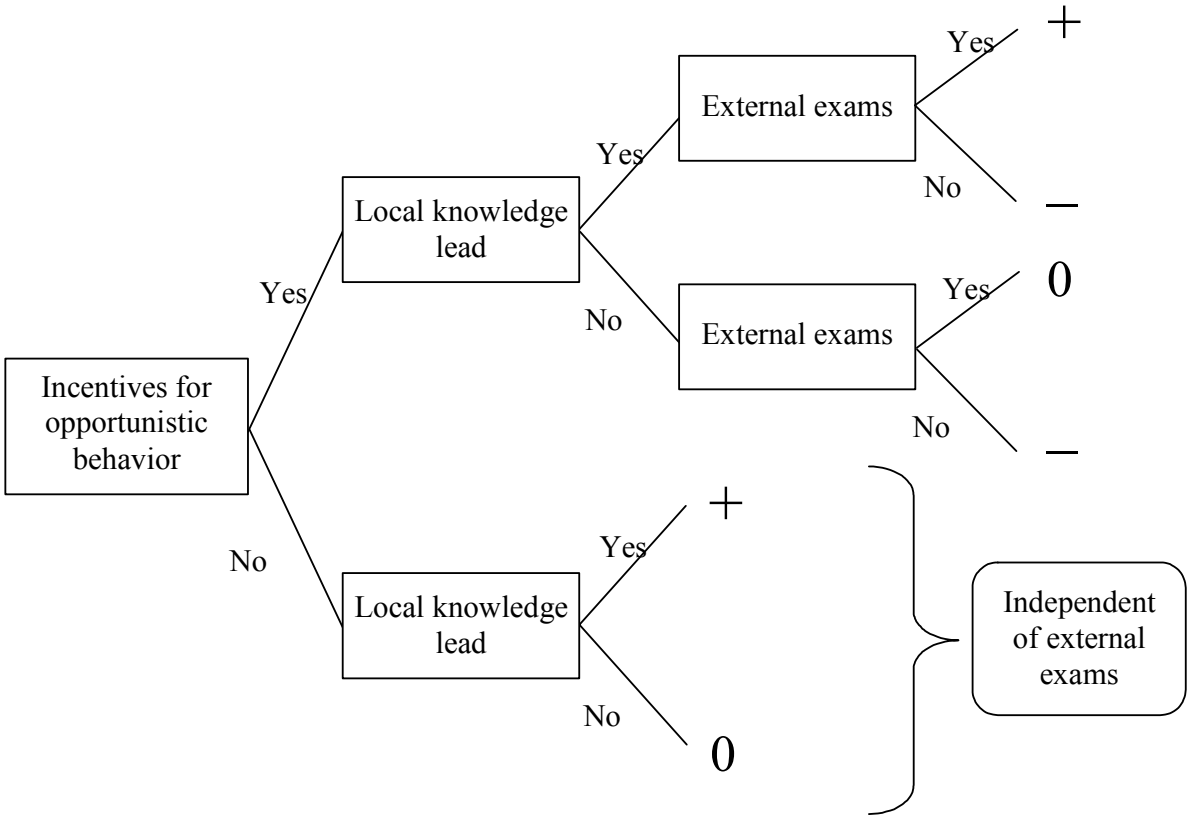
School autonomy or the decentralization of decision-making power can be understood as the delegation of a task by a principal, who wishes to facilitate the provision of knowledge in the school system, to agents, namely the schools (cf. Wößmann 2005a). As discussed in the previous chapter, principal-agent relationships need not always be a "problem": in the absence of divergent interests or asymmetric information, agents can be expected to behave in conformity with the 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). Only where both divergent interests and asymmetric information are present do agents have incentives and opportunities to act in an opportunistic way without risk that such behaviour will be noticed and sanctioned.

The danger of opportunism by decentralized decision-makers is thus limited to those decision-making areas in which their interests diverge from the objective to enhance students' knowledge. This is, for instance, possible whenever the decision concerns the financial position or the workload to be fulfilled by the schools. In such cases, it is rational for the school decision-makers to favor their own interests over the promotion of student achievement as long as possible monitoring agencies such as school boards or parents have imperfect information about the actual behaviour of the schools. In view of the decentralized character of educational provision, there is almost always a high degree of information asymmetry about school behaviour. Nevertheless, it can be at least partially overcome by external exams that supply comparable information about student achievement.

An additional crucial point is that in many decision-making areas, local decision-makers may know much better than a central agency ever could how education services can be most efficiently provided. For example, teachers are likely to have superior knowledge of how to teach their specific students a specific subject. This local knowledge lead can make provision by local agents more efficient than by central planning authorities. But the decisive factor is whether these local decision-makers also have the incentive to exploit their local knowledge in providing educational services. This will be the case only when others become aware of whether they have made the effort to utilize their local knowledge – i.e., only when information asymmetries are bridged, for instance by external exams.

Figure 5 presents the expected effects on student achievement of school autonomy in various decision-making areas characterized by the presence or absence of incentives for opportunistic behaviour and of local knowledge leads.

Figure 5: Effects of autonomy on student achievement depending on external exams



"Incentives for opportunistic behaviour" and "local knowledge lead" are features of the respective decision-making area which can be organized either autonomously or non-autonomously.

+: Autonomy is performance-enhancing. –: Autonomy is performance-reducing. 0: No performance difference between autonomous and central decision-making.

Source: Wößmann (2005a).

In areas where no incentives for opportunistic behaviour exist because the interests of agent and principal are aligned, the expected effects of school autonomy on student achievement are straightforward. If local decision-makers have a knowledge lead in such areas, school autonomy has a positive effect on student achievement because the advantages of local decision-making (better knowledge) exist while the disadvantages (opportunism) do not. If local decision-makers have no knowledge lead in these areas, there will be no difference between decentralized and centralized decision-making. In both cases, it makes no difference for the effect of school autonomy on student achievement whether there are external exams or not, because by definition there is no risk of opportunistic behaviour which would have to be averted.

External exams change the expected effect of school autonomy on student achievement only in decision-making areas that offer incentives for opportunistic behaviour due to the diverging interests of agent and principal. In areas without a local knowledge lead and consequently with no benefits of decentralized decision-making, school autonomy has a negative impact on student achievement without external exams due to local opportunistic behaviour. But with external exams, the risks of negative performance effects due to local opportunistic behaviour are averted, so that student achievement will not differ between systems with autonomous and centralized decision-making.

In decision-making areas containing both incentives for opportunistic behaviour and benefits of superior local knowledge, external exams can avert the disadvantages of opportunistic behaviour, so that the local knowledge lead produces an overall positive effect of school autonomy on student achievement. Without external exams, the advantage of superior local knowledge must be weighed against the disadvantage of opportunistic behaviour, and the net effect of school autonomy depends on the relative size of these two partial effects. It is therefore not obvious whether these decision-making areas yield a slightly positive effect, no effect, or an overall negative effect of school autonomy. Previous empirical results suggest that the negative opportunism effect tends to outweigh the positive knowledge effect, as depicted by the negative net effect in Figure 5. In this case, external exams turn an originally negative effect of school autonomy on student achievement completely around to become a positive effect.

In sum, theory suggests that external exams and school autonomy are in many cases complementary, so that the one is only beneficial if the other is also in place. As a consequence, external exams or other well-implemented accountability systems may be a pre-requisite for decentralized systems of autonomous schools to perform well.

4.2 Existing Evidence

The general pattern of results on school autonomy from previous international student achievement tests is that students perform significantly better in schools that have autonomy in process and personnel decisions (Wößmann 2001, 2003a; Fuchs and Wößmann 2007). These decisions include such areas as the purchase of supplies and budget allocations within schools, hiring and rewarding teachers (within a given budget), and choosing textbooks and instructional methods. The positive performance effects of school autonomy in these kinds of decision-making areas are also found in international tests in primary school (Fuchs and Wößmann 2005).

The existing cross-country evidence also reveals that there are important interaction effects between school autonomy and the accountability introduced by external exams (cf. Wößmann 2007b for a survey). The results show that school autonomy is more beneficial in systems with external exit exams (Wößmann 2005a; Fuchs and Wößmann 2007). In several decision-making areas, external exams even turn an initially negative autonomy effect into a positive effect. For example, in TIMSS and TIMSS-Repeat as well as in PISA 2000, school autonomy regarding teacher salaries has a negative effect on student achievement in systems without external exams. This effect is reversed in systems with external exams so that salary autonomy of schools has positive effects on student achievement.

Similar cases where external exams turn a negative autonomy effect around into a positive effect have been found for such decision-making areas as school autonomy in determining course content and teacher influence on resource funding. More generally, in several additional decision-making areas the general pattern of results suggests that school autonomy is better for student achievement when external exit exams are in place (Wößmann 2005a).

4.3 New Results

Table 3 reports the average effects of different forms of school autonomy on student achievement in PISA 2003. These specifications do not yet consider interactions with accountability measures. Column (1) replicates the results of the basic model, which included two country-level measures of school autonomy. On average, students in countries where most schools have autonomy in staffing decisions, such as Finland, Switzerland, and the United Kingdom, outperform students in countries such as Austria, the Czech Republic, Greece, Norway, Poland, and Sweden where most schools do not have staffing autonomy, after controlling for the standard set of background and other institutional factors.

Table 3: Autonomy

Level at which autonomy is measured:	Country	School	
	(1)	(2)	(3)
Autonomy in formulating budget	-25.056**	-6.640***	-7.261***
	(10.661)	*(1.456)*	*(1.518)*
Autonomy in staffing decisions	29.310*	6.551***	5.731***
	(14.685)	*(1.344)*	*(1.393)*
Autonomy in hiring teachers			4.278***
			(1.532)
Autonomy in establishing starting salaries			-2.109
			(1.478)
Autonomy in determining course content			1.840
			(1.341)
Students	219,794	219,794	219,794
Schools	8,245	8,245	8,245
Countries	29	29	29
R^2	0.386	0.383	0.383

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. Controls include: external exit exams, private operation, government funding, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering in parentheses (column 1: clustering at country level; columns 2 and 3: clustering at school level). Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

By contrast, students in countries with large shares of schools that have autonomy in formulating their own budget, such as Greece, the Netherlands, and New Zealand, perform significantly worse (on average and conditional on the other factors of the model) than students in countries such as Austria, Germany, and Luxembourg where hardly any school has this autonomy.[9] A measure of school autonomy in deciding on budget allocations within the school does not enter significantly when added to the model. It should be noted, though, that this indicator may not be very informative in PISA 2003, because less than five percent of schools report that they lack autonomy on this dimension.

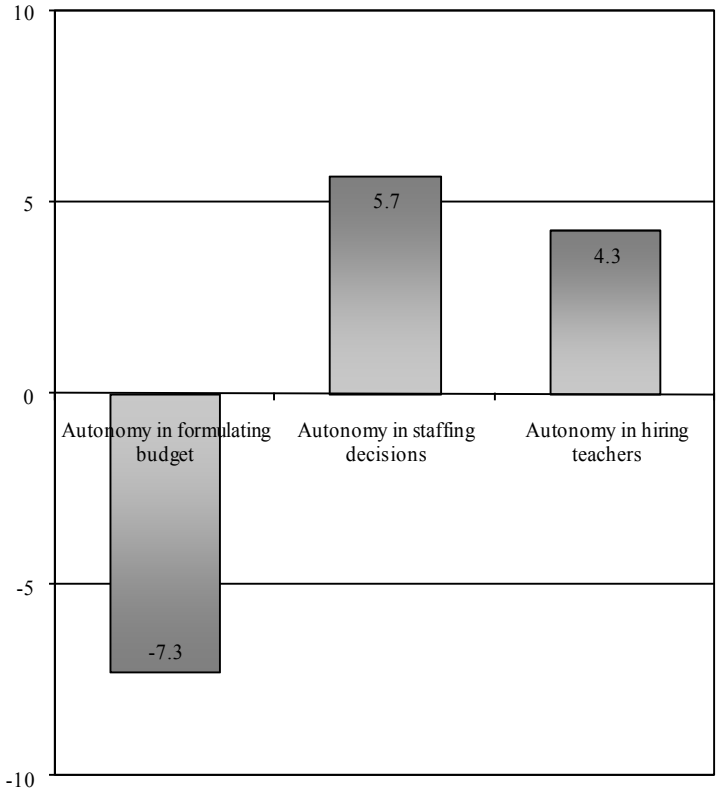
In light of the theoretical background discussed above, these results suggest that the positive effect of superior local knowledge exceeds the negative effect of local opportunism in the case of autonomy in staffing decisions. School leaders seem best able to select the right teachers for their schools, and they do not seem to have strongly divergent interests from advancing student achievement in this decision-making area. By contrast, the negative effect of opportunistic behaviour seems to be bigger than the positive effect of local knowledge leads in the case of autonomy in formulating the budget. In fact, central agencies with budget specialists may have even better knowledge for setting budget levels. At the same time, local interests may diverge from the advancement of educational goals when financial issues are at stake.

[9] The results are based on principals' reports whether formulating the budget was not a main responsibility of their school. Using another questionnaire item, reporting whether the school's governing board exerts a direct influence on decision-making about budgeting (which is not exclusive to alternative bodies), as an alternative measure of budgeting autonomy yields similar results, although usually at lower levels of significance.

The specification presented in column (2) measures the two autonomy variables used in the basic model at the level of schools rather than countries. The pattern of results is the same – autonomy in formulating the budget being negatively associated with student achievement, staffing autonomy positively – although the size of the estimated effects is smaller.

Column (3) adds three additional school-level measures of autonomy to the specification. The first is a more specific measure of autonomy in staffing decisions, namely whether selecting teachers is primarily a school responsibility. This measure asks about teachers specifically, rather than staff in general. In addition, whereas the measure of staffing autonomy reported above did not exclude schools where other bodies also influenced staffing decisions, this measure excludes such schools. This alternative measure of hiring autonomy yields similar results to the more general measure of staffing autonomy, with schools that can hire their own teachers performing significantly better than schools without autonomy in hiring teachers. Interestingly, this effect is independent of the effect of the previous measure of staffing autonomy, suggesting that absence of external interference in school staffing decisions provides additional advantages over more limited control. Countries where virtually all schools can directly hire their own teachers include the Czech and Slovak Republics, Denmark, Hungary, Iceland, Netherlands, New Zealand, Poland, Sweden, the United Kingdom, and the United States, while few schools in Greece, Italy, Luxembourg, Portugal, and Turkey can do so. The statistically significant mean autonomy effects of column (3) are presented graphically in Figure 6.

Figure 6: Autonomy



Estimated effect of each institution on PISA test scores. Source: Column (3) of Table 3.

The other two measures of autonomy in the specification do not enter statistically significantly in the model (which is also true if they are measured at the country level). On average, students in schools that have the main responsibility for establishing teachers' starting salaries – which is true for most schools in the Netherlands and the United Kingdom and for many schools in the Czech Republic, Sweden, and the United States – do not perform significantly different from students in schools without this autonomy – as is generally the case for schools in Austria, Belgium, Germany, Greece, Italy, Norway, and Portugal.[10] Similarly, there is no significant difference on average between students in schools that do not have autonomy to determine the content of their courses – as is the case in Greece and Luxembourg – and students in schools that have the main responsibility for determining course content – as is common in Japan, Korea, the Netherlands, New Zealand, Poland, and the United Kingdom. However, in both cases the average effects foreshadow important differences in the effects of autonomy between systems with and without external exit exams.[11]

4.4 Interaction between Autonomy and Accountability

The theoretical background provided above suggests that whenever there are divergent interests between the principal and the agent in a decision-making area, the effect of autonomy may depend on whether measures are in place that hold schools accountable for their decisions. To capture such interdependence, Table 4 reports results of specifications that include interaction terms between autonomy variables and two measures of accountability. The estimated interaction effects show whether the effect of school autonomy in various decision-making areas differs between school systems with and without accountability devices. The accountability measure used in column (1) is external exit exams. The results suggest that there are indeed significant interactions between autonomy and accountability, which with one exception are positive.

[10] Note that autonomy in hiring and in determining starting salaries is highly collinear with autonomy in firing and in determining salary increases, respectively, which does not allow for including them together. Therefore, the effects of the former may capture some of the potential effects of the latter.

[11] We do not include a measure of autonomy in choosing which textbooks are used, because results proved extremely sensitive to individual countries. For example, the coefficient estimate was significantly positive in the OECD sample, turned significantly negative once Greece was dropped from the analysis, and turned insignificant once other individual countries were also dropped. Similarly, the results of the interaction between textbook autonomy and external exit exams proved very sensitive; without Greece, it showed a result pattern similar to the one reported for budgeting and salary autonomy below.

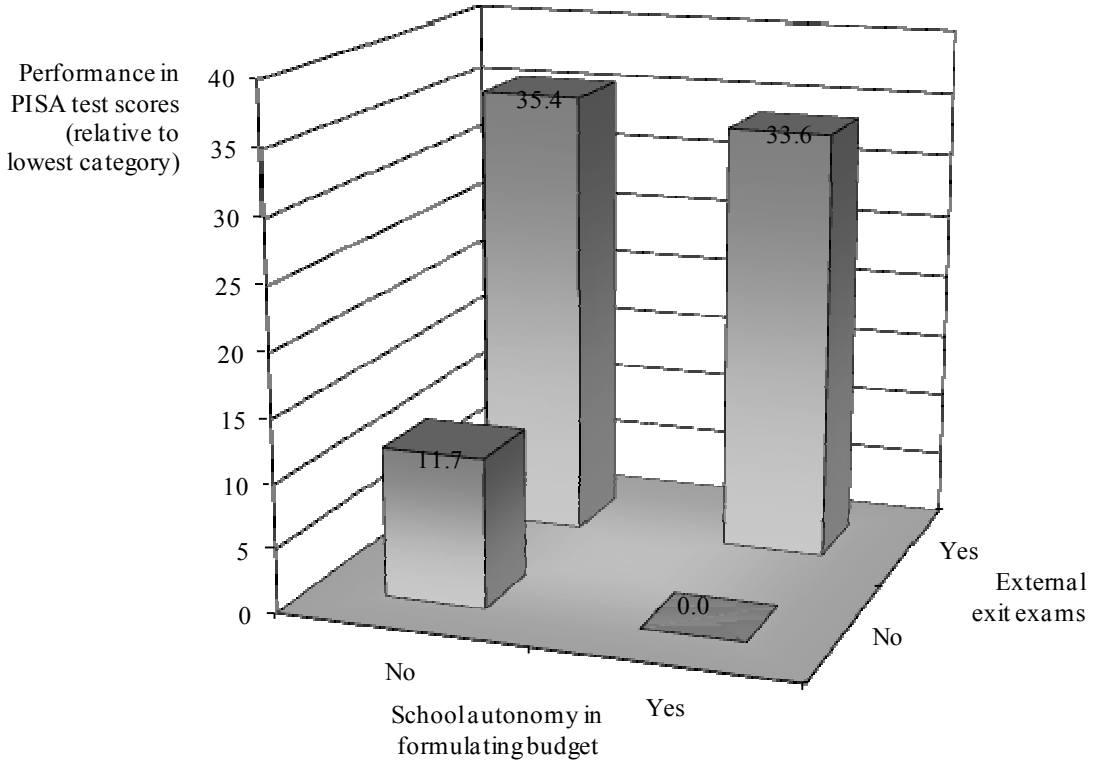
Table 4: Interaction between autonomy and accountability

Accountability measure for interaction:	External exit exams	Assessments used to compare school to district/nation	
Level at which accountability is measured:	Country	Country	School
	(1)	(2)	(3)
External exit exams	23.666**	13.590*	13.609
	(9.714)[a]	*(7.158)*[a]	*(8.509)*[a]
Assessments used to compare school to district or national performance		16.834	7.027**
		(21.318)[a]	*(2.921)*
Autonomy in formulating budget	-11.724***	-6.068***	-11.757***
	(2.192)	*(1.439)*	*(1.912)*
Accountability x Autonomy in formulating budget	9.978***	35.459***	10.883***
	(3.156)	*(5.648)*	*(2.879)*
Autonomy in hiring teachers	24.403***	2.374	13.053***
	(2.194)	*(1.539)*	*(1.840)*
Accountability x Autonomy in hiring teachers	-35.007***	-68.137***	-19.097***
	(3.298)	*(6.417)*	*(2.722)*
Autonomy in establishing starting salaries	-8.272**		
	(3.310)		
Accountability x Autonomy in establishing starting salaries	7.925*		
	(4.060)		
Autonomy in determining course content	-0.517		
	(1.931)		
Accountability x Autonomy in determining course content	4.409		
	(2.879)		
Students	219,794	219,794	219,794
Schools	8,245	8,245	8,245
Countries	29	29	29
R^2	0.386	0.387	0.384

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. The autonomy variables are measured at the school level. Controls include: private operation, government funding, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering at the school level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent. [a] Clustering of standard errors at country level.

The first interaction considered is between autonomy in formulating the school budget and external exit exams. The results, which are also displayed graphically in Figure 7, show that in systems without external exams, school autonomy in formulating the budget has a negative effect on student achievement. In systems with external exit exams, student achievement is generally higher than in systems without external exit exams, both in cases with and without school autonomy. In addition, however, the negative effect of budgetary autonomy on student achievement vanishes in systems with external exit exams. Put differently, the positive effect of external exit exams on student achievement is significantly stronger when schools have autonomy in formulating their budget.

Figure 7: External exit exams and school autonomy in formulating budget



Source: Column (1) of Table 4.

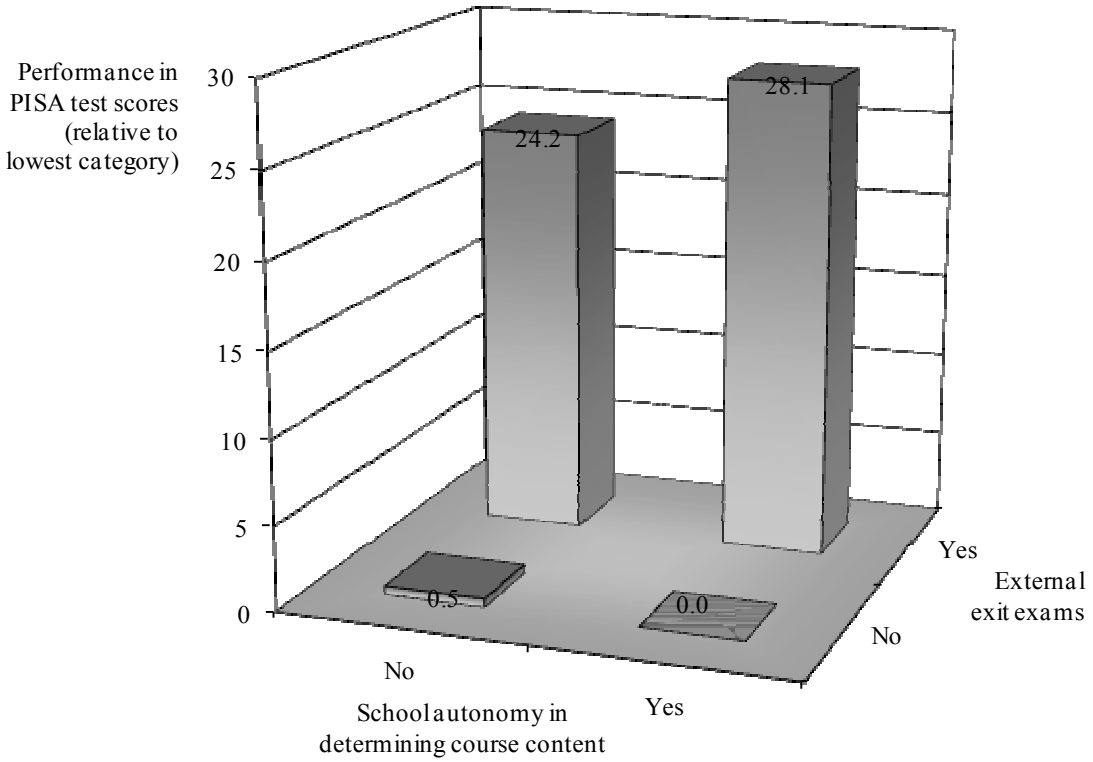
Decisions on formulating the school budget thus appear to involve strong incentives for opportunistic behaviour. Without external exit exams, the negative effect of opportunistic decisions taken by the schools dominates, as local opportunistic behaviour cannot be externally observed and thus cannot be sanctioned. Hence school decision-makers do not feel obliged to make budget decisions in a way that contributes to enhancing student achievement, but can use their decision-making autonomy to promote other interests. The incentives for opportunistic behaviour are to some extent reduced when external exit exams hold schools accountable for their budgetary decisions. External exit exams provide information about whether the schools perform well or not, so that supervisory authorities and parents can draw consequences from poor school behaviour. As a consequence, with external exit exams, any remaining negative effect of opportunistic behaviour and any positive effect of local knowledge leads cancel out, so the combined effect of budgetary autonomy is about zero.

The same pattern of results emerges for school autonomy in establishing teacher salaries. Salary autonomy has a negative effect on student achievement without external exit exams which disappears once external exit exams are in place. Again, it seems that schools behave opportunistically if they are given salary autonomy but are not held accountable for their decisions, while negative and positive effects of salary autonomy seem to cancel out once external exit exams are in place.

In the case of autonomy in determining course content (depicted in Figure 8), the negative effect of autonomy in systems without external exit exams is very small and not statistically significant. This effect

turns to be moderately positive in systems with external exit exams (the interaction term reaches statistical significance at the 12.5% level). This pattern of results suggests that the decision-making area of determining course contents entails both incentives for local opportunistic behaviour and local knowledge leads. The incentives for local opportunistic behaviour may stem from the fact that content decisions influence the workload of teachers, while the local knowledge lead may stem from the fact that teachers probably know best what specific course contents would be best suited for their specific students. Without external exit exams, the two effects cancel out. But when external exit exams limit the negative effects of opportunism, the positive effects of using local knowledge dominate.

Figure 8: External exit exams and school autonomy in determining course content



Source: Column (1) of Table 4.

While in the previous three cases of autonomy, the interaction between autonomy and accountability is positive, the result is different in the case of autonomy in hiring teachers (with a similar pattern of results emerging for the alternative measure of staffing autonomy). In countries without external exit exams, students in schools with hiring autonomy perform better than students in schools without hiring autonomy. In countries with external exit exams, the opposite is true. In the case of mathematics achievement on previous international studies, hiring autonomy has already been an exception to the general rule of positive interactions between accountability and autonomy (cf. Wößmann 2005a). Such a result is difficult to interpret in the framework of the principal-agent model of the educational process. One possible explanation for the pattern would be a selection effect in systems without external exit exams, in that better teachers evade non-autonomous schools and sort into more autonomous schools, an effect that may be less pronounced in the more transparent external-exam systems.

Column (2) of Table 4 presents interaction effects of autonomy with the second measure of school accountability available in this study, namely the use of student achievement assessments to compare schools to district or national performance. The patterns of the interaction effects of school autonomy in budget formulation and in teacher hiring are the same as for external exit exams.[12] This confirms the robustness of the general finding. Moreover, given that the use of assessments for school comparisons is measured at the school level, we can also estimate this specification with a school-level measure of accountability. As the results reported in column (3) reveal, the pattern of results is robust to the school-level measurement of accountability.

The general pattern of results suggests that the effects of school autonomy on student achievement depend on whether schools are held accountable for their decisions. As a general rule, school autonomy seems to be more beneficial when measures of school accountability, especially external exit exams, are in place. Accountability and autonomy seem to be complementary in any decision-making area that includes scope for opportunism and local knowledge leads.

[12] The interactions of the second accountability measure with autonomy in establishing salaries and determining course content are not statistically significant.

5. CHOICE

This chapter analyzes the effects of different aspects of choice and competition on student achievement. After presenting some theoretical background and reviewing existing international evidence, it presents new cross-country evidence on how private school operation, government funding of schools, and parental choice among public schools affect the achievement of students. In addition, we present new results on how the effects of choice interact with the existence of accountability and autonomy, a question not previously examined.

5.1 Theory: Competition Created by Choice, Private Operation, and Public Funding

There has been much recent debate about the merits of demand-sensitive schooling (OECD 2006a). Economic theory suggests that additional choice – both among public schools and between public and private schools – can improve student outcomes by allowing consumers (i.e. parents) to choose the suppliers of schooling that offer the best performance. Assuming parents value academic outcomes, the resulting competition among schools to attract students should enhance overall student achievement.

Privately operated schools are often predicted to be more efficient than publicly operated schools not only because market forces create incentives for performance-conducive qualitative innovation and efficient resource use, but also because private schools typically face fewer regulations than do government-run schools (e.g., Chubb and Moe 1990; Hanushek with others 1994; Shleifer 1998). The existence of private schools may also improve the performance of nearby public schools with which they compete, because losing students will ultimately reduce public school budgets. In the same way, parental choice among public schools is often expected to have positive effects on student outcomes to the extent that public school budgets reflect enrollment. At the same time, if choice among public schools is limited to fiscally independent units such as school districts, any competitive effects may be limited.

In terms of the relative merits of public and private funding (as opposed to operation) of schools, it is sometimes argued that private or parent-based funding can increase accountability and provide incentives for efficient behaviour from the demand side (e.g., Jimenez and Paqueo 1996). It is not obvious, though, that this potential benefit of private involvement would augment the benefit of private provision and parental choice among schools, which should already create performance-conducive incentives.

In fact, this last point suggests an opposite case favoring public funding, if combined with the idea that some families will lack sufficient resources to choose privately operated schools if they are also privately funded (Wößmann 2006). As long as there are credit constraints that prevent poor families from borrowing against possible future income gains of their children due to improved educational performance (cf. Loury 1981; Gradstein, Justman, and Meier 2004), poor families' choices of schools that require private funding will be constrained. Generous public funding of privately operated schools can relax such credit constraints, thereby allowing greater choice for all families and increasing schools' incentives to behave efficiently.

5.2 Existing Evidence

The available cross-country evidence on the effects of choice on student achievement is limited to the effects of private involvement in the operation and financing of schools. At the level of individual schools, students perform better across all the countries participating in the PISA 2000 tests if their specific school is privately managed (Fuchs and Wößmann 2007). This pattern is not uniform across countries, however,

as revealed when using the data from international achievement tests to estimate the effect within countries (cf. Wößmann 2006). Toma (1996; cf. also 2005) similarly estimates the effect of private school operation in five countries using the 1981 second international mathematics test, noting that the positive effect of private provision is independent of whether the countries tend to finance the schools publicly or not. Estimating the effect of private school operation in eight countries in PISA 2000, Vandenberghe and Robin (2004) find positive effects only in some countries, but they do not account for differences in the source of school funding. Because these studies are all based on observational data, they may suffer from selection bias if students with better (or worse) aptitudes for learning or families with greater (or lower) commitment to education are more likely to choose private schools, and if the available control variables do not fully account for these differences.[13]

Just as importantly, however, studies that compare the relative performance of private and public schools within a country may miss an important aspect of the effect of choice, because the competition created by private schools may affect the performance of nearby public schools. Both private and public schools may perform at a higher level because of the existence of private competition. If public schools behave differently because there are private schools nearby, private involvement could enhance overall achievement even if performance does not differ between individual private and public schools.

These systemic effects are best captured by measuring the effect of the share of privately managed schools on overall student achievement at the system level.[14] The international evidence on system-wide positive effects of competition from privately managed schools is substantially stronger than the evidence comparing private and public schools within the same system. In TIMSS, students perform substantially better in countries where more schools are privately managed and where a higher share of public educational spending goes to private institutions (Wößmann 2001, 2003a). Similarly, students in countries with a larger share of privately managed schools perform substantially better in PISA 2000 (Wößmann 2006). At the same time, across countries, larger shares of public funding (as opposed to management) are associated with better student achievement in PISA 2000. Thus, countries which combine relatively high shares of private operation with relatively high shares of government funding do best among all possible operation-funding combinations, while countries which combine public operation with private funding do worst.

Furthermore, Wößmann (2006) finds that the achievement advantage of privately operated schools over publicly operated schools at the school level is particularly strong in countries with large shares of public funding. This suggests that public funding may help additional families to choose privately managed schools, increasing the extent of choice and competition in the system. The existing international evidence therefore suggests that school systems based on public-private partnerships in which the government finances schools but contracts their operation out to the private sector are the most effective in terms of fostering students' educational achievement.

---

13 There is also a lot of national evidence suggesting that school choice can improve student achievement. For evidence that student achievement in privately managed schools exceeds achievement in publicly managed schools see, among others, Howell, Wolf, Campbell, and Peterson (2002), Hoxby (2003), and Neal (1997) for the United States, Bradley and Taylor (2002) and Levaĉić (2004) for England, Sandström and Bergström (2005) and Björklund, Edin, Freriksson, and Krueger (2004) for Sweden, and Angrist, Bettinger, Bloom, King, and Kremer (2002) for Colombia. Some of the empirical contributions also show that the existence of privately managed schools improves the performance of nearby public schools that face their competition (e.g., Hoxby 2003; Sandström and Bergström 2005; Björklund, Edin, Freriksson, and Krueger 2004). Furthermore, Hoxby (2000) presents U.S. evidence that more competition between public schools within the public system can improve student achievement.

14 Wößmann (2007c) finds positive effects of larger shares of private school operation in a cross-regional study of the school systems of the different German states.

5.3 New Results

We start our choice analyses with measures of the private vs. public involvement in the operation and funding of schools. In the PISA school background questionnaire, principals of tested schools report whether their school is a private school, which is managed directly or indirectly by a non-government organization (e.g. a church, trade union, business, or other private institution), or a public school, which is managed directly or indirectly by a public education authority, government agency, or governing board appointed by government or elected by public franchise. More than three quarters of 15-year-old students in the Netherlands attend privately operated schools. Private school shares in Belgium, Ireland, and Korea are also well above one half. By contrast, the share of privately operated schools in Greece, Iceland, Italy, New Zealand, Norway, Poland, Sweden, and Turkey is below five percent. Principals also report the share of their schools' total funding that comes from different government sources, as opposed to parental fees and other private contributions. While the share of government funding lies below 60 percent on average in Korea, Mexico, and Turkey, many countries such as Finland, Germany, Iceland, Luxembourg, the Netherlands, Norway, Poland, Sweden, and Switzerland have an average share of government funding above 95 percent.

Column (1) of Table 5, which replicates the basic model, measures both the share of privately operated schools and the average share of government funding at the country level. The aggregation to the country level circumvents problems of self-selection of students into private and public schools within countries and captures potential systemic effects of private competition on the performance public schools. The results show that a larger share of privately operated schools is associated with better student achievement. At the same time, students perform better where the average share of government funding is larger. As suggested by the theoretical background above, both private operation and government funding increase the extent of choice in the system, and the result seems to be better learning outcomes for students. Both effects are quite large. For example, the estimated difference in achievement between a system like the Netherlands with three quarters of schools privately operated and systems such as Iceland, Norway, and Poland with hardly any private schools is equivalent to more than what students on average learn during two years.

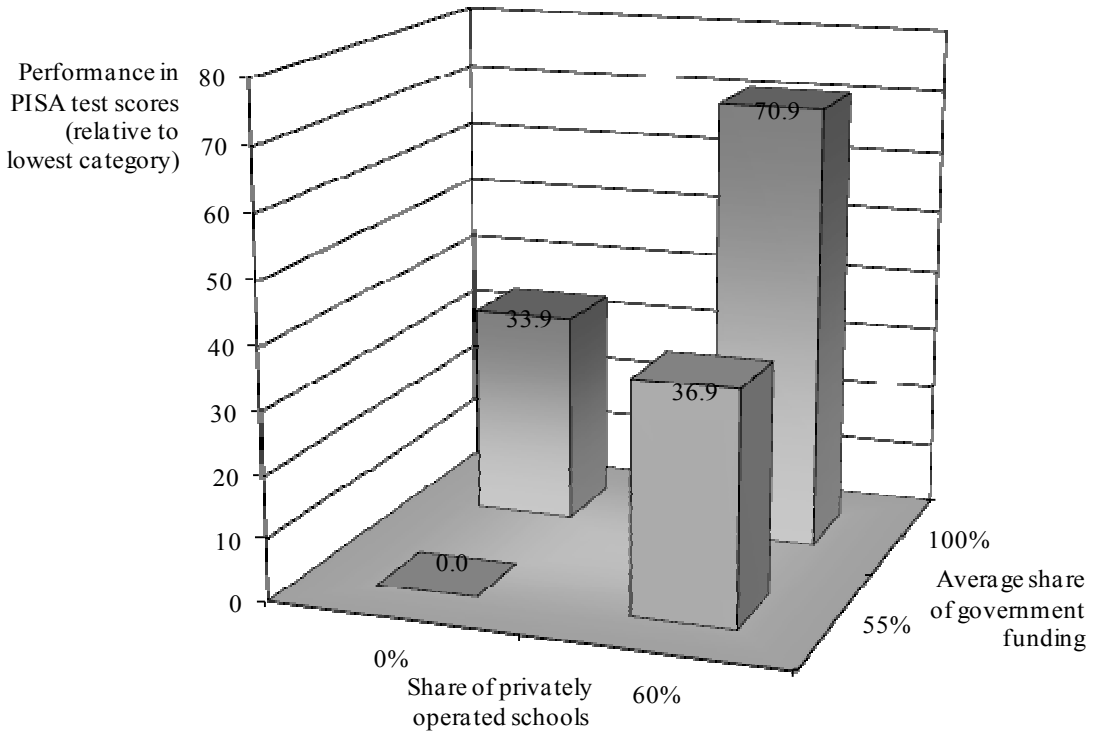
Table 5: Private operation and government funding

Level at which choice is measured:	Country			School
	(1)	(2)[a]	(3)	(4)
Private operation	61.563***	72.722***	38.385***	17.836***
	(10.419)	*(15.420)*	*(13.033)*	*(1.810)*
Government funding	75.437***	81.245***	81.124***	12.531***
	(20.901)	*(19.839)*	*(22.995)*	*(3.411)*
Difference in government funding between public and private schools			-30.239**	
			(12.665)	
Students	219,794	219,794	202,646	219,794
Schools	8,245	8,245	7,731	8,245
Countries	29	29	27	29
R^2	0.386	0.386	0.394	0.377

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. Controls include: external exit exams, autonomy in formulating budget, autonomy in staffing decisions, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering in parentheses (columns 1-3: clustering at country level; column 4: clustering at school level). Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent. [a] PISA measure of private operation instrumented with measure from official enrollment statistics.

Figure 9 provides a graphical depiction of this result. It presents the relative performance at the first decile of the international distribution (below which are only 10 percent of countries) and at the ninth decile (above which are only 10 percent of countries) of both private operation and government funding. For private operation, these are roughly 0% and 60%, respectively, and for government funding, these are 55% and 100%, respectively. The most performance-enhancing combination of public-private partnerships in the school system is where most schools are privately operated, but all are fully publicly funded.

Figure 9: Private operation and government funding



Source: Based on column (1) of Table 5. The two percentage values constitute the first and ninth decile on the two institutional measures, respectively.

In this specification, the share of enrollment in private schools in each country is constructed from the reports of the school principals of the 15-year-olds tested in the PISA study. As an alternative, there is an OECD (2006b) indicator of the share of students in lower secondary education that is enrolled in private schools, which stems from official enrollment statistics of the countries. This alternative measure may provide more encompassing information on the average competitive climate in each country than the measure based on the PISA sample, and it has a different source of measurement error. Using the alternative measure of private operation instead of the PISA-based measure yields qualitatively similar results (with a coefficient on private operation of 68.4 and on government funding of 60.4). This is not surprising, given that the two measures are strongly correlated at the country level (correlation coefficient of 0.758), a relationship that suggests that the PISA-based measure is in fact reliable.

With two measures of private operation that have different sources of measurement error, we can reduce biases due to measurement error and obtain an improved estimator by instrumenting the one

measure with the other. As column (2) of Table 5 shows, our results are very robust to this instrumental variable specification. The coefficient on the instrument in the first stage is close to 1 (0.94) and its *F*-statistic is 60.1, both of which enhance the credibility of the results.

The measure of government funding used in the specifications reported above includes all funding that does not come from parental fees, contributions and donations by benefactors, and other sources. When restricting the non-private funding share to only that part which is paid directly by parents in the form of student fees or school charges, the same negative effect of private funding is evident.

The theoretical background above suggested that government funding can increase the extent of choice and competition in a system in particular when it enables poor families to choose privately operated schools. Thus, the difference in government funding between publicly and privately operated schools in a country measures to what extent privately operated schools are indeed on a par with publicly operated schools in terms of the availability of government funding, and thus in terms of accessibility to the general public. In other words, it is a measure of the fairness of competition between public and private schools. In countries such as Finland, Korea, the Netherlands, the Slovak Republic, and Sweden, privately operated schools receive about the same share of government funding than publicly operated schools on average. The difference is also very small at around 10 percent in Belgium and Ireland, the other two countries (apart from Korea and the Netherlands) with very large shares of privately operated schools. By contrast, the difference in the share of government funding between publicly operated and privately operated schools is around 90% in Greece, the United Kingdom, and the United States, where most privately operated schools receive virtually no government funding at all.

Column (3) of Table 5 reports the results of adding the difference in government funding between public and private schools to the model presented in column (1). The results reveal that students in countries where privately operated schools receive less government funding than publicly operated schools perform significantly worse than students in countries where public funding is equalized between privately and publicly operated schools. This difference in the relative accessibility of privately operated schools accounts for nearly half of the superior performance of students in countries with larger shares of privately operated schools. The total effect of inequality in government funding between school types should be estimated in a model that does not control for the share of schools that are privately operated, because the former will influence the size of the latter. In such a model, the difference in student achievement between a country that has full government funding of public schools but provides no government funding to private schools and a country that puts both types of school on par in their share of government funding is estimated to be 47.3 PISA test-score points, or more than two grade-level equivalents. In short, a level playing field between public and private schools in terms of government funding seems to create an environment of choice and competition that raises student achievement.

Column (4) of Table 5 reports results of a model that measures both whether a school is publicly or privately operated and the share of government funding at the school level.[15] The positive effects of private operation and government funding are robust to this specification, though the magnitude of the coefficient estimates is substantially reduced.[16] The smaller size of the effect of government funding in this specification may be attributable to the selection bias due to credit constraints suggested above. Within each country, children from rich families, who may have higher educational achievement for other reasons such as a more conducive educational climate at home, may tend to select into schools that require large

15 Adding an interaction term between private operation and government funding at the school level yields a positive but statistically insignificant coefficient.

16 In fact, when adding both the country-level and the school-level measures of the two variables in the same model, the school-level measures are statistically insignificant and the whole effects are captured by the country-level measures.

shares of private funding. This selection process, which makes privately funded schools appear better than they are, operates within countries but not between countries.

The fact that the effect of private operation is larger when measured at the country level than at the school level may be explained by systemic effects of competition from privately operated schools on publicly operated schools in the same system. Such systemic effects would affect the average achievement of a system, but would not necessarily show up in the difference between individual privately and publicly operated schools. Thus, a large part of the country-level effect of private operation may stem not from better achievement within privately operated schools, bur rather from better achievement of all schools, public and private, exposed to the competition of private schools.

While it is reasonably straightforward to measure the extent of private school choice, measuring the extent of choice among public schools is more problematic. The student background questionnaire in PISA 2003 provides two measures that may serve as proxies for public school choice. Students are asked for the reasons why they attend their school. One option is that this is the local school for students who live in their area. This may proxy for the fact that students are required to attend the school in their local catchment area and thus indicate a lack of parental choice among schools. However, three caveats are in order. First, attending the local school does not necessarily mean a lack of choice, but may just mean that the local school happened to be the school of choice. Second, reporting that being the local school is a reason for attendance may also indicate that the student has strong social attachments to his or her local community, which may directly affect student achievement. And third, even if choice among public schools is restricted by catchment areas, there may be substantial choice among public schools if the population is mobile and considers school quality in decisions about where to live. Across the OECD countries, less than 10 percent of students in Austria and Italy report that they attend their school because it is the local school, while most students report doing so in Iceland and Norway.

Students also report whether they attend their school because this school is known to be a better school than others in the area. Because this item explicitly refers to a comparison of the specific school to other schools, it may indicate exerted choice among schools. But again, there is a caveat: People who explicitly exert choice may also differ in other regards from people who do not make explicit choices, even though both may have had the same opportunity to exert choice. Few students in Finland, Iceland, Norway, Sweden, and Switzerland report that they attend their school because it is known to be better than others, while roughly every second student says so in Australia, Ireland, New Zealand, Turkey, and the United Kingdom.

Column (1) of Table 6 reports results of a specification that adds these two proxies of public school choice, measured as averages at the country level, to our basic model. Both measures do not enter statistically significantly in the model, and their point estimates are opposite of what would be expected if public school choice played a performance-enhancing role. It seems that either the degree of choice within the public school sector does not drive international differences in student achievement, or the two available indicators are poor proxies for the underlying concept.

Table 6: Public school choice

Level at which choice is measured:	Country	School		
	(1)	(2)	(3)	(4)
Attending school because local	22.938	6.973***	10.430***	7.788***
	(17.806)	*(0.772)*	*(0.963)*	*(0.883)*
Attending school because better	-43.492	9.312***	5.997***	7.937***
	(27.435)	*(0.797)*	*(0.988)*	*(0.961)*
Urban			13.050***	12.181***
			(2.113)	*(2.000)*
Urban x Attending school because local			-10.277***	-10.148***
			(1.761)	*(1.605)*
Urban x Attending school because better			8.812***	7.773***
			(1.905)	*(1.834)*
Country fixed effects	no	no	no	yes
Students	219,794	219,794	219,794	219,794
Schools	8,245	8,245	8,245	8,245
Countries	29	29	29	29
R^2	0.389	0.389	0.390	0.417

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. Controls include: external exit exams, autonomy in formulating budget, autonomy in staffing decisions, private operation, government funding, 15 student characteristics, 16 family background measures, 7 measures of school location and resources, expenditure per student, GDP, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering in parentheses (column 1: clustering at country level; columns 2-4: clustering at school level). Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

A second indicator of the extent to which students are required to attend their local school can be constructed based on the responses of school principals. They report in the PISA school background questionnaire whether residence in a particular area is a prerequisite or high priority for admission to their school. The two measures of admission based on local residence (from the principals) and attending a given school because it is the local school (from the students) seem to be measuring the same concept, as the cross-country correlation is 0.84. Using the one measure as an instrument for the other measure can again improve the estimator by reducing measurement error bias. While the first stage of the instrumental variable regression reveals a strong relationship between the two measures (the coefficient of the instrument is 0.71, the F-statistic is 115.1), the effect of being restricted to attending the local school in the second stage is statistically insignificant and close to zero (coefficient estimate 6.4, standard error 20.5).

Column (2) of Table 6 measures the two proxies of public school choice based on student responses at the individual level. In this specification, students who attend their school because it is the local school and students who attend their school because it is known to be better than others both show higher educational achievement. In this specification, however, it is not clear to what extent the estimated effects capture the effects of public school choice or the effects of being locally attached and of coming from a family that exerts choice.

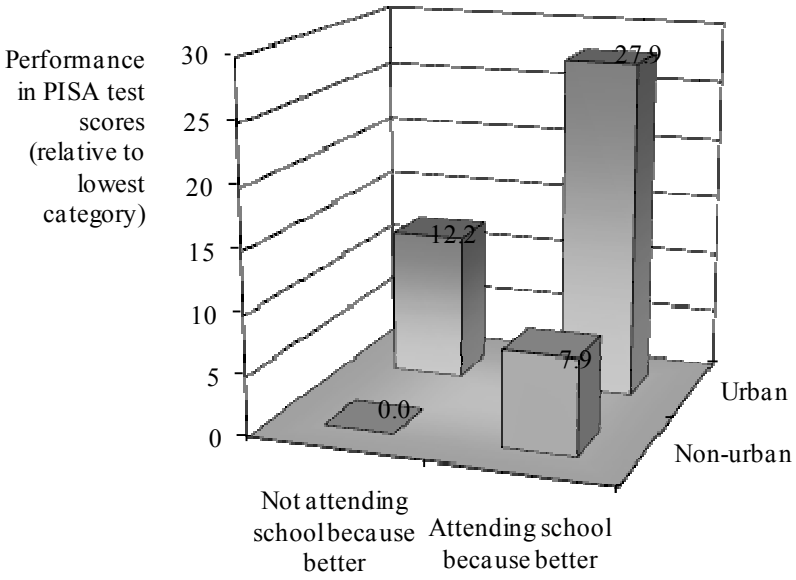
One way to disentangle the effect of choice among public schools from likely biases is to compare how the variables operate in rural and urban areas. Public school choice can only have performance-enhancing competitive effects if there are multiple schools available from which to choose. This is a given in urban areas, but is not necessarily the case in rural areas. We combine the two largest response categories of our control variable measuring the size of the community in which schools are located,

together indicating cities with at least 100,000 inhabitants, as a proxy for the density of schools available for choice (relative to villages and smaller towns with fewer or no schools to choose from).

Column (3) of Table 6 reports results of a specification that includes interaction terms between this indicator of large urban areas and the two proxies of public school choice. Column (4) reports the same specification with country fixed effects, which do not change the results qualitatively. The first thing to note is that on average, students perform significantly better in urban areas. This in itself may partly be the result of having a greater choice of schools, although it may also capture any other difference between rural and urban areas that is not captured by our control variables. More importantly, the interaction terms are both statistically significant: The positive effect of attending the school because it is better than others is larger in urban areas, while the positive effect of attending the school because it is the local one becomes negative in urban areas.

Graphical depictions are helpful in interpreting these results. Figure 10 depicts the interaction between urban areas and attending a school because it is seen as better than others in the area. The main difficulty in interpreting the effect of choosing a better school is that it may just capture the selectivity of more involved parents exerting choice. However, as long as this selectivity is the same in rural and in urban areas, the *difference* in the effect of choosing a better school between urban (15.7 = 27.9-12.2) and rural (7.9 = 7.9-0.0) areas should indicate the true effect of having more public schools to choose from, because there is a greater choice of schools in urban areas. Because the model controls for the main effect of being in an urban area, general differences between families in urban and rural areas do not bias the estimate. As Figure 10 shows, the effect of attending a school because it is considered to be a better one is larger in urban areas than in rural areas. This difference of 7.8 PISA test-score points is probably our best estimate of the effect of being able to choose among public schools because it circumvents many of the usual selectivity problems. (Econometrically, this approach is equivalent to a differences-in-differences estimator, where the first difference is between schools in urban and non-urban areas and the second difference is between affirmative and negative responses on the indicator of public choice.)

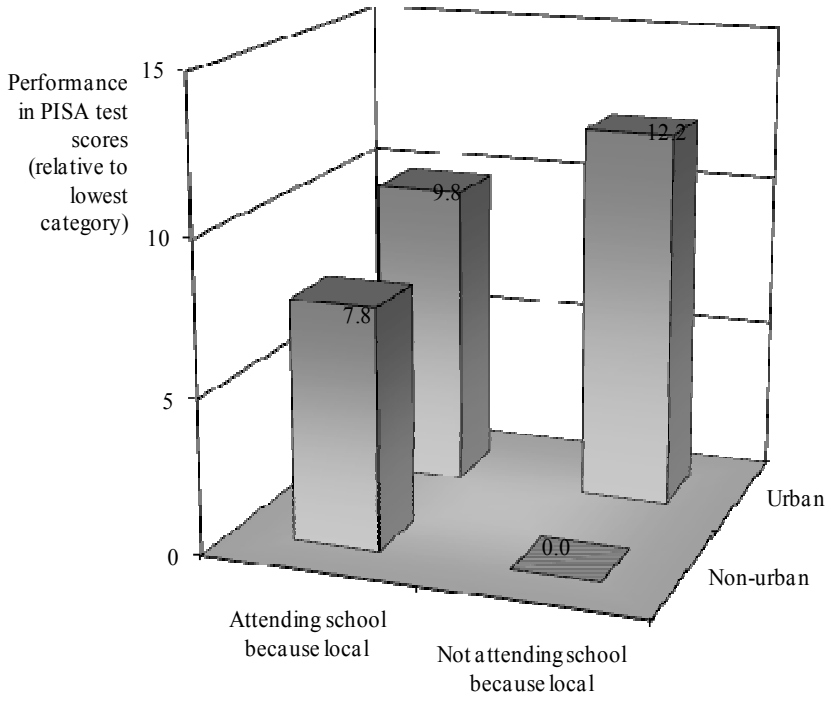
Figure 10: Choice of better schools in rural and urban areas



Source: Column (4) of Table 6.

Figure 11 depicts the interaction between living in an urban area and attending a school because it is the local one, our rough proxy for a lack of public school choice. As discussed before, a main concern with this indicator is that local school attendance may also proxy for effects of a student's general involvement in the local community. Even in non-urban areas, where it is likely that there is only one school within a reasonable commuting distance, reporting that one attends this one school because it is the local one is associated with achievement that is higher by 7.8 PISA test-score points. Again, however, as long as the benefits of local attachment are similar in rural and urban areas, comparing the effect of the public school choice proxy in urban and rural areas will eliminate the effects of local attachment. Attending a school because it is the local one in an urban area means that choice was restricted, while attending a school because it is the local one in a rural area more likely means that there were no additional schools from which to choose. The effect of attending a school because it is local in rural areas is 10.1 PISA test-score points larger than in urban areas. If we assume that local attachment effects are the same across areas, this implies that restricting public school choice reduces student achievement by 10.1 PISA test-score points. Even if we assume that local attachment effects are only given in rural areas and no such effects are given in urban areas, the effect of not attending the local school within urban areas where there are several schools to choose from is a statistically significant 2.4 PISA test-score points.

Figure 11: Attending the local school in rural and urban areas



Source: Column (4) of Table 6.

In sum, there is strong evidence that the extent to which public schools have to compete with private schools increases student outcomes substantially. Student achievement is additionally enhanced where there is more choice because government funding policies create a level playing for privately and publicly operated schools. The extent to which choice among public schools can add to these positive effects of

private school choice is less clear, but there is some indication that being able to choose a better school and not being forced to attend the local school yield additional gains in student achievement.

5.4 Interaction between Choice and Accountability

Some interesting interactions between the effects of choice and accountability may be expected at the system level. For example, choice-based systems may function better if system-wide accountability systems create comparable information on educational achievement. Unfortunately, the coefficient estimates for interaction terms between country-level variables that would test this hypothesis prove highly sensitive to the inclusion of other interaction terms in the model, suggesting that the available degrees of freedom are insufficient to identify interactions between choice and accountability at the country level.

We therefore only pursue the following school-level question in this section: Do the effects of different forms of accountability differ between public and private schools? In particular, does exposure to external accountability help or hinder the achievement of privately operated schools?

Table 7 reports the results of including interaction terms between the different school-level accountability policies and the indicator of whether the school is privately operated. The two columns report the results of one single specification, with the first column reporting the main effect (in effect capturing the effect of accountability in publicly operated schools) and the second column the interaction of the specific accountability measure with the indicator of privately operated schools (in effect capturing the difference in the effect of accountability between publicly and privately operated schools).

Table 7: Interaction between private operation and accountability

	Main effect	Interaction with private operation
	(1)	
Private operation	4.149	
	(3.912)	
Assessments used to make decisions about students' retention/promotion	11.715***	1.947
	(1.853)	*(3.551)*
Assessments used to group students	-7.050***	5.113
	(1.424)	*(3.520)*
Monitoring of teacher lessons by principal	3.357**	-2.080
	(1.479)	*(3.526)*
Monitoring of teacher lessons by external inspectors	0.716	15.330***
	(1.647)	*(3.622)*
Assessments used to compare school to district or national performance	-3.119**	15.480***
	(1.390)	*(3.619)*
Students	219,794	
Schools	8,245	
Countries	29	
R^2	0.382	

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. All six institutional variables are measured at the school level. Controls include: external exit exams, school-level government funding, autonomy in formulating budget, autonomy in staffing decisions, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering at the school level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

The results reveal that the effects of the three forms of accountability policies that are internal to the school – use of assessments for retention and promotion, use of assessments to group students, and monitoring of teacher lessons by the principal of the school – do not differ between publicly and privately operated schools. But the effects of the two forms of accountability policies that are external to the school – monitoring of teacher lessons by external inspectors and use of assessments to compare the school to district or national performance – are significantly larger in privately operated schools. In fact, their effects are close to zero in public schools, but strongly positive in private schools.

These results suggest that private schools in particular benefit from the accountability created by external inspection and performance comparisons with other schools. Relatively autonomous private schools seem to require external accountability, and parents seem to require the information generated by policies of external accountability in order to make well-informed choices.

5.5 Interaction between Choice and Autonomy

An equivalent school-level question can be addressed for the interaction between choice and autonomy: Do privately operated schools work differently in an environment where all schools have autonomy to respond to the competitive forces resulting from parental choices, as compared with an environment in which schools have less autonomy? To answer this question, Table 8 includes interaction terms between the two country-level measures of autonomy used in our basic model and the school-level indicator of private operation.

Table 8: Interaction between private operation and system-level autonomy

	Main effect	Interaction with private operation
	(1)	
Private operation	13.973***	
	(1.911)	
Autonomy in formulating budget	-22.220***	38.126***
	(2.956)	*(8.958)*
Autonomy in staffing decisions	32.955***	20.806***
	(2.581)	*(6.819)*
Students	219,794	
Schools	8,245	
Countries	29	
R^2	0.379	

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. The two autonomy variables are measured at the country level, private operation is measured at the school level. Controls include: external exit exams, school-level government funding, 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering at the school level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

The results show that there are strong positive interactions between school-level private operation and both country-level measures of autonomy. That is, privately operated schools perform even better if schools in the system are generally autonomous, be it in formulating the budget (a decision-making area in which the main effect of autonomy is negative) or in staffing decisions (where the main effect is positive). These results suggest that the incentives created by parental choice of private schools work particularly well if (private and public) schools in the system have autonomy to respond to the parental demands. In such systems, privately operated schools face particularly strong incentives to perform well.

6. NON-COGNITIVE SKILLS

We have focused so far on the effects of accountability, autonomy, and choice on cognitive outcomes, especially student achievement in mathematics as measured by the PISA 2003 test. This chapter provides a complementary analysis of how the same three institutional features affect students' non-cognitive outcomes.

6.1 Background: Economic Outcomes and Policy Determinants of Non-Cognitive Skills

"Non-cognitive skills" is an overarching term used to refer to a range of behaviours, habits, and attitudes that are not measured by conventional tests of cognitive ability. A recent and growing literature suggests that the labor-market benefits of high-quality schooling accrue not only due to the improved cognitive skills measured by such tests but also to changes in non-cognitive skills (cf. Duncan and Dunifon 1998; Dunifon, Duncan, and Brooks-Gunn 2001; Heckman 2000; Cunha, Heckman, Lochner, and Masterov 2006). For example, a recent study by Heckman, Stixrud, and Urzua (2006) found that higher levels of educational attainment improved non-cognitive skills and that such skills had effects on wages that were similar in size to the effects associated with cognitive skills. Differences in non-cognitive skills have also been shown to play an important role in explaining the relatively poor performance of holders of the General Educational Development (GED) credential in the U.S. labor market (Heckman and Rubinstein 2001).

There are several channels through which non-cognitive skills may influence economic success. They are known to contribute to academic achievement (Wolfe and Johnson 1995; Duckworth and Seligman 2005). Better non-cognitive skills may also lead students to complete more schooling, as suggested by the fact that gaps in non-cognitive skills between men and women seem able to explain gender gaps in college attendance rates in the United States (Jacob 2002). Finally, non-cognitive skills seem to raise wages by directly increasing productivity on the job (Heckman, Stixrud, and Urzua 2006).

Although the importance of non-cognitive skills for labor market outcomes is by now well-established, there is very little evidence available on how policy shapes the development of those skills (Deke and Haimson 2006). Yet non-cognitive skills are particularly interesting from a policy standpoint because they may be more malleable than cognitive skills – and therefore may be more responsive to differences in school quality (Heckman, Stixrud, and Urzua 2006).

In short, there is a clear need for evaluations of educational interventions that account for their effects on non-cognitive traits that influence subsequent educational success and labor market performance. Both existing earlier research and the results of this report established that the institutional arrangements of accountability, autonomy, and choice exert an overall positive influence on cognitive outcomes. But how do these institutions affect non-cognitive outcomes?

Economic theory suggests two competing hypotheses as to their potential impact. The first stems from the possibility that schools face a tradeoff between fostering the development of cognitive and non-cognitive skills. Schools that devote additional resources and attention to raising student achievement as measured by cognitive tests may pay less attention to students' development in other areas. If this is the

case, institutions that lead schools to emphasize cognitive achievement could lead to simultaneous declines in other outcomes (cf. Holmstrom and Milgrom 1991 for such a multitask principal-agent model).

The second hypothesis is that schools that are incentivized by their institutional environment to foster better cognitive outcomes will become more effective in ways that also improve non-cognitive skills. For example, parents who care about non-cognitive as well as cognitive skills will exert their choices considering both aspects, creating incentives to further both types of skills at the same time. Also, if non-cognitive and cognitive skills are complementary so that non-cognitive skills are instrumental in fostering cognitive skills, schools that are incentivized to achieve high cognitive skills will view the improvement of non-cognitive skills as one way of advancing cognitive outcomes.

6.2 Measures of Non-Cognitive Skills and Their Implication for the Empirical Model

Non-cognitive skills are difficult to define and to measure, which may help explain their neglect in analyses of earnings, schooling, and other lifetime outcomes. Many different aspects of personality are often lumped together under the general heading of non-cognitive skills. In this chapter, we analyze four outcome variables derived from the PISA 2003 background questionnaires that can proxy for skills in these areas: first, an index derived by PISA based on school principals' assessments of the enthusiasm and cooperation of their students ("Morale and Commitment"); second, an index derived by PISA based on school principals' assessments of absenteeism, disruption of classes, lack of respect, the use of alcohol and illegal drugs, and students' intimidating and bullying other students ("Non-disruptive Behaviour"); third, an index derived by PISA based on students' reports on noise and disorderly conduct during mathematics lessons ("Disciplinary Climate"); and fourth, students' self-reported tardiness. A more detailed description of the measures of non-cognitive skills is provided in Appendix A.5.

These indicators are of particular interest because, in each case, similar variables have been shown to be associated with students' long-term outcomes (Deke and Haimson 2006). At the same time, it is important to keep in mind several caveats. First, the variables used as proxies for non-cognitive skills certainly do not capture all aspects of non-cognitive skills that are important for individual economic and social success. Non-cognitive skills have many facets which are not ideally reflected by the variables measured in PISA 2003. For instance, social skills are hardly captured by any of the variables in the database, although they are widely believed to be quite important for long-term outcomes. The variables considered here are more closely related to discipline and work ethic, and thus do not capture the full range of non-cognitive skills that play a role in the labor market and in society as a whole.

Second, cross-cultural differences in dealing with conflicts, admitting and expressing grievances, and voicing dissent, for instance between Asian and European cultures, likely limit the cross-cultural comparability of the measured non-cognitive variables. Third, measurement error may be introduced through response biases, in particular the tendency of respondents to answer questions in a socially desirable manner. For instance, it is possible that students do not admit to coming late to school, and principals may not be honest in their responses about students' use of alcohol and illegal drugs or even their morale because they try to present themselves or their school in a more positive light.

Econometric techniques can only solve these problems up to a point. We evade problems of cross-cultural comparability by including country fixed effects in all estimated regressions of non-cognitive skills. This serves to remove all between-country variation. Yet it comes at the cost that between-country variation in the institutional features is also removed, so that our basic model cannot be estimated for non-cognitive dependent variables. In other words, using country fixed effects means that all institutional variables need to be measured at the school rather than country level. This is particularly problematic for the choice variables, for which the problem of selection bias is severe. For instance, students who report

that they attend a particular school because it is better than alternatives may differ from those who do not in many respects. The regression coefficient on this variable therefore needs to be interpreted with caution.

Despite these caveats, the PISA 2003 database clearly presents a welcome opportunity to address a question that has so far not been investigated: How do the institutional arrangements of accountability, autonomy, and choice affect non-cognitive outcomes?

6.3 Results

Table 9 presents the estimation results. All analyses include country fixed effects, and student weights are again computed such that each country contributes equally to the analyses. The indices of Morale and Commitment, Non-disruptive Behaviour, and Disciplinary Climate are each standardized to have a mean of 500 and a standard deviation of 100. Positive values on the indices indicate more positive non-cognitive outcomes in the sense that student morale and commitment to learning and disciplinary climate are better and that there are less student related factors hindering learning. In contrast, increasing values of tardiness reflect more negative non-cognitive outcomes.

Table 9: Accountability, autonomy, choice, and non-cognitive outcomes

Dependent variable:	Morale and Commitment	Non-disruptive Behaviour	Disciplinary Climate	Tardiness[a]
	(1)	(2)	(3)	(4)
Assessments used to make decisions about students' retention/promotion	0.008	3.065	3.457	0.874
	(4.510)	*(4.411)*	*(3.740)*	*(1.751)*
Assessments used to group students	4.310	-2.525	4.099	-3.598***
	(2.984)	*(3.024)*	*(2.713)*	*(1.310)*
Monitoring of teacher lessons by principal	12.852***	9.385**	-1.945	-1.362
	(3.518)	*(3.660)*	*(3.425)*	*(1.540)*
Monitoring of teacher lessons by external inspectors	7.240**	2.605	7.651**	-2.185
	(3.217)	*(3.507)*	*(3.040)*	*(1.386)*
Assessments used to compare school to district or national performance	12.155***	-1.539	-3.850	-1.607
	(3.238)	*(3.247)*	*(2.843)*	*(1.282)*
Standardized tests used at least monthly	11.166	-2.972	3.439	2.346
	(6.845)	*(8.195)*	*(5.088)*	*(2.590)*
Autonomy in formulating budget	-3.294	-6.972*	2.590	0.623
	(4.036)	*(3.627)*	*(3.453)*	*(1.576)*
Autonomy in staffing decisions	7.435**	-4.088	-1.766	-1.390
	(3.285)	*(3.712)*	*(3.072)*	*(1.432)*
Autonomy in hiring teachers	4.624	14.684***	-0.438	2.751
	(4.520)	*(4.735)*	*(4.311)*	*(1.914)*
Autonomy in establishing starting salaries	3.135	-0.964	-1.804	0.01342
	(3.976)	*(3.705)*	*(3.646)*	*(1.587)*
Autonomy in determining course content	-0.971	-5.429	0.211	0.903
	(3.268)	*(3.352)*	*(3.139)*	*(1.511)*
Private operation	1.653	15.277***	-2.886	-6.563***
	(5.049)	*(5.788)*	*(4.813)*	*(2.266)*
Government funding	-5.683	-11.476	9.428	6.717**
	(8.323)	*(9.242)*	*(6.974)*	*(3.183)*
Attending school because local	8.175	3.702	6.851	-2.259**
	(7.500)	*(8.129)*	*(6.971)*	*(1.135)*
Attending school because better	79.333***	70.582***	68.161***	-8.159***
	(9.312)	*(10.209)*	*(8.977)*	*(1.218)*
Urban	-8.037	-42.384***	-32.656***	37.253***
	(8.634)	*(8.488)*	*(7.784)*	*(2.379)*
Urban x Attending school because local	-15.740	3.575	8.487	-2.976
	(9.732)	*(9.905)*	*(9.078)*	*(1.897)*
Urban x Attending school because better	19.706	27.285***	31.797***	-7.704***
	(12.348)	*(13.833)*	*(11.065)*	*(1.928)*
Level of analysis	Schools	Schools	Schools	Students
Observations (students)	–	–	–	215,122
Observations (schools)	7,985	7,990	8,190	8,195
Countries	29	29	29	29
R^2	0.285	0.262	0.315	–

Sample: OECD countries. Columns 1-3: Least-squares regressions, equal country weights. Column 4: Ordered probit regression, equal country weights; see Table C.4 in Appendix C for the interpretation of significant coefficients. Controls include: country fixed effects, 15 student characteristics, 16 family background measures, 7 measures of school location and resources, imputation dummies, and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering at the school level in parentheses. Significance level: *** 1 percent, ** 5 percent, * 10 percent. [a] Reported coefficients multiplied by 100.

With few exceptions, the institutional features associated with higher cognitive achievement also tend to be associated with better non-cognitive outcomes whenever the estimated effects are statistically significant. Students in schools where teachers' lessons are monitored by principals, for example, exhibit higher levels of morale and commitment and better behaviour, as reported by their principal, than students in other schools. Similarly, students in schools where teachers' lessons are monitored by external inspectors show higher morale and commitment and report that students in their classroom are better behaved. Principals also report higher levels of morale and commitment among their students where assessments are used to compare the school's performance to other schools in the district or nation. There is a statistically significant association between the use of assessments to group students (a measure that has a negative impact on mathematics test scores) and tardiness: Students in schools with this accountability device report to be late less often.[17] It may be that grouping students by ability level generates higher levels of student engagement – a factor that could contribute to the practice's enduring popularity – but that it does so at the expense of their academic progress.

In schools with greater autonomy in hiring and staffing decisions (which were found to be positively associated with cognitive skills), school principals also report a higher level of student morale and commitment and less student behaviour hindering learning. There also is a tendency for schools with autonomy in formulating the school budget (which was found to be negatively associated with cognitive skills) to have a higher degree of disruptive behaviour that hinders students' learning. The measures of non-cognitive skills reported by students are not significantly associated with any autonomy variable.

As discussed above, particular caution is required when interpreting the coefficients on the variables measuring parental choice among schools because of the problem of selection bias. Students who report that they attend their school because it is known to be better than others in the area score substantially better on all four non-cognitive variables examined. They report a better disciplinary climate and claim that they are late for school less often. Principals in these schools also judge their students' morale and commitment more favorably and report fewer problems with disruptive student behaviour. All these differences are quite large, at more than two thirds of a standard deviation in the case of the three standardized outcomes. While it is important to keep in mind the role that self-selection may play in generating these results, the fact that the interaction terms between attending a better school and urban areas is significant and positive suggests that choice also plays a role.

In privately operated schools, there seem to be fewer factors of disruptive behaviour hindering students' learning, and students seem to be more disciplined in the sense that they are less tardy. There is also some evidence that a higher share of government funding is associated with more tardiness, a finding that is difficult to interpret.

The analysis of this chapter was guided by two competing hypotheses. The first hypothesis was that the institutional devices of accountability, autonomy, and choice foster cognitive student achievement, but at the expense of non-cognitive outcomes. The second hypothesis was that these institutions improve both students' cognitive and non-cognitive skills. Our results, though they come with many caveats, are much more consistent with the second hypothesis.

[17] Because of the ordinal nature of the non-cognitive variable "tardiness", weighted ordered probit regression is used. Table C.4 in Appendix C reports marginal effects of the ordered probit model which allow a more detailed interpretation.

7. CONCLUSION

All over the world, societies worry about the state of their school systems. Do they work efficiently to advance the cognitive and non-cognitive skills of students? In the endeavors to reform the school systems, three institutional measures have recently taken center stage: accountability, autonomy, and choice. The rationale of such market-oriented reforms is that school systems based on informed choice between autonomous schools improve student achievement by creating incentives for students, parents, teachers, schools, and administrators to provide the best learning environment for students. However, such reforms are not without criticism and opposition. Do they work? Do school systems based on choice among autonomous and accountable schools really perform better?

This report uses the cross-country variation in student achievement and in the three institutional features available in the PISA 2003 database to shed light on this question by performing cross-country student-level multiple regression analyses. The empirical facts provide a clear answer: Various forms of school accountability, autonomy, and choice policies combine to lift student achievement to substantially higher levels. Of course, there are many nuances in the detailed results presented above which paint a much richer picture of how specific aspects of these three institutional features affect student achievement. But as a general rule, students in school systems based on accountability, autonomy, and choice perform substantially better on cognitive skills in mathematics, science, and reading as tested in PISA 2003 than do students in school systems with less accountability, autonomy, and choice. Furthermore, the improved cognitive skills do not come at the cost of neglect for non-cognitive skills. Quite to the contrary, many aspects of accountability, autonomy, and choice are also associated with superior non-cognitive skills such as higher student morale and commitment, lower disruptive behaviour, better disciplinary climate, and less tardiness, as measured by the PISA 2003 background questionnaires.

Accountability measures aimed at students, teachers, and schools can complement each other to improve student outcomes. External exit exams and the use of assessments for decisions about student promotion and retention incentivize students to increase their achievement, while the use of assessments to group students reduces performance. Regular standardized testing is only beneficial where clear standards and goals are set by external exit exams. Student achievement increases also when teachers are held accountable because their principals and external inspectors monitor their lessons. Likewise, students perform better if their schools are held accountable because assessments are used to compare them to district or national performance.

On average, students perform better if schools have autonomy to decide on staffing and to hire their own teachers, while student achievement is lower when schools have autonomy in areas with large scope for opportunistic behaviour, such as formulating their own budget. But school autonomy in formulating the budget, in establishing teacher salaries, and in determining course content are all significantly more beneficial in systems where external exit exams introduce accountability. Autonomy in staffing decisions is an exception where the opposite seems to be the case.

Students perform substantially better in systems where private school operation creates choice and competition. At the same time, student achievement increases along with government funding of schools. A level playing field in terms of access to government funding for public and private schools proves particularly performance enhancing. The evidence is less clear on whether choice among public schools has any significant effect on student achievement across countries, although in urban areas where there are

more schools to choose from, student achievement is higher for students who are not restricted to attend the local school and who report that they attend their school because it is better than alternatives. Within countries, the superior performance of privately operated schools seems to hinge on the simultaneous existence of policies that introduce external accountability and on the autonomy that schools in the system have to respond to private competition.

In sum, the international evidence presented in this report shows that along several dimensions, accountability, autonomy, and choice interact to determine student achievement. This is particularly true for the dependence of autonomy effects on accountability, but also for interactions of choice with external accountability and with autonomy. It seems, therefore, that school accountability, autonomy, and choice are interrelated policies that can be mutually reinforcing.

The evidence presented in this report can help countries learn from one another. The cross-country analyses exploit the unique opportunity offered by the substantial institutional variation in accountability, autonomy, and choice that exists across countries, but usually not within individual countries. Thus, countries without experience with one or more of the institutional measures examined here can be informed by the experiences of other countries. The rich evidence also contains lessons about which particular forms of accountability, autonomy, or choice will be most valuable in particular contexts. Accountability policies can be aimed at students, teachers, or schools. Schools may be autonomous in decision-making areas such as budgeting, staffing, salaries, or course content, but not in others. And school choice encompasses many aspects of private school operation, the funding of public and private schools, and choice among public schools. No single country has experience in all of these areas, and no single country has yet established itself as possessing the one best school system which others would do well to emulate. In a rapidly changing and globalizing world, the need for educational improvement is universal. And international evidence can inform policy reforms worldwide in ways that national experiences never can.

APPENDIX A: DATABASE AND DESCRIPTIVE STATISTICS

This Appendix describes the PISA 2003 database and its measures of cognitive skills, how we used the PISA data to construct a student-level micro database for the estimation, details of the available measures of school accountability, autonomy, and choice, and an overview of the extensive background controls included in the analysis.

A.1 The PISA 2003 Database and Its Measures of Cognitive Skills

The 2003 round of the OECD Programme for International Student Assessment (PISA 2003) was conducted in 41 developed and emerging countries, 30 of which are OECD countries.[18] PISA 2003 assessed the mathematical, scientific, and reading literacy as well as the problem solving skills of the student population in each participating country. The term "literacy" signifies that not only the knowledge of the students in each of the three domains, for example based on national curricula, is assessed but also their ability to use the acquired knowledge to meet real-life challenges. As in the first PISA study conducted in 2000, the target population was the 15-year-old students in each country, regardless of the grade they currently attended. Thus, in most of the countries assessed, the target population comprises young people near the end of their compulsory schooling, independent of how many years of schooling are foreseen for 15-year-olds by the structure of the national school systems. Table A.2 reports the countries participating in the PISA 2003 study.

The PISA sampling procedure ensured that a representative sample of the target population was tested in each country. Most countries employed a two-stage sampling technique. The first stage drew a (usually stratified) random sample of schools in which 15-year-old students were enrolled. In most countries, the probability of the schools to be selected was proportional to their size as measured by the estimated numbers of 15-year-old students enrolled in the school. The second stage randomly sampled 35 of the 15-year-old students in each of these schools, with each 15-year-old student in a school having equal selection probability. In schools with less than 35 students in the targeted age group, all of these students were selected into the sample. Generally, a minimum of 150 schools had to be sampled (or all schools if there were less than 150 schools in a country) and a minimum of 4,500 students had to be assessed in each country. The final sample size varied considerably between the participating countries, ranging 3,350 students in 129 schools in Iceland and 29,983 students in 1,124 schools in Mexico (Luxembourg tested all 3,923 target-aged students in all its 29 applicable schools).

The performance tests were paper and pencil tests, lasting a total of two hours for each student. Test items included both multiple-choice items and open ended questions. The PISA tests were constructed to test a range of relevant skills and competencies that reflected how well young adults are prepared to analyze, reason, and communicate their ideas effectively. Each subject was tested using a broad sample of tasks with differing levels of difficulty to represent a coherent and comprehensive indicator of the continuum of students' abilities. The main focus of the PISA 2003 study was on mathematical literacy, with about 70 per cent of the testing time devoted to this domain. The test items were presented to the students in the form of test booklets that consisted of different clusters of test items. Each student was

[18] For detailed information on the PISA study and its database, see OECD (2004, 2005a, 2005b) and the PISA homepage at http://www.pisa.oecd.org.

given one of 13 different test booklets that varied in the composition of the test items representing the four tested domains. PISA used item response theory scaling and calculated five plausible values for proficiency in each of the tested domains for each participating student. The performance in each domain was mapped on a scale with an international mean of 500 and a standard deviation of 100 test-score points across the OECD countries.

A.2 Construction of a Student-Level Micro Database for the Estimation

PISA 2003 does not only provide achievement data for representative samples of students in the participating countries but also a rich array of background information on each student as well as on his or her school. In separate background questionnaires, students were asked to provide information on their personal characteristics and family backgrounds, and school principals provided information on their schools' resource endowments and institutional settings.

Combining the available data, we constructed a dataset containing 219,794 students in 29 OECD countries. France had to be dropped from the sample because no school-level background information was provided for any of the schools sampled in this country. We also constructed a second dataset which consisted of both OECD and non-OECD countries. This second dataset contains 265,878 students in 37 countries. Liechtenstein, Macao, and Serbia/Montenegro had to be discarded from the dataset because fundamental country-level variables were not available in an internationally comparable way.

The datasets combine students' test scores in mathematical literacy and the other testing domains with students' characteristics, family-background data, school-related variables of resource availability, and school-level measures of accountability, autonomy, and choice. For estimation purposes, a variety of qualitative variables were transformed into dummy variables. We imputed missing observations on the questionnaire items with advanced micro-econometric techniques (cf. Appendix B.3 for the imputation technique and how the imputations are controlled for in the actual estimations).

We combine the rich PISA data at the student and school level with additional country-level data. GDP per capita in 2003, measured in purchasing power parities (PPP), is provided by version 6.2 of the Penn World Tables (Heston, Summers, and Aten 2002). Cumulative educational expenditure per student between age 6 and 15 in 2002, measured in PPP are provided in OECD (2006b) and other versions of the OECD's Education at a Glance.[19] The number of years spent in separate school systems after the occurrence of the first selection in the education process is taken from OECD (2006b) for OECD countries and from the UNESCO World Database on Education for partner countries. The data on the existence of curriculum-based external exit exams is an updated version of the data used by Bishop (2006), Wößmann (2003b), and Fuchs and Wößmann (2007), which is collected from reviews of comparative-education studies and educational encyclopedia, interviews with representatives of the national education systems, government documents, and background papers.

Table A.1 reports international descriptive statistics for all the variables employed in this paper. It also includes information on the amount of original versus missing data for each variable. Table A.2 presents country means of selected key variables for each participating country.

[19] For the three countries with missing data in OECD (2006b) or other versions of the OECD's Education at a Glance, we use comparable data for these countries based on information from the World Development Indicators of the World Bank and data from both sources for countries where both are available to predict the missing data for the three countries by ordinary least squares.

A.3 Data on Accountability, Autonomy, and Choice

With the exception of the external exit exams, the measures of school accountability, autonomy, and choice are almost entirely taken from the school background questionnaires of the PISA 2003 study.

Measures of accountability include aspects of student testing, teacher monitoring, and school accountability. School principals report how often 15-year-old students are generally assessed in their school using standardized tests, with answer options ranging from never over 1 to 2 times a year and 3 to 5 times a year to monthly and more than once a month. As a measure of regular standardized testing in a school, we use an indicator of whether standardized tests are used at least monthly. School principals also report on whether assessments of 15-year-old students are used in their school for different purposes, including use to make decisions about students' retention or promotion; to group students for instructional purposes; and to compare the school to district or national performance.[20] In terms of teacher monitoring, principals report whether (a) principal or senior staff observations of lessons and whether (b) observation of classes by inspectors or other persons external to the school have been used during the last year to monitor the practice of mathematics teachers at their school.

Measures of school autonomy include responses of school principals to several items asking who has the main responsibility for different types of decisions regarding the management of the school. In particular, principals ticked whether any of the following was not a main responsibility of their school (as opposed to being a responsibility of either the school's governing board, the principal, department heads, or teachers): formulating the school budget; selecting teachers for hire; establishing teachers' starting salaries; and determining course content.[21] In addition, principals reported whether the school's governing board exerts a direct influence on decision making about staffing in their school (with other non-exclusive answer possibilities including such bodies as regional education authorities, parent groups, and teacher groups, among others). We use this as a more general measure of autonomy in staffing decisions in addition to the measure of autonomy in hiring teachers.

Measures of school choice include the availability of private schools and some proxies for the parental choice among public schools more generally. Principals reported whether their schools is a public or a private school, where a public school was defined as "a school managed directly or indirectly by a public education authority, government agency, or governing board appointed by government or elected by public franchise", while a private school was defined as "a school managed directly or indirectly by a non-government organization; e.g. a church, trade union, business, or other private institution." Principals also reported about what percentage of their schools' total funding for a typical school year comes from government sources, including departments, local, regional, state, and national governments (as opposed to student fees or school charges paid by parents; contributions by benefactors, donations, bequests, sponsorships, and parent fund raising; and other sources). Finally, principals reported how much consideration is given to residence in a particular area when students are admitted to their school, with answer options ranging from prerequisite over high priority to considered and not considered. We use an indicator for whether a particular residence was prerequisite or high priority for admission as a measure for lack of parental choice among schools. Similarly, the students were asked whether it were reasons why

---

[20] Principals also report on whether assessments are used to inform parents about their child's progress, but with 97 percent replying positively, there is hardly any international variation in this variable.

[21] There were also items on autonomy in firing teachers and in determining teachers' salary increases. However, these two are extremely collinear with autonomy in hiring teachers and in establishing teachers' starting salaries, respectively, with cross-country correlations as high as 0.963 and 0.971, respectively. Therefore, only one autonomy variable each was used, and the results should be interpreted as capturing autonomy in the joint decision-making areas of hiring/firing teachers and determining starting salaries as well as salary increases, respectively.

they attend the specific school (a) that this is the local school for students who live in this area, and (b) that this school is known to be a better school than others in the area. We use the former as an indicator of lack of parental choice among schools, and the latter as an indicator of exerted choice among schools.

A.4 Background Controls

Since PISA 2003 collected background information about the students, their families, and schools, it is possible to control for influencing factors at these levels. The 42 variables included as controls in the model are reported in Table C.1 (descriptive statistics are given in Table A.1). These include 15 measures of student characteristics, including student gender, student age, the age at which the student started primary education, a dummy indicating whether the student attended pre-primary education for more than one year,[22] two dummies for grade repetition, a set of dummies representing the grade that the student currently attends, two indicators for the immigrant status of the student,[23] and two indicators for the language spoken at home.[24]

The controls also include 16 measures of family background: the family structure as indicated by whether the student lived together with both parents, with only one parent,[25] or in a patchwork family, four indicators on the parents' working status,[26] three indicators of the highest occupational status of the parents,[27] five indicators of the number of books in the students' home,[28] and the PISA index of Economic, Social and Cultural Status (ESCS).[29]

---

22 We also tested including an indicator for attending pre-primary education for one year or less, but the coefficient estimate turned out to be not different from zero relative to no pre-primary attendance.

23 The immigrant status of the students was captured by the following categories: "native" students (those students born in the country of assessment or who had at least one parent born in the country); "first generation" students (those born in the country of assessment but whose parent(s) were born in another country); and "non-native" students (those students born outside the country of assessment and whose parents were also born in another country). In the analysis, "native" students served as the residual category.

24. The language spoken at home most of the time was captured by the following four categories: "test language"; "other official national languages"; "other national dialects or languages"; and "other languages". Only the latter two dummies were included in the analysis, with the first two serving as residual categories.

25 We also tested including living with the mother or the father separately, but the coefficient estimates turned out to be statistically non-distinguishable.

26 The four indicators of parents' working status are: both parents working full time; one parent working full time and the other half time; at least one parents working full time; at least one parent working half time. Other possible combinations of working part time or looking for a job act as the residual category.

27 The highest occupational status of both parents was scaled in four categories: blue collar high skilled; white collar low skilled; white collar high skilled, and blue collar low skilled, which serves as the residual category.

28 The categories of books at home are: 1-10, 11-25, 26-100, 101-200, 201-500, and more than 500 books, with the first category acting as the residual category.

29 The ESCS index is derived from the highest occupational status of parents, the highest educational level, and an estimate related to household possessions. We also tested additionally including indicators of parental education, but their effect seems to be fully captured by the included ESCS index.

The model includes 9 school-level measures of school location and resources: three indicators of the size of the community in which the school is located,[30] the average class size in mathematics, two indicators of the availability of instructional material, instruction time in mathematics, and the shares of teachers in the school who are fully certified and who have a tertiary degree in pedagogy. In addition, the model includes the country-level variables GDP per capita and expenditure per student, as described above.

A.5 Measures of Non-Cognitive Skills

While the main focus of the PISA 2003 study is on an assessment of students' cognitive skills in the domains of mathematics, science, reading, and problem solving, the PISA 2003 database also provides some measures of non-cognitive skills, which are derived from students' and school principals' reports in the school and student background questionnaires. In this report, four different measures of non-cognitive skills are used.

First, in the school background questionnaire, principals were asked whether their students enjoyed being in school, whether they worked with enthusiasm, whether they took pride in their school, whether they valued academic achievement and the education they could receive in this school, whether they were cooperative and respectful, and whether they did their best to learn as much as possible. PISA combined these variables into an index of "school principal's assessment of student morale and commitment" ("Morale and Commitment") using item response theory (IRT) scaling. Higher values on this index indicate a higher level of (perceived) student morale and commitment.

Second, principals assessed the extent to which students' learning in their school was hindered by student absenteeism, disruption of classes by students, class skipping, lack of respect, use of alcohol and illegal drugs, and students intimidating or bullying other students. From the principals' responses to these variables, PISA derived an index ("Non-disruptive Behaviour") using IRT scaling. Higher values on this index indicate that student learning is hindered to a lower degree.

Third, in the student background questionnaire, students were asked to assess the disciplinary climate during mathematics lessons. In particular, they reported whether students listened to what the teacher says, whether there was noise and disorder, whether the teacher had to wait for a long time for students to quieten down, whether students could work well, and whether students did not start working for a long time after the lesson begins. The index "disciplinary climate during mathematics lessons" ("Disciplinary Climate") was derived by PISA through IRT scaling, and positive values on this index indicate students' perceptions of a positive disciplinary climate.

To facilitate the comparison of the magnitude of estimated coefficients from regressions with cognitive test scores as dependent variables with those of non-cognitive dependent variables, all three indices were standardized to have a mean of 500 and a standard deviation of 100. For all indices, positive values indicate a positive assessment of non-cognitive skills. Observations with missing values on these indices were dropped from the analysis (5.6% of the observations for "Morale and Commitment", 6.0% for "Non-disruptive Behaviour", and 0.6% for "Disciplinary Climate").

The fourth measure of non-cognitive skills is tardiness. Students were asked to report how many times they arrived late for school in the last full two weeks they were in school. The possible answer categories and the international share of students reporting each category are "none" (64.0%), "1 or 2 times" (24.4%), "3 or 4 times" (6.5%), and "5 or more times" (5.1%). The 2.2% observations with missing answers were dropped from the analysis.

[30] The coefficient estimates on location of the school in a small town (3,000 to 15,000 people) and in a town (15,000 to 100,000 people) turned out to be statistically non-distinguishable, so we combined these two categories into one.

A.6 Tables of Descriptive Statistics

Table A.1: Descriptive statistics of the international dataset

	Incl. imputations		Only original data		
	Mean	Std. Dev.	Mean	Std. Dev.	Imputations
TEST SCORES					
Math	499.626	100.365	499.626	100.365	0.0%
Science	499.239	105.252	499.239	105.252	0.0%
Reading	494.137	100.457	494.137	100.457	0.0%
ACCOUNTABILITY					
External exit exams					
In mathematics	0.650		0.650		0.0%
In science	0.557		0.557		0.0%
Assessments used to					
Decide about students' retention/promotion	0.779		0.782		6.5%
Group students	0.474		0.473		3.5%
Compare school to district/national performance	0.475		0.475		3.4%
Monitoring of teacher lessons					
By principal	0.607		0.607		3.6%
By external inspectors	0.245		0.245		3.9%
Standardized tests used at least monthly	0.056		0.056		4.5%
AUTONOMY					
Autonomy in formulating budget	0.789		0.788		3.3%
Autonomy in staffing decisions	0.406		0.405		3.8%
Autonomy in hiring teachers	0.671		0.670		2.9%
Autonomy in establishing starting salaries	0.287		0.287		3.3%
Autonomy in determining course content	0.748		0.748		3.2%
CHOICE					
Private operation (PISA)	0.178		0.174		5.6%
Private operation (EAG)	0.143		0.143		0.0%
Government funding	0.862		0.861		8.8%
Diff. in gov. funding b/w public + private schools	0.355		0.355		0.0%
Attending school because local	0.478		0.474		4.7%
Attending school because better	0.273		0.272		4.7%
STUDENT CHARACTERISTICS					
Female	0.496		0.496		0.3%
Age (years)	15.780	0.290	15.780	0.291	0.3%
Preprimary education (more than 1 year)	0.679		0.680		2.4%
School starting age	6.021	0.826	6.032	0.863	11.7%
Grade repetition in primary school	0.076		0.074		13.3%
Grade repetition in secondary school	0.068		0.062		16.0%
Grade					
7th grade	0.006		0.006		0.5%
8th grade	0.047		0.047		0.5%
9th grade	0.359		0.359		0.5%
10th grade	0.527		0.526		0.5%
11th grade	0.061		0.061		0.5%
12th grade	0.001		0.001		0.5%
Immigration background					
Native student	0.916		0.916		2.7%
First generation students	0.037		0.037		2.7%
Non-native students	0.047		0.047		2.7%

(continued on next page)

Table A.1 (continued)

	Incl. imputations		Only original data		
	Mean	Std. Dev.	Mean	Std. Dev.	Imputations
Language spoken at home					
Test language or other official national language	0.922		0.921		4.3%
Other national dialect or language	0.033		0.032		4.3%
None of above	0.047		0.046		4.3%
FAMILY BACKGROUND					
Living with					
No parent	0.017		0.018		6.1%
Single mother or father	0.189		0.201		6.1%
Patchwork family	0.060		0.064		6.1%
Both parents	0.733		0.717		6.1%
Parents' working status					
Both full-time	0.391		0.391		2.0%
One full-time, one half-time	0.179		0.179		2.0%
At least one full time	0.293		0.293		2.0%
At least one half time	0.065		0.065		2.0%
Other (less than one half but not both missing)	0.071		0.071		2.0%
Parents' job					
Blue collar low skilled	0.095		0.095		4.2%
Blue collar high skilled	0.139		0.139		4.2%
White collar low skilled	0.234		0.234		4.2%
White collar high skilled	0.532		0.533		4.2%
Books at home					
1-10 books	0.092		0.093		2.9%
11-25 books	0.141		0.142		2.9%
26-100 books	0.314		0.310		2.9%
101-200 books	0.203		0.198		2.9%
201-500 books	0.156		0.159		2.9%
More than 500 books	0.095		0.098		2.9%
Index of socio-economic & cultural status (ESCS)	0.000	1.000	-0.001	1.007	1.8%
SCHOOL LOCATION AND RESOURCES					
School's community location					
Village or rural area (<3,000)	0.108		0.108		2.8%
Town (3,000-100,000)	0.568		0.568		2.8%
City (100,000-1,000,000)	0.213		0.213		2.8%
Large city with > 1 million people	0.112		0.112		2.8%
Class size (mathematics)	23.222	7.352	23.206	7.621	7.8%
Shortage of instructional materials					
Not at all	0.381		0.380		3.2%
Strongly	0.069		0.070		3.2%
Instruction time (mathematics, minutes per week)	197.801	89.921	197.874	93.651	7.9%
Teacher education (share at school)					
Fully certified teachers	0.907		0.908		19.0%
Tertiary degree in pedagogy	0.654		0.668		34.0%
GDP per capita (1,000 $)	23.009	8.926	23.009	8.926	0.0%
Educational expenditure per student (1,000 $)	56.947	25.507	56.947	25.507	0.0%

Sample: OECD countries. Number of observations in sample incl. imputations: 219,794 students. Mean: International mean (weighted by sampling probabilities). – Std. Dev.: International standard deviation (only for continuous variables). Imputations: Percentage of students with missing and thus imputed data, weighted by sampling probabilities.

Table A.2: OECD country means of test scores, accountability, autonomy, and choice (continued on next page)

	TEST SCORES			ESCS	ACCOUNTABILITY							
				Socio-	External exit exams		Assessment for			Monitoring of teacher lessons		Standardized
	Math	Science	Reading	economic status	Math	Science	promotion	grouping	comparison	by principal	by external inspectors	tests (monthly)
Australia	524.08	525.38	525.67	0.23	0.81	0.81	0.62	0.78	0.55	0.63	0.08	0.02
Austria	505.10	490.98	490.91	0.05	0.00	0.00	0.93	0.32	0.12	0.78	0.37	0.01
Belgium	529.09	508.20	506.99	0.14	0.00	0.00	0.99	0.20	0.10	0.58	0.48	0.04
Canada	532.64	518.00	527.65	0.44	0.51	0.51	0.95	0.72	0.70	0.87	0.10	0.02
Czech Republic	516.06	522.18	488.04	0.15	1.00	1.00	0.92	0.35	0.50	0.99	0.31	0.02
Denmark	513.74	474.41	491.21	0.20	1.00	1.00	0.04	0.14	0.06	0.63	0.11	0.02
Finland	544.17	547.53	542.90	0.24	1.00	1.00	0.95	0.17	0.56	0.34	0.04	0.00
Germany	503.08	502.62	491.70	0.15	0.44	0.44	0.96	0.36	0.21	0.69	0.26	0.02
Greece	444.55	480.66	471.58	-0.16	0.00	0.00	0.99	0.11	0.12	0.07	0.16	0.19
Hungary	490.34	504.02	481.87	-0.07	1.00	1.00	0.95	0.35	0.86	0.96	0.26	0.02
Iceland	514.71	494.50	491.73	0.69	1.00	0.00	0.15	0.56	0.84	0.47	0.02	0.00
Ireland	503.48	506.20	515.82	-0.08	1.00	1.00	0.44	0.78	0.17	0.07	0.05	0.03
Italy	465.77	486.30	474.94	-0.11	0.00	0.00	0.84	0.51	0.33	0.16	0.01	0.17
Japan	533.64	548.14	499.04	-0.08	1.00	1.00	0.90	0.45	0.18	0.56	0.15	0.03
Korea	541.63	538.46	534.71	-0.10	1.00	1.00	0.25	0.63	0.62	0.90	0.62	0.04
Luxembourg	493.28	482.81	478.58	0.19	1.00	1.00	1.00	0.30	0.22	0.42	0.07	0.02
Mexico	384.86	403.53	399.53	-1.14	0.00	0.00	0.93	0.59	0.55	0.72	0.36	0.17
Netherlands	538.06	524.91	513.96	0.08	1.00	1.00	0.97	0.89	0.63	0.58	0.33	0.13
New Zealand	524.08	521.81	521.99	0.21	1.00	1.00	0.78	0.74	0.87	0.94	0.52	0.22
Norway	495.35	484.93	499.68	0.61	1.00	0.30	–	0.38	0.64	0.26	0.07	0.00
Poland	490.10	497.86	496.48	-0.21	1.00	1.00	0.84	0.33	0.71	0.97	0.14	0.04
Portugal	466.14	468.46	477.76	-0.64	0.00	0.00	0.97	0.26	0.33	0.05	0.10	0.00
Slovak Republic	498.63	494.67	469.24	-0.09	1.00	1.00	0.97	0.55	0.46	0.98	0.25	0.03
Spain	485.57	487.48	481.68	-0.30	0.00	0.00	1.00	0.48	0.18	0.15	0.14	0.13
Sweden	509.59	506.33	514.32	0.25	1.00	0.00	0.39	0.45	0.73	0.58	0.16	0.05
Switzerland	526.09	513.11	498.61	-0.06	0.00	0.00	0.95	0.28	0.19	0.42	0.59	0.02
Turkey	423.80	434.64	441.68	-0.99	1.00	1.00	0.71	0.51	0.59	0.89	0.40	0.14
United Kingdom	508.02	518.20	506.81	0.11	1.00	1.00	0.68	0.94	0.89	0.91	0.61	0.01
United States	483.49	491.59	494.87	0.29	0.09	0.09	0.76	0.66	0.91	1.00	0.37	0.02

Table A.2 (continued)

	AUTONOMY					CHOICE					
	In formulating budget	In staffing decisions	In hiring teachers	In establishing starting salaries	In determining course content	Private operation (PISA)	Private operation (EAG)	Government funding	Public-private diff. gov. fund.	Attending school because local	Attending school because better
Australia	0.89	0.21	0.62	0.20	0.79	0.38	0.35	0.71	–	0.55	0.52
Austria	0.14	0.03	0.22	0.00	0.61	0.08	0.08	–	–	0.10	0.27
Belgium	0.81	0.62	0.83	0.00	0.55	0.69	0.57	0.89	0.11	0.27	0.32
Canada	0.75	0.59	0.81	0.32	0.45	0.07	–	0.92	0.40	0.72	0.36
Czech Republic	0.83	0.05	0.98	0.69	0.75	0.07	0.02	0.95	0.33	0.31	0.31
Denmark	0.91	0.74	0.97	0.21	0.76	0.22	0.23	0.93	0.23	0.64	0.17
Finland	0.80	0.88	0.70	0.10	0.92	0.07	0.04	1.00	0.02	0.81	0.10
Germany	0.09	0.28	0.18	0.02	0.48	0.08	0.07	0.96	0.20	0.42	0.26
Greece	1.00	0.09	0.04	0.00	0.00	0.04	0.05	0.88	0.90	0.50	0.28
Hungary	0.87	0.79	1.00	0.38	0.80	0.11	0.07	0.91	0.15	0.14	0.35
Iceland	0.94	0.36	1.00	0.19	0.86	0.00	0.01	1.00	0.55	0.84	0.12
Ireland	0.77	0.52	0.86	0.04	0.38	0.61	0.00	0.93	0.08	0.61	0.48
Italy	0.26	0.16	0.07	0.02	0.84	0.05	0.03	0.72	0.61	0.07	0.17
Japan	0.47	0.22	0.29	0.27	1.00	0.27	0.06	0.74	0.57	0.20	0.18
Korea	0.92	0.26	0.33	0.15	0.99	0.56	0.20	0.52	-0.08	0.43	0.28
Luxembourg	0.05	0.51	0.00	0.05	0.05	0.14	0.20	0.97	0.10	0.32	0.27
Mexico	0.84	0.34	0.75	0.47	0.70	0.16	0.13	0.39	0.45	0.15	0.30
Netherlands	1.00	0.71	1.00	0.88	0.97	0.77	0.76	0.96	0.00	0.29	0.20
New Zealand	0.99	0.73	1.00	0.19	0.94	0.05	0.16	0.78	0.66	0.58	0.46
Norway	0.73	0.10	0.64	0.01	0.48	0.01	0.02	1.00	0.11	0.93	0.06
Poland	0.30	0.02	1.00	0.21	1.00	0.01	0.02	0.96	0.61	0.79	0.18
Portugal	0.83	0.28	0.08	0.01	0.36	0.06	0.12	0.84	0.21	0.51	0.24
Slovak Republic	0.84	0.23	1.00	0.60	0.65	0.12	0.05	0.93	-0.01	0.28	0.28
Spain	0.86	0.18	0.36	0.06	0.65	0.38	0.32	0.86	0.29	0.48	0.28
Sweden	0.88	0.11	1.00	0.71	0.92	0.04	0.06	1.00	0.01	0.74	0.12
Switzerland	0.64	0.81	0.93	0.13	0.39	0.06	0.07	0.95	0.77	0.65	0.11
Turkey	0.51	0.35	0.07	0.05	0.36	0.03	–	0.55	0.56	0.35	0.48
United Kingdom	0.90	0.88	0.99	0.80	0.94	0.06	0.06	0.93	0.86	0.61	0.51
United States	0.85	0.77	0.98	0.69	0.81	0.06	0.09	0.88	0.91	–	–

Country means, based on non-imputed data for each variable, weighted by sampling probabilities. ESCS = PISA index of Economic, Social and Cultural Status. Institutional measures are shares within each country (in percent). – = not available.

Table A.3: Non-OECD country means of test scores, accountability, autonomy, and choice

	TEST SCORES			ESCS	ACCOUNTABILITY							
					External exit exams		Assessment for			Monitoring of teacher lessons		Standardized tests (monthly)
	Math	Science	Reading	Socio-economic status	Math	Science	promotion	grouping	comparison	by principal	by external inspectors	
Brazil	355.52	391.76	403.51	-0.96	0.00	0.00	0.83	0.45	0.38	0.50	0.12	0.22
Hong Kong (China)	549.43	539.11	509.21	-0.76	1.00	1.00	0.96	0.63	0.23	0.92	0.26	–
Indonesia	360.09	394.56	380.62	-1.27	1.00	1.00	0.84	0.46	0.51	0.92	0.75	0.04
Latvia	483.03	489.12	490.88	0.11	1.00	1.00	0.94	0.40	0.80	1.00	0.41	0.25
Russian Federation	469.11	489.73	442.30	-0.10	1.00	1.00	0.97	0.56	0.70	1.00	0.74	0.08
Thailand	417.14	428.52	419.91	-1.19	1.00	1.00	0.72	0.77	0.59	0.87	0.49	0.00
Tunisia	358.92	384.77	374.44	-1.35	1.00	1.00	0.84	0.44	0.73	0.74	0.80	0.19
Uruguay	421.85	438.25	434.98	-0.35	0.00	0.00	0.91	0.29	0.18	0.92	0.52	0.02

	AUTONOMY					CHOICE					
	In formulating budget	In staffing decisions	In hiring teachers	In establishing starting salaries	In determining course content	Private operation (PISA)	(EAG)	Government funding	Public-private diff. gov. fund.	Attending school because local	better
Brazil	0.59	0.18	0.39	0.17	0.88	0.15	0.09	0.79	0.90	0.39	0.41
Hong Kong (China)	0.98	0.72	0.91	0.28	0.98	0.91	1.00	0.90	0.05	0.51	0.39
Indonesia	0.97	0.14	0.50	0.52	0.98	0.46	0.36	0.33	0.36	0.49	0.52
Latvia	0.79	0.67	0.99	0.37	0.56	0.01	0.00	0.96	0.82	0.48	0.41
Russian Federation	0.48	0.14	0.99	0.49	0.83	0.00	0.00	0.92	0.92	0.51	0.33
Thailand	0.80	0.50	0.26	0.22	0.99	0.12	0.06	0.83	0.45	0.62	0.54
Tunisia	0.33	0.02	0.01	0.29	0.11	–	0.01	0.71	–	0.49	0.38
Uruguay	0.28	0.15	0.20	0.20	0.26	0.14	0.12	0.79	0.92	0.56	0.23

Country means, based on non-imputed data for each variable, weighted by sampling probabilities. ESCS = PISA index of Economic, Social and Cultural Status.
Institutional measures are shares within each country (in percent). – = not available.

APPENDIX B: ECONOMETRIC MODELING

The basic setup of the empirical model, estimating international education production functions by cross-country student-level multiple regressions, is described in Section 2.2 in the main text. This Appendix discusses details of the econometric model, including the potential for bias when using cross-country data in cross-sectional analyses, econometric complications resulting from the hierarchical data structure such as the multi-level structure of the error term and the use of sampling weights, and model implications of data imputation.

B.1 Cross-Country Data and Potential Bias

The econometric estimation of the PISA dataset is restricted by its cross-sectional nature, which does not allow for panel or value-added estimations (cf., e.g., Hanushek 2002; Todd and Wolpin 2003). Because of unobserved student abilities, cross-sectional analyses can give rise to omitted variable bias when the variables of interest are correlated with the unobserved characteristics. In this report, we hope to minimize such biases due to unobserved student heterogeneity by including a huge set of observed abilities, characteristics, and institutions which reduce potential biases. Estimates based on cross-sectional data will be unbiased under the conditions that the explanatory variables of interest are unrelated to features that still remain unobserved, that they are exogenous to the dependent variable, and that they and their impact on the dependent variable do not vary over time. We view the variables of student characteristics, family background, and school location and resources included in our model as control variables which do not necessarily lend themselves to causal interpretation.

Many of the institutional features of an education system may be reasonably assumed to be exogenous to individual students' performance. The cross-country nature of the data allows the systematic utilization of country differences in institutional settings of the educational systems, which would be neglected in within-country specifications. At the country level, explanatory variables are included to control for country differences with respect to educational expenditure and the development stage of a country. However, a caveat applies here in that a country's institutions may be related to unobserved, e.g. cultural, factors which in turn may be related to student performance. To the extent that this may be an important issue, caution should prevail in drawing causal inferences and policy conclusions from the presented results.

In terms of time variability, changes in institutions generally occur only gradually and evolutionary rather than radically, particularly in democratic societies. Consequently, the institutional structures of education systems are highly time-invariant and thus most likely constant, or at least rather similar, during a student's life in secondary school. We therefore assume that the educational institutions observed at one point in time persist unchanged during the students' secondary-school life and thus contribute to students' achievement levels, and not only to the change from one grade to the next. A level-estimation approach thus seems well-suited for determining the total association between institutions and student achievements. Still, institutional structures may differ between primary and secondary school, so that issues of omitted prior inputs in a students' life may still bias estimated institutional effects, generally in an attenuating way.

B.2 Micro-Econometric Issues of Hierarchically Structured Data: Multi-Level Error Components and Sampling Weights

The complex survey structure and design of the PISA 2003 study requires a non-trivial structure of the error term ε_{isc} of the estimation equation (see equation (1b) in the main text). Since PISA employed a

two stage sampling design, where in the first stage schools and in the second stage a sample of students was drawn from these schools, the primary sampling unit (PSU) in PISA is the school. As shown by Moulton (1986), the hierarchical structure of the data requires the addition of higher-level error components to avoid spurious results. Therefore, the error term ε in all the econometric equations estimated in this report has a country-level and a school-level element in addition to the individual student element:

$$\varepsilon_{isc} = \eta_c + \nu_s + \upsilon_i \qquad \text{(A1)}$$

where η is a country-specific error component, ν is a school-specific error component, and υ is a student-specific error component.

Clustering-robust linear regression (CRLR) is used to estimate standard errors that recognize this clustering of the survey design by allowing any given amount of correlation within PSUs in the error variance-covariance matrices (cf. Deaton 1997). The CRLR method relaxes the classical assumption of independence across individual observations and requires only that the observations be independent across the PSUs, i.e. across schools.

This assumption results in a CRLR approach which employs a covariance matrix of the following form:

$$V = \begin{pmatrix} \Sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \Sigma_l \end{pmatrix} \qquad \text{(A2)}$$

with $\sum_i$ (i=1,…,l) as the covariance matrices of the least square regression within each school (PSU). Assuming that PSUs are independent from one another leads to the block diagonal matrix V with PSUs as diagonal elements and results in consistent and efficient coefficient estimates (cf. White 1984).

In addition, the PISA 2003 study uses a stratified sampling design in each country which demands the use of sampling weights to obtain consistent student population estimates (allowing for different sampling probabilities). This is a direct consequence of the fact that PISA over-samples some sub-groups of the student population and thus students have different sampling probabilities for different strata with respect to student or family characteristics.

By using a weighted least squares (WLS) regression approach with students' sampling probabilities as weights, the estimation produces coefficient estimates which are equal to the estimates for a complete census enumeration of the whole student population in a country (DuMouchel and Duncan 1983; Wooldridge 2001). To avoid that the coefficient estimates are driven by the student population size of a country, the sampling weight is normalized in a way that all countries contribute equally to the coefficient estimates of the international education production function.

B.3 Data Imputation and Its Implications for the Estimation Model

Like in any survey dataset, there are missing data in the PISA 2003 dataset. Although this problem is minor for almost any single variable as can be seen from Table A.1, it becomes more problematic when

estimating international educational productions. Given the large set of explanatory variables considered and given that each variable has missing values for some students, dropping all student observations that have a missing value on at least one variable would mean a severe reduction in sample size. Data on teacher education are not available for up to a third of the students, and data on school starting age and grade repetition are missing for 11.7 percent to 16.0 percent. While the percentage of missing values for the other variables individually ranges from 0.0 percent to 8.8 percent (cf. Table A.1), the percentage of students with a missing value on least one variable of the baseline model is 63.4%. That is, the sample size in the baseline model would be reduced to 80,338 students in 24 countries.

Apart from the general reduction in sample size which would reduce the statistical power of the estimation, dropping all students with a missing value on at least one variable would delete information available on other explanatory variables for these students and introduce bias if values are not missing at random. Thus, data imputation is the only viable way of performing the broad-based analyses of this report.

We impute missing values using a conditional mean imputation method (cf. Little and Rubin 1987), which predicts the conditional mean for each missing observation on the explanatory variables using non-missing values of the specific variables and a set of explanatory variables observed for all students. Specifically, in order to obtain a complete dataset for all students for whom performance data are available, we imputed missing values of explanatory variables using a set of "fundamental" explanatory variables F that were available for all students. These fundamental variables F include gender, age, five grade dummies, four dummies on the students' family structure, five dummies for the number of books at home, GDP per capita as a measure of the country's level of economic development, and the country's educational expenditure per student.[31]

For each student i with missing data on a specific variable M, the set of "fundamental" explanatory variables F with data available for all students was used to impute the missing data in the following way. Let S denote the set of students j with available data for M. Using the students in S, the variable M was regressed on F:

$$M_{j \in S} = F_{j \in S}\phi + \varepsilon_{j \in S} \tag{A3}$$

Then, the coefficients ϕ from these regressions and the data on F_i were used to impute the value of M_i for the students with missing data:

$$\widetilde{M}_{i \notin S} = F_{i \notin S}\phi \tag{A4}$$

The imputation method for implied variables was WLS estimation for continuous variables, ordered probit estimation for ordinal variables, and probit estimation for dichotomous variables. For continuous variables, predicted values were then filled in for missing data. For ordinal and dichotomous variables, in each category the respective predicted probability was filled in for missing data. We perform the imputation once for the sample of OECD countries and once for the extended sample that includes non-OECD countries.

[31] The small amount of missing data on the variables in F was imputed by the use of median imputation on the lowest available level (school or country).

Generally, data imputation introduces measurement error in the explanatory variables, which should make it more difficult to observe statistically significant effects.[32] However, if values are not missing conditionally at random, estimates could still be biased. For example, if among observationally similar students the probability of a missing value for a variable depends on an unobserved student characteristic that also influences achievement, imputation would predict the same value of the variable for students with a missing value that was observed for the other students, which would result in biased coefficient estimates.

To account for this possibility of non-randomly missing observations and to make sure that the results are not driven by imputed data, we include a vector of imputation dummy variables as controls in the estimation. This vector contains one dummy for each variable of the model that takes the value of 1 for observations with missing and thus imputed data and 0 for observations with original data. The vector allows the observations with missing data on each variable to have their own intercepts. We additionally include interaction terms between each variable and its imputation dummy, which allows observations with missing data to also have their own slopes for the respective variable. These imputation controls make the results robust against possible bias arising from imputation errors in the variables. Thus, the models actually estimated in this report have the following structure:

$$\begin{aligned} T_{isc} &= B_{isc}\alpha + R_{sc}\beta + I_{sc}\gamma \\ &\quad + D^{B}_{isc}\mu_1 + \left(D^{B}_{isc}B_{isc}\right)\mu_2 + D^{R}_{sc}\mu_3 + \left(D^{R}_{sc}R_{sc}\right)\mu_4 + D^{I}_{sc}\mu_5 + \left(D^{I}_{sc}I_{sc}\right)\mu_6 + \varepsilon_{isc} \end{aligned} \tag{A5}$$

which adds the vectors of imputation dummies *D* and their interactions with the variables to equation (1b).

[32] In an analysis of the PISA 2000 data, Fuchs and Wößmann (2007) employ an adjustment mechanism for standard errors suggested by Schafer and Schenker (2000) that accounts for the degree of variability and uncertainty in the imputation process as well as for the share of missing data and find that all qualitative results are highly robust to the alternatively computed standard errors.

APPENDIX C: ADDITIONAL TABLES

Table C.1: Full results of the basic model

Subject:	Mathematics		Science	
Country sample:	OECD	Extended	OECD	Extended
	(1)	(2)	(3)	(4)
INSTITUTIONS				
External exit exams	13.724*	11.155*	15.745**	13.824**
	(7.496)	*(6.192)*	*(6.992)*	*(5.205)*
Autonomy in formulating budget	-25.056**	-28.596**	-17.723	-17.655*
	(10.661)	*(10.728)*	*(11.515)*	*(10.377)*
Autonomy in staffing decisions	29.310*	34.974**	21.216	23.177*
	(14.685)	*(13.710)*	*(14.733)*	*(13.051)*
Private operation	61.563***	61.405***	38.985***	42.757***
	(10.419)	*(10.317)*	*(8.517)*	*(8.747)*
Government funding	75.437***	80.114***	58.538**	54.644***
	(20.901)	*(17.352)*	*(21.958)*	*(16.757)*
STUDENT CHARACTERISTICS				
Female	-17.524***	-16.399***	-12.066***	-10.084***
	(0.644)	*(0.575)*	*(0.801)*	*(0.709)*
Age (years)	19.076***	15.961***	18.252***	15.786***
	(1.082)	*(1.026)*	*(1.438)*	*(1.317)*
Preprimary education (more than 1 year)	5.760***	8.251***	2.816***	3.840***
	(0.700)	*(0.627)*	*(0.903)*	*(0.786)*
School starting age	-2.218***	-0.469	-3.325***	-1.208**
	(0.517)	*(0.469)*	*(0.635)*	*(0.561)*
Grade repetition in primary school	-36.216***	-31.896***	-30.594***	-28.507***
	(1.438)	*(1.199)*	*(2.175)*	*(1.753)*
Grade repetition in secondary school	-34.412***	-32.037***	-33.262***	-32.509***
	(1.617)	*(1.393)*	*(2.193)*	*(1.909)*
Grade				
7th grade	-51.695***	-59.770***	-41.481***	-53.120***
	(4.081)	*(3.002)*	*(5.813)*	*(3.652)*
8th grade	-30.897***	-34.999***	-29.844***	-31.090***
	(2.214)	*(1.892)*	*(2.912)*	*(2.376)*
9th grade	-14.089***	-15.304***	-12.880***	-12.500***
	(1.249)	*(1.033)*	*(1.433)*	*(1.189)*
11th grade	-11.172***	-9.019***	-1.972	-1.324
	(2.034)	*(1.971)*	*(2.177)*	*(2.091)*
12th grade	0.668	0.222	4.107	5.883
	(4.752)	*(4.675)*	*(5.824)*	*(5.695)*
Immigration background				
First generation students	-7.975***	-3.022**	-8.947***	-5.882***
	(1.540)	*(1.390)*	*(2.141)*	*(1.807)*
Non-native students	-8.373***	-2.037	-12.164***	-7.014***
	(1.660)	*(1.592)*	*(2.233)*	*(2.032)*
Language spoken at home				
Other national dialect or language	-20.162***	-25.815***	-26.014***	-30.647***
	(2.887)	*(2.698)*	*(3.281)*	*(3.182)*
Foreign language	-7.084***	-14.195***	-19.060***	-23.633***
	(1.699)	*(1.740)*	*(2.511)*	*(2.366)*

(continued on next page)

Table C.1 (continued)

Subject:	Mathematics		Science	
Country sample:	OECD	Extended	OECD	Extended
	(1)	(2)	(3)	(4)
FAMILY BACKGROUND				
Living with				
Single mother or father	20.253***	16.160***	18.769***	15.498***
	(1.839)	*(1.491)*	*(2.717)*	*(2.147)*
Patchwork family	23.096***	17.944***	22.569***	17.139***
	(2.030)	*(1.695)*	*(2.991)*	*(2.420)*
Both parents	28.221***	22.796***	25.112***	21.404***
	(1.820)	*(1.470)*	*(2.719)*	*(2.119)*
Parents' working status				
Both full-time	-2.072	-1.583	-2.460	-3.754**
	(1.328)	*(1.081)*	*(1.879)*	*(1.499)*
One full-time, one half-time	7.118***	6.760***	5.771***	5.223***
	(1.063)	*(0.883)*	*(1.465)*	*(1.213)*
At least one full time	14.340***	12.172***	14.013***	11.298***
	(1.172)	*(1.016)*	*(1.633)*	*(1.391)*
At least one half time	9.219***	8.429***	7.030***	6.379***
	(1.132)	*(0.962)*	*(1.525)*	*(1.286)*
Parents' job				
Blue collar high skilled	0.579	0.085	2.951**	1.162
	(0.984)	*(0.861)*	*(1.403)*	*(1.157)*
White collar low skilled	3.136***	2.635***	4.582***	3.065***
	(0.939)	*(0.877)*	*(1.370)*	*(1.184)*
White collar high skilled	9.103***	9.007***	10.565***	9.009***
	(1.001)	*(0.923)*	*(1.477)*	*(1.272)*
Books at home				
11-25 books	5.674***	4.048***	7.152***	4.772***
	(0.980)	*(0.810)*	*(1.441)*	*(1.124)*
26-100 books	23.995***	23.114***	26.251***	23.588***
	(1.019)	*(0.865)*	*(1.412)*	*(1.155)*
101-200 books	34.151***	34.900***	37.094***	36.677***
	(1.125)	*(0.989)*	*(1.542)*	*(1.307)*
201-500 books	51.471***	53.029***	57.250***	55.797***
	(1.233)	*(1.091)*	*(1.669)*	*(1.444)*
More than 500 books	52.737***	53.272***	58.324***	56.724***
	(1.408)	*(1.252)*	*(1.912)*	*(1.669)*
ESCS	18.421***	16.682***	19.267***	17.625***
	(0.532)	*(0.451)*	*(0.651)*	*(0.550)*
GDP per capita (1,000 $)	-1.951*	-0.738	-1.427	-0.039
	(1.016)	*(0.941)*	*(0.901)*	*(0.900)*
OECD member		37.219***		26.914***
		(7.411)		*(6.972)*

(continued on next page)

Table C.1 (continued)

Subject:	Mathematics		Science	
Country sample:	OECD	Extended	OECD	Extended
	(1)	(2)	(3)	(4)
SCHOOL LOCATION AND RESOURCES				
School's community location				
Town (3,000-100,000)	2.143	0.497	3.354*	1.387
	(1.546)	*(1.400)*	*(1.897)*	*(1.629)*
City (100,000-1,000,000)	9.482***	8.112***	10.209***	8.563***
	(1.917)	*(1.734)*	*(2.294)*	*(1.978)*
Large city with > 1 million people	8.680***	10.016***	9.523***	8.242***
	(2.412)	*(2.153)*	*(2.656)*	*(2.305)*
Educational expenditure per student (1,000 $)	1.030**	0.565	0.787*	0.323
	(0.407)	*(0.349)*	*(0.404)*	*(0.355)*
Class size (mathematics)	1.562***	1.156***	1.660***	1.291***
	(0.068)	*(0.057)*	*(0.078)*	*(0.063)*
Shortage of instructional materials				
Not at all	-9.993***	-8.265***	-10.242***	-7.463***
	(2.587)	*(1.882)*	*(2.548)*	*(1.840)*
Strongly	6.914***	6.629***	8.859***	7.548***
	(1.312)	*(1.222)*	*(1.432)*	*(1.311)*
Instruction time (minutes per week)	0.036***	0.038***	0.018***	0.026***
	(0.005)	*(0.004)*	*(0.006)*	*(0.005)*
Teacher education (share at school)				
Fully certified teachers	8.665**	3.946	15.224***	11.047***
	(3.444)	*(3.015)*	*(3.654)*	*(3.174)*
Tertiary degree in pedagogy	4.596**	2.890	4.294*	1.728
	(1.961)	*(1.813)*	*(2.258)*	*(2.072)*
Students	219,794	265,878	118,809	143,528
Schools	8,245	9,904	8,194	9,844
Countries	29	37	29	37
R^2	0.386	0.461	0.348	0.389

Dependent variable: PISA 2003 international test score. Least-squares regressions weighted by students' sampling probability. All five institutional variables are measured at the country level. The models additionally control for imputation dummies and interaction terms between imputation dummies and the variables. Robust standard errors adjusted for clustering at the school level in parentheses (clustering at country level for all country-level variables, which are all institutional variables, GDP per capita, OECD member, and expenditure per student). Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

Table C.2: Robustness specifications of the basic model: Country and grade sample

	Excl. Mexico and Turkey	Excl. grades 6 and 12	Only two largest grades per country	Drop grade controls
	(1)	(2)	(3)	(4)
External exit exams	13.828*	13.843*	13.574*	11.948*
	(7.489)	*(7.487)*	*(7.338)*	*(6.973)*
Autonomy in formulating budget	-26.705**	-25.430**	-26.402**	-19.613*
	(11.321)	*(10.677)*	*(10.372)*	*(10.576)*
Autonomy in staffing decisions	32.159*	29.502*	30.317**	25.283*
	(15.879)	*(14.719)*	*(14.791)*	*(13.974)*
Private operation	58.917***	61.717***	62.935***	61.337***
	(10.838)	*(10.465)*	*(10.985)*	*(11.168)*
Government funding	60.789**	74.842***	74.417***	61.668***
	(26.669)	*(20.886)*	*(20.763)*	*(19.104)*
Observations (students)	184,956	216,993	206,694	219,794
Clustering units (countries)	27	29	29	29
R^2	0.346	0.377	0.353	0.383

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. All five institutional variables are measured at the country level. Controls include: 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. The extended country sample specifications include an OECD dummy. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

Table C.3: Robustness specifications of the basic model: Controls and imputations

	Controls for tracking	Control for Europe dummy	Without imputation dummies	Imputation of constant
	(1)	(2)	(3)	(4)
External exit exams	14.108*	14.820*	11.188	11.738
	(7.263)	*(7.152)*	*(6.781)*	*(7.828)*
Autonomy in formulating budget	-24.419*	-37.843***	-20.344*	-22.485*
	(14.015)	*(10.933)*	*(11.507)*	*(11.812)*
Autonomy in staffing decisions	30.192**	25.388*	32.548*	31.407*
	(13.908)	*(13.864)*	*(16.103)*	*(15.752)*
Private operation	56.771***	80.924***	67.172***	58.735***
	(11.913)	*(9.832)*	*(11.623)*	*(10.309)*
Government funding	71.274***	163.281***	92.995***	65.464***
	(21.107)	*(28.866)*	*(20.626)*	*(17.945)*
Years tracked	-1.777			
	(2.613)			
Number of tracks	2.580			
	(3.550)			
Europe		-34.412***		
		(7.730)		
Observations (students)	219,794	219,794	219,794	219,794
Clustering units (countries)	29	29	29	29
R^2	0.387	0.392	0.350	0.379

Dependent variable: PISA 2003 international mathematics test score. Sample: OECD countries. Least-squares regressions weighted by students' sampling probability. All five institutional variables are measured at the country level. Controls include: 15 student characteristics, 16 family background measures, 9 measures of school location and resources, expenditure per student, GDP per capita, imputation dummies, and interaction terms between imputation dummies and the variables. The extended country sample specifications include an OECD dummy. Robust standard errors adjusted for clustering at the country level in parentheses. Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

Table C.4: Ordered probit regression results for non-cognitive skills

In ordered probit regressions, the interpretation of regression coefficients is not straightforward. In particular, from the sign of an ordered probit regression coefficient, it is impossible to make any statements about what happens to the probabilities of the middle categories. To evaluate the effect of a significant continuous regressor, we compute the marginal effect of this regressor evaluated at the sample means of all regressors (Table C.4b). The effects of dummy variables are evaluated by comparing the probabilities that result when the dummy variable takes its two different values, holding the other regressors at their sample means (Tables C.4a).

Table C.4a: Effects of dummy variables

"In the last two full weeks you were in school, how many times did you arrive late for school?"	Prob [y=1] none	Prob [y=2] 1 or 2 x	Prob [y=3] 3 or 4 x	Prob [y=4] ≥ 5 times
Assessments used to group students=0	0.6421	0.2520	0.0622	0.0437
Assessments used to group students=1	0.6548	0.2455	0.0591	0.0406
Change	-0.0127	0.0065	0.0031	0.0031
Private operation=0	0.6441	0.2513	0.0616	0.0430
Private operation=1	0.6682	0.2386	0.0559	0.0373
Change	-0.0241	0.0127	0.0057	0.0057
Attend school because local=0	0.6444	0.2512	0.0616	0.0429
Attend school because local=1	0.6528	0.2468	0.0596	0.0409
Change	-0.0087	0.0044	0.0020	0.0020
Urban=0	0.6922	0.2253	0.0503	0.0322
Urban=1	0.5515	0.2936	0.0851	0.0698
Change	0.1407	-0.0683	-0.0348	-0.0376
Urban, does not attend school because better	0.5427	0.2971	0.0874	0.0730
Urban, does attend school because better	0.6015	0.2721	0.0722	0.0542
Non-urban, does not attend school because better	0.6843	0.2298	0.0522	0.0338
Non-urban, does attend school because better	0.7127	0.2134	0.0457	0.0281

Dependent variable: tardiness. Interpretation: If assessments are used to group students in a school, the probability that students have not been late in the last two weeks is higher than in schools without such an accountability system, and the probability of being late is lower, holding all other regressors at their sample means.

Table C.4b: Marginal effect of government funding

"In the last two full weeks you were in school, how many times did you arrive late for school?"	Prob[y=1] None	Prob [y=2] 1 or 2 times	Prob [y=3] 3 or 4 times	Prob [y=4] ≥ 5 times
Government funding	-0.0249	0.1292	0.060	0.006

Dependent variable: tardiness. Interpretation: The higher the share of government funding, the lower the probability that students have not been late in the last two full weeks they were in school, holding all other regressors at their sample means.

REFERENCES

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, Michael Kremer (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review* 92 (5): 1535-1558.

Barro, Robert J., Xavier Sala-i-Martin (2004). *Economic Growth*. Second edition. Cambridge, MA: MIT Press.

Betts, Julian R. (1998). The Impact of Educational Standards on the Level and Distribution of Earnings. *American Economic Review* 88 (1): 266-275.

Bishop, John H. (1992). The Impact of Academic Competencies on Wages, Unemployment, and Job Performance. *Carnegie-Rochester Conference Series on Public Policy* 37: 127-194.

Bishop, John H. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *American Economic Review* 87 (2): 260-264.

Bishop, John H. (1999). Are National Exit Examinations Important for Educational Efficiency? *Swedish Economic Policy Review* 6 (2): 349-398.

Bishop, John H. (2006). Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure. In: Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Volume 2, pp. 909-944. Amsterdam: North-Holland.

Bishop, John H., Joan Y. Moriarty, Ferran Mane (2000). Diplomas for Learning, Not Seat Time: The Impact of New York Regents Examinations. *Economics of Education Review* 19 (4): 333-349.

Bishop, John H., Ludger Wößmann (2004). Institutional Effects in a Simple Model of Educational Production. *Education Economics* 12 (1): 17-38.

Björklund, Anders, Per-Anders Edin, Peter Freriksson, Alan Krueger (2004). Education, Equality and Efficiency: An Analysis of Swedish School Reforms during the 1990s. IFAU Report 2004:1. Uppsala: Institute for Labour Market Policy Evaluation.

Bradley, Steve, Jim Taylor (2002). The Effect of the Quasi-Market on the Efficiency-Equity Trade-off in the Secondary School Sector. *Bulletin of Economic Research* 54 (3): 295-314.

Carnoy Martin, Susanna Loeb (2003). Does External Accountability Improve Student Outcomes? A Cross-State Analysis. *Education Evaluation and Policy Analysis* 24 (4): 305-331.

Chubb, John E., Terry M. Moe (1990). *Politics, Markets, and America's Schools*. Washington, D.C.: Brookings Institution Press.

Costrell, Robert M. (1994). A Simple Model of Educational Standards. *American Economic Review* 84 (4): 956-971.

Cunha, Flavio, James J. Heckman, Lance Lochner, Dimitriy V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In: Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Volume 1, pp. 697-812. Amsterdam: North-Holland.

Deaton, Angus (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

Deke, John, Joshua Haimson (2006). Valuing Student Competencies: Which Ones Predict Postsecondary Educational Attainment and Earnings, and for Whom? Princeton, NJ: Mathematica Policy Research.

Duckworth, Angela L., Martin E. P. Seligman (2005). Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents. *Psychological Science* 16 (12): 939-944.

DuMouchel, William H., Greg J. Duncan (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association* 78 (383): 535-543.

Duncan, Greg J., Rachel Dunifon (1998). 'Soft-skills' and Long-run Labor Market Success. *Research in Labor Economics* 17: 123-149.

Dunifon, Rachel, Greg J. Duncan, Jeanne Brooks-Gunn (2001). As Ye Sweep, So Shall Ye Reap. *American Economic Review* 91 (2): 150-154.

Fuchs, Thomas, Ludger Wößmann (2005). The Determinants of Differences in Primary-School Learning across Countries. Mimeo, Ifo Institute for Economic Research at the University of Munich.

Fuchs, Thomas, Ludger Wößmann (2007). What Accounts for International Differences in Student Performance? A Re-examination using PISA Data. *Empirical Economics* 32 (2-3): 433-464.

Gradstein, Mark, Moshe Justman, Volker Meier (2004). *The Political Economy of Education: Implications for Growth and Inequality*. Cambridge, MA: MIT Press.

Hanushek, Eric A. (1994). Education Production Functions. In: Torsten Husén, T. Neville Postlethwaite (eds.), *International Encyclopedia of Education*, 2nd edition, Volume 3, pp. 1756-1762. Oxford: Pergamon.

Hanushek, Eric A. (2002). Publicly Provided Education. In: Alan J. Auerbach, Martin Feldstein (eds.), *Handbook of Public Economics, Volume 4*, pp. 2045-2141. Amsterdam: North-Holland.

Hanushek, Eric A., with Charles S. Benson, Richard B. Freeman, and others (1994). *Making Schools Work: Improving Performance and Controlling Costs*. Washington, D.C.: Brookings Institution Press.

Hanushek, Eric A., Dennis D. Kimko (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review* 90 (5): 1184-1208.

Hanushek, Eric A., Margaret E. Raymond (2004). The Effect of School Accountability Systems on the Level and Distribution of Student Achievement. *Journal of the European Economic Association* 2 (2-3): 406-415.

Hanushek, Eric A., Ludger Wößmann (2006). Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries. *Economic Journal* 116 (510): C63-C76.

Hanushek, Eric A., Ludger Wößmann (2007a). The Role of School Improvement in Economic Development. NBER Working Paper 12832. Cambridge, MA: National Bureau of Economic Research.

Hanushek, Eric A., Ludger Wößmann (2007b). *Education Quality and Economic Growth*. Washington, DC: World Bank.

Heckman, James J. (2000). Policies to Foster Human Capital. *Research in Economics* 54 (1): 3-56.

Heckman, James J., Yona Rubinstein (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review* 91 (2): 145-149.

Heckman, James J., Jora Stixrud, Sergio Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour. *Journal of Labor Economics* 24 (3): 411-482.

Heston, Alan, Robert Summers, Bettina Aten (2006). Penn World Table Version 6.2. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Philadelphia: University of Pennsylvania.

Holmstrom, Bengt, Paul Milgrom (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 7 (Special Issue): 24-52.

Howell, William G., Patrick J. Wolf, David E. Campbell, Paul E. Peterson (2002). School Vouchers and Academic Performance: Results from Three Randomized Field Trials. *Journal of Policy Analysis and Management* 21 (2): 191-217.

Hoxby, Caroline M. (1999). The Productivity of Schools and Other Local Public Goods Producers. *Journal of Public Economics* 74 (1): 1-30.

Hoxby, Caroline M. (2000). Does Competition among Public Schools Benefit Students and Taxpayers? *American Economic Review* 90 (5): 1209-1238.

Hoxby, Caroline M. (2003). School Choice and School Competition: Evidence from the United States. *Swedish Economic Policy Review* 10 (3): 9-65.

Jacob, Brian A. (2002). Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education. *Economics of Education Review* 21 (6): 589-598.

Jacob, Brian A. (2005). Accountability, Incentives and Behaviour: The Impact of High-stakes Testing in the Chicago Public Schools. *Journal of Public Economics* 89 (5-6): 761-796.

Jacob, Brian A., Steven D. Levitt (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics* 118 (3): 843-877.

Jimenez, Emmanuel, Vicente Paqueo (1996). Do Local Contributions Affect the Efficiency of Public Primary Schools? *Economics of Education Review* 15 (4): 377-386.

Jürges, Hendrik, Kerstin Schneider, Felix Büchel (2005). The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany. *Journal of the European Economic Association* 3 (5): 1134-1155.

Laffont, Jean-Jacques, David Martimort (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press.

Lazear, Edward P. (2003). Teacher Incentives. *Swedish Economic Policy Review* 10 (3): 179-214.

Levaĉić, Rosalind (2004). Competition and the Performance of English Secondary Schools: Further Evidence. *Education Economics* 12 (2): 177-193.

Little, Roderick J.A., Donald B. Rubin (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Loury, Glenn C. (1981). Intergenerational Transfers and the Distribution of Earnings. *Econometrica* 49 (4): 843-867.

McIntosh, Steven, Anna Vignoles (2001). Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes. *Oxford Economic Papers* 53 (3): 453-481.

Moulton, Brent R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics* 32 (3): 385-397.

Mulligan, Casey B. (1999). Galton versus the Human Capital Approach to Inheritance. *Journal of Political Economy* 107 (6, pt. 2): S184-S224.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, John H. Tyler (2000). How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings? *Journal of Policy Analysis and Management* 19 (4): 547-568.

Neal, Derek (1997). The Effects of Catholic Secondary Schooling on Secondary Achievement. *Journal of Labor Economics* 15 (1): 98-123.

Nechyba, Thomas J. (2003). Centralization, Fiscal Federalism, and Private School Attendance. *International Economic Review* 44 (1): 179-204.

Organisation for Economic Co-operation and Development (OECD) (2000). *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. Paris: OECD.

Organisation for Economic Co-operation and Development (2004). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

Organisation for Economic Co-operation and Development (2005a). *PISA 2003 Technical Report*. Paris: OECD.

Organisation for Economic Co-operation and Development (2005b). *PISA 2003 Data Analysis Manual*. Paris: OECD.

Organisation for Economic Co-operation and Development (2006a). *Demand-Sensitive Schooling? Evidence and Issues*. Paris: OECD, Centre for Educational Research and Innovation (CERI).

Organisation for Economic Co-operation and Development (2006b). *Education at a Glance: OECD Indicators 2006*. Paris: OECD.

Sandström, F. Mikael, Fredrik Bergström (2005). School Vouchers in Practice: Competition Will Not Hurt You. *Journal of Public Economics* 89 (2-3): 351-380.

Schafer, Joseph L., Nathalie Schenker (2000). Inference With Imputed Conditional Means. *Journal of the American Statistical Association* 95 (449): 144-154.

Schütz, Gabriela, Martin R. West, Ludger Wößmann (2007). *School Accountability, Autonomy, Choice and the Equity of Student Achievement: International Evidence from PISA 2003*. OECD Education Working Paper No. 12, EDU/WKP(2007)7. Paris: OECD.

Shleifer, Andrei (1998). State versus Private Ownership. *Journal of Economic Perspectives* 12 (4): 133-150.

Stiglitz, Joseph E. (2002). Information and the Change in the Paradigm in Economics. *American Economic Review* 92 (3): 460-501.

Todd, Petra E., Kenneth I. Wolpin (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *Economic Journal* 113 (485): F3-F33.

Toma, Eugenia F. (1996). Public Funding and Private Schooling across Countries. *Journal of Law and Economics* 39 (1): 121-148.

Toma, Eugenia F. (2005). Private Schools in a Global World: 2004 Presidential Address. *Southern Economic Journal* 71 (4): 693-704.

Vandenberghe, Vincent, Stephane Robin (2004). Evaluating the Effectiveness of Private Education across Countries: A Comparison of Methods. *Labour Economics* 11 (4): 487-506.

White, Halbert (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

Wolfe, Raymond N., Scott D. Johnson. (1995). Personality as a Predictor of College Performance. *Educational and Psychological Measurement* 55 (2): 177-185.

Wooldridge, Jeffrey M. (2001). Asymptotic Properties of Weighted *M*-Estimators for Standard Stratified Samples. *Econometric Theory* 17 (2): 451-470.

Wößmann, Ludger (2001). Why Students in Some Countries Do Better: International Evidence on the Importance of Education Policy. *Education Matters* 1 (2): 67-74.

Wößmann, Ludger (2002). *Schooling and the Quality of Human Capital*. Berlin: Springer.

Wößmann, Ludger (2003a). Schooling Resources, Educational Institutions and Student Performance: the International Evidence. *Oxford Bulletin of Economics and Statistics* 65 (2): 117-170.

Wößmann, Ludger (2003b). Central Exit Exams and Student Achievement: International Evidence. In: Paul E. Peterson, Martin R. West (eds.), *No Child Left Behind? The Politics and Practice of School Accountability*, pp. 292-323. Washington, D.C.: Brookings Institution Press.

Wößmann, Ludger (2005a). The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics* 13 (2): 143-169.

Wößmann, Ludger (2005b). Educational Production in Europe. *Economic Policy* 20 (43): 445-504.

Wößmann, Ludger (2006). Public-Private Partnerships and Schooling Outcomes across Countries. CESifo Working Paper 1662. Munich: CESifo.

Wößmann, Ludger (2007a). International Evidence on Expenditures and Class Size: A Review. *Brookings Papers on Education Policy* 2006/2007: 245-272.

Wößmann, Ludger (2007b). International Evidence on School Competition, Autonomy and Accountability: A Review. *Peabody Journal of Education* 82 (2-3): 473-497.

Wößmann, Ludger (2007c). Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries. Harvard University, Program on Education Policy and Governance Research Paper PEPG 07-02. Cambridge, MA: Harvard University.