Rolf Strietholt
Wilfried Bos
Jan-Eric Gustafsson
Monica Rosén (Eds.)

Educational Policy Evaluation through International Comparative Assessments

Waxmann 2014
Münster • New York

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in
the Deutsche Nationalbibliografie; detailed bibliographic data
are available in the Internet at http://dnb.d-nb.de

Print-ISBN 978-3-8309-3091-4
E-Book-ISBN 978-3-8309-8091-9

www.waxmann.com
info@waxmann.com

Cover design: Anne Breitenbach, Tübingen
Typesetting: Stoddart Satz- und Layoutservice, Münster
Print: Hubert & Co., Göttingen

Printed on age-resistant paper,
acid-free as per ISO 9706

Contents

PART A
CONCEPTUAL AND METHODOLOGICAL FOUNDATIONS

Eric A. Hanushek and Ludger Woessmann

Institutional Structures of the Education System and Student Achievement: A Review of Cross-country Economic Research*

"If custom and law define what is educationally allowable within a nation,
the educational systems beyond one's national boundaries
suggest what is educationally possible."
Arthur W. Foshay (1962) on the first pilot study
of international student achievement

Virtually all nations of the world today realize the research and policy value of student performance data that come from testing the cognitive skills of students. While there is wide variation across nations in testing – differing by subject matter, grade level, purpose, and quality of testing – the idea of assessing what students know as opposed to how long they have been in school has diffused around the world, in part at the instigation of international development and aid agencies. Somewhat less known is that comparative cross-national testing has been going on for a long time. Nations participated in common international assessments of mathematics and science long before they instituted national testing programs. These common international assessments provide unique data for understanding both the importance of various factors determining achievement and the impact of skills on economic and social outcomes.

International consortia were formed in the mid-1960s to develop and implement comparisons of educational achievement across nations. Since then, the math, science, and reading performance of students in many countries have been tested on multiple occasions using (on each occasion) a common set of test questions in all participating countries. By 2013, three major international testing programs are surveying student performance on a regular basis: the Programme for International Student Assessment (PISA) testing math, science, and reading performance of 15-year-olds on a three-year cycle since 2000, the Trends in International Mathematics and Science Study (TIMSS) testing math and science performance (mostly) of eighth-graders on a four-year cycle since 1995, and the Progress in International Reading

* This is an extract from a much longer article previously published under the title "The Economics of International Differences in Educational Achievement" in the Handbook of the Economics of Education, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, Amsterdam: North Holland, pp. 89–200, Copyright Elsevier (2011). We are grateful to Elsevier for granting us the right to reproduce the material here.

Literacy Study (PIRLS) testing primary-school reading performance on a five-year cycle since 2001.

In a variety of cases, these international assessments actually substitute for national testing. The international testing provides information on educational outcomes where otherwise only small, unrepresentative samples of outcome data are available. Indeed, the simplest of international comparisons has spurred not only governmental attention but also immense public interest as is vividly documented by the regular vigorous news coverage and public debate of the outcomes of the international achievement tests in many of the participating countries. For example, the results of the first PISA study made headlines on the front pages of tabloids and more serious newspapers alike: the *Frankfurter Allgemeine Zeitung* (Dec. 4, 2001) in Germany titled "Abysmal marks for German students", *Le Monde* (Dec. 5, 2001) in France titled "France, the mediocre student of the OECD class", and *The Times* (Dec. 6, 2001) in England titled "Are we not such dunces after all?"

The research based on the international assessments goes in two different directions: research designed to understand the underlying determinants of cognitive skills and research focused on the consequences of skill differences. Here, we simply focus on surveying the literature on institutional structures of the education system as one group of determinants of international educational achievement, covering both evidence within different countries and evidence across countries. For research on student background and school inputs as two other groups of possible determinants, see sections 4.2 and 4.3 in Hanushek and Woessmann (2011). For the second line of research, see their section 5. Furthermore, their Sections 2 and 3 provide a brief economic motivation to frame the discussions and an overview and critical assessment of the different available international datasets on educational achievement.

The standards of evidence throughout empirical economics have changed in recent years, sometimes dramatically. The character of change also enters directly into our consideration of cross-country analyses. The analytical designs employed in the cross-country analyses we discuss have developed over time in a way that parallels much of the related micro-econometric work within individual countries. The initial publications of comparative tests across nations by the organizations that conducted the different studies tended to report bivariate associations. Subsequent analyses performed multiple regressions in the form of educational production functions that tried to address the most obvious perils of bias from intervening factors by adding corresponding control variables. While initial studies estimated international educational production functions at the aggregate country level, subsequent studies exploited the full variation of the international micro data.

More recently, several studies have started to employ econometric techniques such as instrumental-variable, regression-discontinuity, differences-in-differences, and different sorts of fixed-effects specifications in order to come closer to identification of causal relationships in the international data on educational achievement. This applies both to the identification of causal effects within countries and to the challenge of overcoming possible bias from unobserved country heterogeneity – e.g., in terms of cultural differences – in cross-country estimation. While these developments are far from complete at this time, we emphasize the issues of identification and interpretation in much of the discussion below.

We limit the coverage of this chapter to studies that make cross-country comparisons. Based on this criterion, we cover only studies that estimate the same specification for different countries or estimate a cross-country specification. Studies that use the international survey data for analysis within a single country will be referenced only insofar as they are directly relevant for the internationally comparative approach.

Unique Advantages of and Concerns with the Use of Cross-Country Data

International achievement data, developed and refined over the past half century, were not collected to support any specific economic research agenda. But, as we shall discuss below, there are a number of research and policy agendas that are uniquely amenable to analysis because of the existence of such data. We argue that such data have made it possible for economists to address a range of fundamental questions that previously resisted satisfactory analysis. And, because the extent and nature of international achievement data still remain largely unknown, it is important to evaluate the advantages and disadvantages of these data in understanding a variety of significant research and policy questions.

In terms of understanding the determinants of educational achievement, the international data have at least six unique advantages over research restricted to single countries or states. First, the data permit exploitation of variation that only exists across countries. For example, systematic institutional variation between countries as found with differences in the competitiveness and flexibility of teacher labor markets, forms of accountability systems, the extent of a private school sector, or the structure of student tracking simply does not exist within most countries. Or, the existence of central exit exams is a national characteristic in nearly all countries, so that the effect of central exams cannot be estimated using national data in these countries unless their status changes over time. The lack of within-country institutional variation makes an empirical identification of the impact of many institutional features of school systems impossible when using national datasets.

Second, even where within-country variation exists, variations across countries in key institutional factors and in characteristics of the schools and population are frequently much larger than those found within any country. From an analytical viewpoint, using such international variation generally implies increased statistical power to detect the impact of specific factors on student outcomes.

Third, the international achievement data based on the same data-collection process provides an opportunity to examine comparable estimates of the determinants and consequences of educational achievement for a diverse set of countries. Such research can thus throw light on whether a result is truly country-specific, applies

more generally, or is simply a spurious result from a particular within-country sample.

Fourth, and related to the previous point, international evidence can identify systematic heterogeneity in effects that differ across countries. For example, such comparative research can delve into why class-size effects on achievement are heterogeneous across countries, perhaps leading to deeper insights about, say, the interaction between curriculum or teacher training and classroom processes.

Fifth, 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 some differences in student, family, 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, it is possible to circumvent these selection problems – in effect measuring the impact of, for example, the share of students in a country attending private schools on student achievement in the country as a whole. Such cross-country evidence will not be biased by standard issues of selection at the individual level.

Sixth, uncovering general equilibrium effects is often impossible in a single country but sometimes feasible across a number of countries. For example, the presence of private schools may influence the behavior 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. 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.

With these research advantages also come concerns and disadvantages. Three stand out. First, the relevant variations are frequently limited by the number of countries with both assessment and other common data. Second, even though each of the assessments collects substantial amounts of ancillary survey information at the individual level, virtually all are single cross-sectional designs with no ability to track individuals. Third, there is frequently a concern that unmeasured "cultural" factors are important in various processes of interest. Each of these makes the identification and estimation of cross-country models difficult and limits the range of analyses currently possible.

Further, while not specific to this cross-country work, some inherently difficult data and modeling problems also remain. The focus of this chapter is measures of educational achievement – skills that are expressed in test scores – rather than quantitative measures of educational attainment. For reasons of availability, the focus of our skill measurement is just on cognitive skills, opening up possible concerns about other skills such as non-cognitive skills. The systematic measurement of such skills has yet to be possible in international comparisons. Furthermore, the research covered refers to basic general skills that are generally learned through the end of secondary school, leaving aside programs of higher education and specific vocational skills. Apart from data availability, this focus is also dictated by a need for international comparability where measures of any quality aspects of higher education are generally unavailable.

International Evidence on Education Production Functions

As is the case in the majority of the literature on educational production, the basic model underlying the literature on determinants of international educational achievement resembles some form of the education production function:

$$T = a_0 + a_1 F + a_2 R + a_3 I + a_4 A + e$$

where T is the outcome of the educational production process as measured, e.g., by test scores of mathematics, science, and reading achievement. The vector F captures facets of student and family background characteristics, R is a vector of measures of school resources, I are institutional features of schools and education systems, and A is individual ability.

When estimating this equation within different countries, studies based on international data face the same methodological challenges as studies restricted to a specific country (see Hanushek 1979, 2002; Todd & Wolpin, 2003) for key issues in empirical identification of education production functions). The fundamental challenge is that most inputs in the education production function are likely not to be exogenous in a statistical sense. Leading concerns derive from omitted variables, selection, and reverse causation. A key candidate of an omitted variable is student ability A, most dimensions of which tend to go unmeasured and are likely correlated with other inputs in important ways. An additional concern for research on most of the international tests is their cross-sectional structure which does not allow for panel or value-added estimations, so that temporally prior inputs are usually unobserved. School inputs will often be the outcome of the choices of parents, administrators, and schools that are correlated with the error term of the production function. The same is true for some institutional characteristics. Given this substantial scope for endogeneity bias, least-squares estimates of the equation need to be interpreted with great care, even when they control for a large set of observable input factors. This has led to the development of more elaborate techniques that try to draw on exogenous variation in the variables of interest.

In the following review of the literature, we will refer to the more descriptive studies only briefly and mostly focus on studies trying to address the key identification issues. There is, however, one specific aspect about making cross-country comparisons of estimates obtained from performing the same estimation in different countries: If one is willing to make the assumption that any bias is constant across

countries, then a cross-country comparison of estimates is feasible, even if interpretation of the size of each estimate is not.

The main challenges change when it comes to studies estimating cross-country associations. As discussed above, there are both unique advantages and specific concerns with using cross-country data to estimate the determinants of educational achievement. At the most general level, cross-country estimation is able to get around the most pressing concerns of bias from selection but introduces new kinds of omitted variable concerns. Within-country variation is often subject to severe selection problems: For example, students who choose to attend a private school may differ along both observable and unobservable dimensions from students taught in neighborhood public schools. While many observable characteristics are often controlled for in econometric analyses, thereby comparing students who are observationally equivalent, within-country estimates may still suffer from selection on unobserved characteristics. In cross-country analyses, one can aggregate the institutional variable of interest up to the country level, thereby circumventing the selection problem. In effect, the cross-country analysis then measures the impact of, for example, the share of students in a country attending private schools on student achievement in the country as a whole. Such cross-country analysis cannot be biased by standard issues of selection at the individual level, as patterns of sorting cancel out at the system level.

The main cost to this – apart from the limited degrees of freedom at the country level – is that unobserved heterogeneity at the country level may introduce new forms of omitted variable bias. For example, cultural factors such as "Asian values" may remain unobserved in the econometric model and correlate both with student outcomes and relevant inputs in the education production function. Education systems – and societies more generally – may also differ in other important dimensions unobserved by the researcher. To address such concerns, the main results of cross-country studies should be checked for robustness to including obvious correlates of the cultural factors as control variables at the country level. Another robustness check is to draw only on variation within major world regions by including regional (continental) fixed effects. More fundamentally, some cross-country studies have started to adopt new techniques directly developed to address such issues of identification in particular contexts, and these studies will be the main focus of the following review.

Early studies that employ the international student achievement tests to estimate similar education production function within different countries include Heynemann and Loxley (1983) and Toma (1996). Early studies using the cross-country variation of international tests to estimate international education productions on country-level observations include Bishop (1997), Hanushek and Kimko (2000), and Lee and Barro (2001). The first economic study to make use of the vast potential of the international micro data on students' achievement, family background, and school inputs and of the broad array of institutional differences that exists across countries to esti-

Table 1: An example of an international education production function: PISA 2003

	Coef.	Std. err.
STUDENT CHARACTERISTICS		
Age (years)	17.593***	*(1.101)*
Female	-17.360***	*(0.639)*
Preprimary education (more than 1 year)	5.606***	*(0.703)*
School starting age	-3.863***	*(0.505)*
Grade repetition in primary school	-35.794***	*(1.410)*
Grade repetition in secondary school	-34.730***	*(1.646)*
Grade		
7th grade	-47.184***	*(4.068)*
8th grade	-28.009***	*(2.239)*
9th grade	-12.486***	*(1.337)*
11th grade	-6.949***	*(2.062)*
12th grade	7.030	*(4.826)*
Immigration background		
First generation student	-9.047***	*(1.544)*
Non-native student	-9.040***	*(1.644)*
Language spoken at home		
Other national dialect or language	-23.736***	*(2.849)*
Foreign language	-8.381***	*(1.665)*
FAMILY BACKGROUND		
Living with		
Single mother or father	19.349***	*(1.842)*
Patchwork family	21.272***	*(2.032)*
Both parents	27.432***	*(1.829)*
Parents' working status		
Both full-time	-2.479*	*(1.325)*
One full-time, one half-time	6.744***	*(1.063)*
At least one full time	13.753***	*(1.173)*
At least one half time	8.416***	*(1.133)*
Parents' job		
Blue collar high skilled	0.431	*(0.970)*
White collar low skilled	2.864***	*(0.933)*
White collar high skilled	8.638***	*(0.988)*
Books at home		
11-25 books	5.554***	*(0.978)*
26-100 books	22.943***	*(1.009)*
101-200 books	32.779***	*(1.117)*
201-500 books	49.834***	*(1.219)*
More than 500 books	51.181***	*(1.399)*
Index of Economic, Social and Cultural Status (ESCS)	18.114***	*(0.524)*
GDP per capita (1,000 $)	-1.890*	*(1.060)*

(continued on next page)

Table 1 (continued)

	Coef.	Std. err.
SCHOOL INPUTS		
School's community location		
Town (3,000-100,000)	3.226*	*(1.531)*
City (100,000-1,000,000)	10.782***	*(1.890)*
Large city with > 1 million people	7.895***	*(2.378)*
Educational expenditure per student (1,000 $)	1.174***	*(0.405)*
Class size (mathematics)	1.474***	*(0.067)*
Shortage of instructional materials		
Not at all	-10.180***	*(2.576)*
Strongly	6.720***	*(1.300)*
Instruction time (minutes per week)	0.035***	*(0.005)*
Teacher education (share at school)		
Fully certified teachers	9.715***	*(3.422)*
Tertiary degree in pedagogy	6.573***	*(2.010)*
INSTITUTIONS		
Choice		
Private operation	57.585***	*(8.355)*
Government funding	81.839***	*(22.327)*
Accountability		
External exit exams	25.338*	*(10.054)*
Assessments used to decide about students' retention/ promotion	12.185***	*(1.631)*
Monitoring of teacher lessons by principal	4.557***	*(1.343)*
Monitoring of teacher lessons by external inspectors	3.796***	*(1.415)*
Assessments used to compare school to district/national performance	2.134*	*(1.259)*
Assessments used to group students	-6.065***	*(1.301)*
Autonomy and its interaction with accountability		
Autonomy in formulating budget	-9.609***	*(2.178)*
External exit exams x Autonomy in formulating budget	9.143***	*(3.119)*
Autonomy in establishing starting salaries	-8.632***	*(3.251)*
External exit exams x Autonomy in establishing starting salaries	5.868	*(3.980)*
Autonomy in determining course content	0.175	*(1.907)*
External exit exams x Autonomy in determining course content	3.224	*(2.858)*
Autonomy in hiring teachers	20.659***	*(2.249)*
External exit exams x Autonomy in hiring teachers	-28.935***	*(3.365)*
Students	219,794	
Schools	8,245	
Countries	29	
R^2 (at student level)	0.390	
R^2 (at country level)	0.872	

Note. Dependent variable: PISA 2003 international mathematics test score. Least-squares regressions weighted by students' sampling probability. 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 private operation, government funding, external exit exams, GDP per capita, and expenditure per student). Significance level (based on clustering-robust standard errors): *** 1 percent, ** 5 percent, * 10 percent.

Source. Own calculations based on Woessmann, Luedemann, Schuetz, and West (2009), who provide additional background details.

mate extensive multivariate cross-country education production functions is Woessmann (2003b). While still subject to the prior issues of cross-country identification, employing the rich student-level data on background factors allows to hold constant a large set of observable factors usually unavailable in national datasets.

Table 1 presents an example estimation of an international education production function.[1] Using student-level data for 29 OECD countries from the 2003 cycle of the PISA test of 15-year-olds, the model expresses individual student achievement in math as a function of large set of input factors. While this is a basic model that does not fully exploit the potential of the international data, the model specification already documents the rich set of background factors available from the student and school background questionnaires. Moreover, the international data display wide variation in many of the potential inputs to achievement, thus allowing for more precise estimation of any effects. At the individual level, the factors include student characteristics such as age, gender, immigration, and preprimary educational attendance and family-background measures such as socio-economic status, parental occupation, family status, and the number of books in the home. At school level, the model includes resource measures such as class size and shortage of materials, instruction time, teacher education, community location, and institutional factors such as a set of measures of teacher monitoring and student assessment, different dimensions of school autonomy, and their interaction with accountability measures. At the country level, this basic model includes a country's GDP per capita, educational expenditure per student, and the institutional factors of external exit exams, share of privately operated schools, and average government funding of schools.

While the cross-sectional nature of this estimation allows for a descriptive interpretation only, it is worth noting that many measures of students' individual and family background are systematically related to their achievement, as are several measures of the institutional structure of the school system. By contrast, the point estimate on class size, the classical measure of quantitative school inputs, is counterintuitive,[2] and the estimates on the more qualitative school inputs, while positive, are more limited than the background and institutional estimates. The model accounts for 39 percent of the achievement variation at the student level and for 87 percent at the country level. That is, while unobserved factors such as ability differences are important at the individual level, the model is able to account statistically for most of the between-country variation in academic achievement. These basic result patterns are broadly common to all studies of international education production functions estimated on the different international student achievement tests. We discuss the literature on the first two groups of determinants – student and family

1 See Woessmann, Luedmann, Schuetz, and West (2009) for additional background and robustness analyses.

2 The coefficient on country-level spending is very small. While it is statistically significant, identification here comes from a very particular margin, as the correlation between spending and per-capita GDP (whose coefficient is negative here) in this model is as high as 0.93. Other studies tend to find a significant positive coefficient on GDP per capita, but not on spending. See Hanushek and Woessmann (2011) for more extensive discussion.

background, as well as school inputs – in greater detail. Here, we focus on institutional structures of the education system as the third group of determinants.

Motivated by the poor results on school inputs, research has increasingly focused on whether non-resource institutional features of school systems affect student outcomes. In this topic, the particular opportunity of cross-country research comes into play: The chief advantage of the international comparative approach stems from its ability to exploit the substantial variation in national education policies across countries (cf. Woessmann, 2007b). By contrast, within-country studies are usually restricted to analysis of much more limited variation in institutional structure. Moreover, by drawing on wider and long-established institutional variation between countries, the international approach can capture general-equilibrium effects of institutional settings, which will not necessarily be the case when a specific educational reform is introduced only on a small scale, or only very recently. Such long-term general-equilibrium effects are usually the ones that economic theory stresses as being particularly important, because persistent institutional changes will alter incentives and thus behavior. By changing prices, available alternatives and competitive pressures for other market participants will have effects on market outcomes beyond the people specifically treated.

Since cross-country studies can address the most obvious issues of selection into treatment by using average measures of institutions at the systemic level, the main challenge for the identification of causal effects lies in unobserved country heterogeneity. Institutions may be correlated with other, unobserved country characteristics that are related to student achievement. While still in its infancy, several methods have been developed to address this problem, tailored to specific worries related to each specific institution. As will be discussed below, the range includes fixed effects for world regions to eliminate the most basic cultural differences; within-country identification where different education systems exist within one country (holding constant differences in language, legal structures, and cultures); differences-in-differences models that identify effects from changes between grades within each country; and the use of historical instruments that gave rise to arguably exogenous variation in institutional structures today.

The following review is structured around five institutional features that have attracted the most attention in the international literature so far: accountability measures, school autonomy, competition and private involvement, school tracking, and the pre-primary education system. Tables 2–4 provide details on the individual studies analyzing institutional features. Table 2 reports evidence within different countries and the other two tables report cross-country evidence. Given that different institutional features tend to be related both to the level and to the equity of outcomes, Table 3 focuses on achievement levels, and Table 4 on the equity of achievement.

Table 2: *Within-country studies on institutions and educational achievement*

Study	Data-set	Countries	Level of analysis	Topic of investi-gation	Measure of insti-tutions	Measure of achieve-ment	Estimation method	Results
8ishop (1995), ch. 6	IAEP-II	Canada, U.S.	Student	Effect of curriculum-based external exams	Central exams, type of school	Math + science	Cross-Section	External exams positively associated with student achievement; also with student, parental, and teacher behavior
Toma (1996)	SIMS	Belgium, France, New Zealand, Ontario (Can.), U.S.	Student	Effects of public funding and private schools	Type of school (public/ private)	Math, beginning and end of school year	Value-added achieve-ment model	Positive effect of private schools; funding not significantly associated with performance; governmental control over private schools negative factor
Vanden-berghe and Robin (2004)	PISA	9 countries	Student	Private vs. public edu-cation	Type of school (public/ private)	Math, science, + reading	Cross-section IV, Heckman two stages, PSM	Significant positive association of private schools with achievement in some but not all countries
Corten and Dronkers (2006)	PISA 2000	19 countries	Student	Low-SES students and private schools	Gover-nance and funding of school.	Math + reading	MLM	Slight advantage of private government-dependent schools, no significant differences between public and private-independent schools
Dronkers and Robert (2008)	PISA 2000	22 countries	Student	Public and private schools	Gover-nance and funding of school.	Reading	MLM	Better performance of government-dependent private schools explained by better school climate

Note. SES = socio-economic status. IV = instrumental variable. PSM = propensity score matching. MLM = multi-level modeling. See Tables 1 and 2 in Hanushek and Woessmann (2011) for acronyms of datasets.

Table 3: Cross-country studies on institutions and levels of educational achievement

Study	Dataset	No. of countries	Level of analysis	Topic of investigation	Measure of institutions	Measure of achievement	Estimation method	Results
Bishop (1995), ch. 4	IAEP-II	15-21	Country	Effects of CBEEE	CBEEE	Math, science, + geography	Cross-section OLS	Student achievement and teacher salaries higher in CBEEE countries; differences in qualifications and spending not significant
Bishop (1997)	TIMSS, IAEP-II	39, Canada	Country, School	Effects of CBEEE	CBEEE	Math + science	Cross-section OLS	Large effect of CBEEE on student achievement; effects on parent, teacher, administrator behavior
Woessmann (2003b)	TIMSS	39	Student	Effects on student performance	Seven different categories	Math + science	Cross-section WCRLR	Large effects of institutional arrangements such as external exit exams, school autonomy, and private competition; far more important than resources
Woessmann (2003a)	TIMSS+ TIMSS-R	39, 38 (54)	Student	Effects of central exit exams	Central exit exams	Math + science	Cross-section WCRLR	Performance of students higher in systems with central exams; positive interaction with autonomy
Woessmann (2005c)	TIMSS+ TIMSS-R + PISA	39, 38 (54), 32	Student	Heterogeneity of central exam effect	Central exit exams, school autonomy	Math + science	Cross-section WCRLR, quantile regr.	Substantial heterogeneity of central exam effects along student, school, and time dimension
Bishop (2006), ch. 3	PISA	41	Country	Effects of MCE and CBEEE	CBEEE	Math, science + reading	Cross-section OLS	Positive effects of CBEEE on student achievement; do not affect school attendance
Fuchs and Woessmann (2007)	PISA	31	Student	Effects on student performance	CBEEE, autonomy, private schools	Math, science, + reading	Cross-section WCRLR, IV	Institutional variation accounts for a quarter of between-country achievement variation; external exams interact positively with autonomy; positive effect of private operation
Sprietsma (2008)	PISA 2003	8	Student	School choice, school selectivity, and student performance	School choice, schools' student selection	Math, reading + science	Cross-section, MLM, quantile regression	Regional intensity of school choice and school selectivity positively related to student achievement; similar effect for low and high performing students
Woessmann (2009b)	PISA	29	Student	Public vs. private school funding and operation	Private operation and funding	Math + reading	Cross-section WCRLR	Negative effects of public operation on student achievement; positive effect of public funding
Woessmann, Luedemann, Schuetz, and West (2009, ch. 2–6	PISA 2003	29, 37	Student	Accountability, autonomy, and choice	Several measures of accountability, autonomy, choice	Math +science	Cross-section WCRLR	Positive effects of several accountability measures on student performance and on role of autonomy; positive effects of share of privately operated schools and of government funding
West and Woessmann (2010)	PISA 2003	29	Student	Effect of competition from private schools on student achievement	Share of privately operated schools	Math, science, + reading	Cross-section WCRLR, IV (instrumenting private school share by historical Catholic share)	Positive causal effect of share of privately operated schools on student achievements, negative effect on costs
Schuetz (2009)	PISA 2003	38	Student	Effect of pre-primary education on later educational achievement	Characteristics of pre-primary education system	Math	Cross-section WCRLR, country fixed effects, (DiD)	Positive association of pre-primary attendance with test scores; systematically stronger in countries with higher spending, larger shares of privately managed institutions, and higher training and relative pay of educators in pre-primary system

Note. CBEEE = curriculum based external exit exams. MCE = minimum competency exams. WCRLR = weighted clustering-robust linear regression. OLS = ordinary least squares. IV = instrumental variable. MLM = multilevel modeling. DiD = differences in differences. See Tables 1 and 2 in Hanushek and Woessmann (2011) for acronyms of datasets.

Table 4: *Cross-country studies on institutions and equity of educational achievement*

Study	Dataset	No. of countries	Level of analysis	Topic of investigation	Measure of institutions	Measure of achievement	Measure of equity	Estimation method	Results
Hanushek and Woessmann (2006)	PISA, PIRLS, TIMSS	45	Country	Early tracking and inequality	Age of first tracking	Math, science, + reading	Standard deviation in test scores	Pooled data, DiD	Significant effect of early tracking on inequality; no clear effect on mean performance
Schuetz, Ursprung, and Woessmann (2008)	TIMSS+ TIMSS-R	54	Student, country	Equality of opportunity	Age of first tracking, pre-school enrollment and duration	Mean math + science	Dependence of test scores on books at home	Cross-section WCRLR, country fixed effects	Late tracking and pre-school duration reduce impact of family background; inverted U-shaped effect of pre-school enrollment; no tradeoff with efficiency
Ammermueller (2005)	PISA, PIRLS	14	Student	Institutions and educational opportunities	Number of school types, instruction time, private school share, autonomy	Reading	Dependence of test scores on student background variables	Pooled data, WCRLR, DiD	Significant negative effect of number of school types and share of private schools on equality of opportunity; positive effect of instruction time
Brunello and Checchi (2007)	IALS, PISA 2003	17, 32	Student	School tracking and equality of opportunity	Age of first selection, length of tracking, share of vocational education	Competences and other indicators	Dependence of test scores on parental education; coefficient of variation in test scores	Cohort study, OLS, probit, multinomial logit	Mixed results; tracking reinforces family-background effects on formal education but weakens them on learning on the job
Waldinger (2006)	PISA +2003, TIMSS, PIRLS	8–14 (DiD), 29 (cross-section)	Student	Tracking and family background	Grade of first tracking	Math + reading	Dependence of test scores on parental background variables	Pooled cross-section data, DiD	Tracking does not increase impact of family background after controlling for pre-tracking differences; but small samples
Woessmann, Luedemann, Schuetz, and West (2009), ch. 7	PISA 2003	27	Student	Accountability, autonomy, and choice	Several measures of accountability, autonomy, choice, and tracking	Math	Dependence of test scores on PISA index of ESCS	Cross-section WCRLR	Public funding, private operation, and later tracking reduce impact of family background; accountability measures mostly equity-neutral
Schneeweis (2010)	TIMSS, TIMSS-R, +2003, PISA +2003	62, 167 country-years	Student/ Country	Educational institutions and integration of migrants	Ethnic segregation, pre-primary enrollment, school starting age, instruction time, external exams	Math + science	Unexplained test score gap of immigrants	Blinder-Oaxaca decomposition, pooled WLS, country fixed effects	Institutions account for 20% of immigrant disadvantage; esp. pre-primary education, young school starting age, low classroom segregation, instruction time

Note. ESCS = economic, social, and cultural status. WCRLR = weighted clustering-robust linear regression. OLS = ordinary least squares. DiD = differences in differences. See Tables 1 and 2 in Hanushek and Woessmann (2011) for acronyms of datasets.

Accountability

Analyses of the impact of curriculum-based external exit exam (CBEEE) systems illustrate the unique power of international production function estimates to address important policy-relevant issues.[3] By signaling student achievement to potential employers on the labor market and institutions of higher education, external school-leaving exams increase students' rewards for learning as well as parents' scope for monitoring the education process, so that they can be understood as an accountability device. (See Bishop, 2006 for a discussion of the underlying theoretical concepts).

Students in countries that have external exit-exam systems very consistently perform significantly and substantially better on the international student achievement tests than students in countries without external exit-exam systems (see Table 3). Using country-level data, John Bishop has shown this for the 1991 IAEP math, science, and geography tests (Bishop, 1995, section 4), the 1991 SIRS reading test (Bishop, 1999), the 1995 TIMSS math and science tests (Bishop (1997)), and the PISA 2000 reading, math, and science tests (Bishop, 2006, section 3). Microeconometric cross-country analyses that extensively control for family-background and school-input factors at the student level have confirmed this result for the 1995 TIMSS tests (Woessmann, 2001, 2003b), the 1999 TIMSS-Repeat tests (Woessmann, 2003a), the 2000 PISA tests (Woessmann, 2005c; Fuchs & Woessmann, 2007), and the 2003 PISA tests (Woessmann, Luedemann, Schuetz, & West 2009; see Table 1). 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, or between 20 and 40 percent of a standard deviation of the respective international tests.[4]

Beyond external exit exams, student achievement in PISA 2000 is also positively associated with teachers' monitoring of student progress by regular standardized tests (Fuchs & Woessmann, 2007). Richer data on additional accountability mechanisms available in PISA 2003 (documented in Table 1) reveal positive associations of student achievement with accountability measures aimed at teachers, such as internal and external monitoring of teacher lessons, and with accountability measures aimed at schools, such as assessments used to compare them to district or national achievement (Woessmann, Luedemann, Schuetz, & West, 2009).

Given the cross-sectional nature of identification, possible unobserved country heterogeneity related to the existence of external exit exams is a concern. To exclude the possibility that external exit exams just capture general cultural features of different world regions shows that results are robust to a regional fixed effects specification that controls for indicators of nine world regions. To ensure that the results do not

3 We concentrate on accountability for achievement that comes through exit exams, because understanding this topic requires analyses spanning jurisdictions with and without such institutions, making it a natural topic for use of international assessments. Of course, many analyses of accountability systems in general have proceeded within individual countries; see Hanushek and Raymond (2004) and Figlio and Loeb (2010).

4 Schneeweis (2010) finds that across countries, central exit exams are negatively related to the achievement gap between migrants and natives.

capture other features of centralization, results also prove robust to including controls for the centralization of school curricula and textbook approval, the share of central government financing, and ethno-linguistic fractionalization as a proxy for the homogeneity of a country's population.

Substantial cultural biases are also ruled out by the fact that the same positive association between central exams and student achievement is found within countries where some regions have external exam systems and other not. Such cross-regional studies exist for Canadian provinces (Bishop, 1997), German states (Jürges, Schneider, & Büchel, 2005; Woessmann, 2010b), and U.S. states (Bishop, Moriarty, & Mane, 2000; Woessmann, 2010b) 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. To probe causality further, Jürges, Schneider, and Büchel (2005) apply a differences-in-differences approach to the German TIMSS 1995 data that exploits the fact that, in some secondary-school tracks, the states with central exit exams have them in math but not in science, finding smaller but still substantial effects.[5]

Woessmann (2005c) exploits the student-level variation within each country to analyze whether external-exam effects are heterogeneous along several dimensions in quantile regressions and interacted specifications. Results using the TIMSS, TIMSS-Repeat, and PISA tests suggest that the effect tends to increase with student ability but does not differ with most family-background measures. It increases during the course of secondary education and with regular standardized examination. Furthermore, as discussed below, the effects of external exams are complementary to several dimensions of school autonomy.

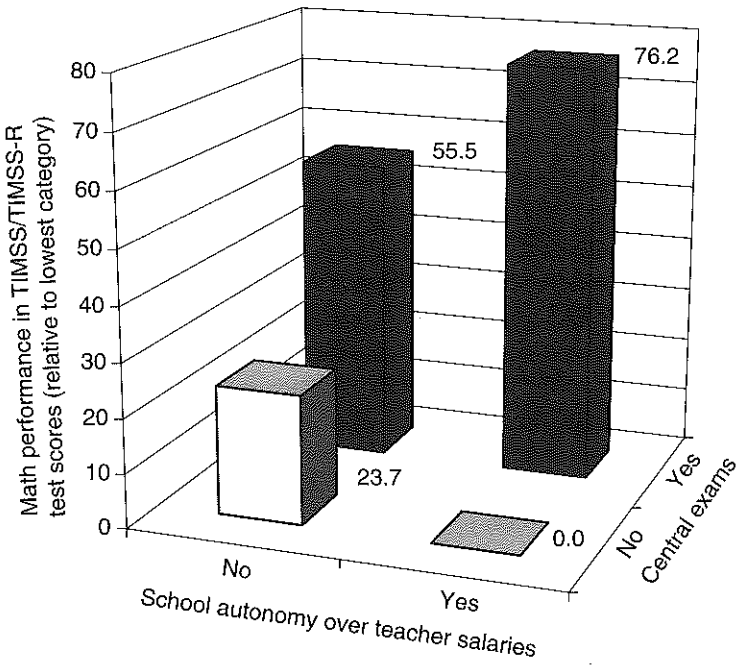
Autonomy

Another institutional feature that is sometimes argued to exert positive effects on student outcomes is school autonomy because local decision-makers tend to have superior information. On the other hand, in decision-making areas where their interests are not strictly aligned with improving student achievement, local decision-makers may act opportunistically unless they are held accountable for the achievement of their students (see Woessmann, 2005c for a discussion in a principal-agent framework).

The school background questionnaires of international tests allow deriving measures of school autonomy in several different decision-making areas. The general pattern of results (cf. Table 3) is that students perform significantly better in schools that have autonomy in process and personnel decisions (Woessmann, 2003b; Fuchs & Woessmann, 2007; Woessmann, Luedemann, Schuetz, & West, 2009). These decisions include such areas as deciding on the purchase of supplies and on budget allo-

5 This approach assumes that there are no spillovers between achievement in maths and in science. Jürges and Schneider (2010) find positive effects of central exit exams on student achievement, but negative effects on self-reported student attitudes toward maths, across German states.

Figure 1: External exams, school autonomy, and student achievement across countries



Note. Performance difference between the four categories relative to the lowest category which is set equal to zero. Based on a cross-country student-level multiple regression using the combined TIMSS and TIMSS-Repeat micro databases that extensively controls for family background, school inputs, and other institutional features.

Source: Woessmann (2005).

cations within schools, hiring and rewarding teachers (within a given budget), and choosing textbooks, instructional methods, and the like. Similarly, students perform better if their teachers have both incentives and the possibility to select appropriate teaching methods. By contrast, school autonomy in budget formation and teacher autonomy over the subject matter to be covered in class – two decision-making areas that are likely subject to substantial opportunism but little superior local knowledge – are negatively associated with student achievement.

The international evidence also points to a significant interaction between the effect of school autonomy with the extent of accountability in the school system (as previously found in Table 1). In some areas, autonomy is negatively associated with student achievement in systems that do not have external exit exams, but the association turns positive when combined with external-exam systems. Reflecting coefficient estimates from a student-level international education production function using the combined TIMSS and TIMSS-Repeat data, Figure 1 depicts school autonomy over teacher salaries as one such example. School autonomy over teacher salaries is *negatively* associated with student achievement in systems without external exams. However, in line with the arguments above, the average level of student achievement is

higher in systems with external exams. But what is more, the association between school autonomy and student achievement turns completely around in systems with external exams: Salary autonomy of schools is *positively* associated with student achievement in external-exam systems. The estimates in Figure 1 are expressed in percentages of a standard deviation on the international test scores, suggesting that the achievement difference between the best and worst institutional setting amounts to three quarters of a standard deviation, a huge effect compared to other educational interventions. Evidence from PISA 2000 corroborates this interaction pattern (Fuchs & Woessmann, 2007). Similar positive interactions between external exams and school autonomy have also been found for such decision-making areas as school autonomy in determining course content and teacher influence on resource funding, among others (see Woessmann, 2005c, for details).

In light of economic theory, this pattern of results is intuitively appealing. It indicates that local autonomy can lead to worse student outcomes if schools do not face incentives to focus attention on these outcomes. By contrast, when external exams hold schools accountable for student achievement, school autonomy leads to better outcomes. However, methodologically this empirical evidence on school autonomy is descriptive and awaits additional work that tries to more explicitly identify exogenous variation in school autonomy.

A recent study by Hanushek, Link, and Woessmann (2013) has made headway in this direction by constructing a panel dataset from the four waves of international PISA tests spanning 2000–2009, comprising over one million students in 42 countries. Relying on panel estimation with country fixed effects, they identify the effect of school autonomy from within-country changes in the average share of schools with autonomy over key elements of school operations. Their results show that autonomy affects student achievement negatively in developing and low-performing countries, but positively in developed and high-performing countries, and confirm that local decision-making works better when there is also external accountability that limits any opportunistic behavior of schools.

Competition from Private Schools

A third institutional feature that has been researched using international data is the relative performance of publicly and privately operated schools and the competition introduced by the latter. (For a general overview of school competition, see Hoxby, 2003 and Rouse & Barrow, 2009).

A first approach is to estimate differences in student achievement between public and private schools in each country, after controlling extensively for student and school background information. The PISA school background questionnaire provides specific school-level information on public versus private management and financing. Public school management is defined as schools managed directly or indirectly by a public education authority or governing board appointed by government or elected by public franchise, whereas private school management is defined as schools man-

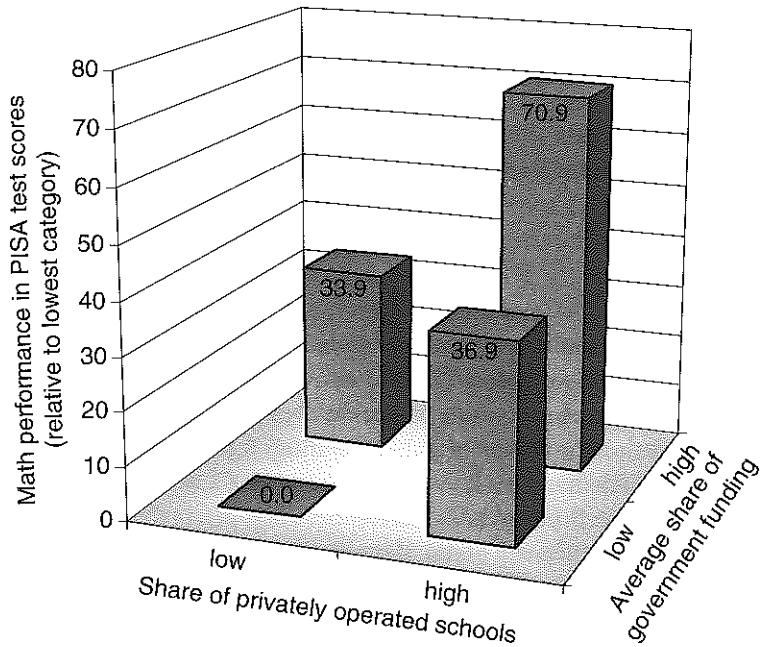
aged directly or indirectly by a non-government organization, for example churches, trade unions, or businesses. The share of public funding of each school is reported as the percentage of total school funding coming from government sources (at different levels), as opposed to such private contributions as fees and donations.

Looking across all countries (Table 2), private school management tends to be positively associated with student achievement, with a difference to publicly operated schools of 16–20 percent of an international standard deviation in the three subjects in PISA 2000 (Fuchs & Woessmann, 2007). A similar result is found in PISA 2003 (Woessmann, Luedemann, Schuetz, & West, 2009). The pattern is not uniform across countries, however, as revealed when estimating the effect within countries (Woessmann, 2009b). Toma (1996; see also 2005) similarly estimates the effect of private school operation in five countries using the 1981 SIMS, 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. Using the same database and distinguishing between privately operated schools that do and do not depend on government funding, Corten and Dronkers (2006) find a positive association of the achievement of students with low socio-economic status with private government-dependent schools, but no significant differences between public and private-independent schools. Dronkers and Robert (2008) find that the better performance of government-dependent private schools can be accounted for by a better school climate.

Using school-level variation of public-private operation in a pooled sample of countries, Woessmann, Luedemann, Schuetz, and West (2009) find positive interactions between private school operation and the average extent of autonomy that schools have in a country. Privately operated schools perform better if schools in the system are autonomous in formulating the budget and in staffing decisions, suggesting 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. Furthermore, they show that the association of student achievement with two measures of external accountability – the monitoring of teacher lessons by external inspectors and assessment-based comparisons of schools to national performance – is stronger in privately operated schools than in publicly operated schools. Private schools may thus benefit particularly from the accountability created by external inspection and performance comparisons with other schools.

Given the problem of non-random selection into private versus public schools within a country, these results based on micro-level variations within countries should be interpreted with caution. While many features of self-selection will be held constant by the extensive family-background controls that most of the studies contain, possible unobserved student heterogeneity may still raise concerns of selection bias. Because issues of self-selection cancel out at the country level, the cross-country estimation approach provides the possibility to address selection concerns by measuring private schooling as a share at the country level. In addition, in contrast to

Figure 2: Private operation, public funding, and student achievement across countries



Note. Performance difference between the four categories relative to the lowest category which is set equal to zero. Based on a cross-country student-level multiple regression using the PISA 2003 micro database that extensively controls for family background, school inputs, and other institutional features. "Low" and "high" refer to the 1st and 9th decile on the international distribution of the two variables (0% and 60% in the case of private operation and 55% and 100% in the case of government funding).

Source: Woessmann, Luedemann, Schuetz, and West (2009).

most within-country studies, studies that measure private-school shares at the country level are able to capture general-equilibrium effects that may arise from private competition. If the existence of private alternatives exerts competitive pressure on nearby public schools, both private and public schools may perform at a higher level due to larger private shares. Consequently, there may be important effects of private schools at the system level even if there is no performance difference between private and public schools at the school level.

Studies that include country-level measures of private school operation (Table 3) consistently find a strong positive association with student achievement (see Woessmann, 2003b for TIMSS 1995; Woessmann, 2009b for PISA 2000; and Woessmann, Luedemann, Schuetz, & West, 2009 for PISA 2003). At the same time, the measure of private funding shares available in PISA is negatively associated with student achievement. This pattern is depicted in Figure 2 and Table 1, which show that students in countries that combine relatively high shares of private operation with relatively high shares of public funding perform highest among the different operation-

funding combinations, while students in countries that combine public operation with private funding perform lowest. On average, the difference between the countries at the first and ninth decile on the international distribution – 60 percentage points in terms of private operation and 45 percentage points in terms of government funding – can account for roughly 0.35 standard deviations in educational achievement each.

The results point towards the importance of distinguishing between the operation and funding dimensions of private involvement. Without public funding, poor families may be constrained in their choices because they do not have the financial means to opt for private schooling. In this case, public funding may help families to exert their choices in terms of privately managed schools. The fact that public funding is positively associated with student achievement may thus also point to positive performance effects of school choice and competition. This line of reasoning is consistent with evidence in Woessmann (2009b) showing that, at the school level, the advantage of privately operated schools over publicly operated schools is particularly strong in countries with large shares of public funding. It is also in line with the finding of Woessmann, Luedemann, Schuetz, and West (2009), who show that students in countries where public funding is equalized between privately and publicly operated schools perform significantly better than students in countries where privately operated schools receive less government funding than publicly operated schools. Thus, a level playing field between public and private schools in terms of government funding may be an important ingredient for the competitive effects of private schools to emerge.

Beyond choice created by private schools, Woessmann, Luedemann, Schuetz, and West (2009) do not find significant associations on average of student achievement with 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. But within urban areas where there are schools to choose from, reduced local attendance and increased choice of better schools are associated with better student achievement. Using subnational regional variation in PISA 2003, Sprietsma (2008) finds a positive association of student achievement with the regional average of students reporting to attend their school because it is known to be a good school, which is interpreted as a measure of quality-based school choice.

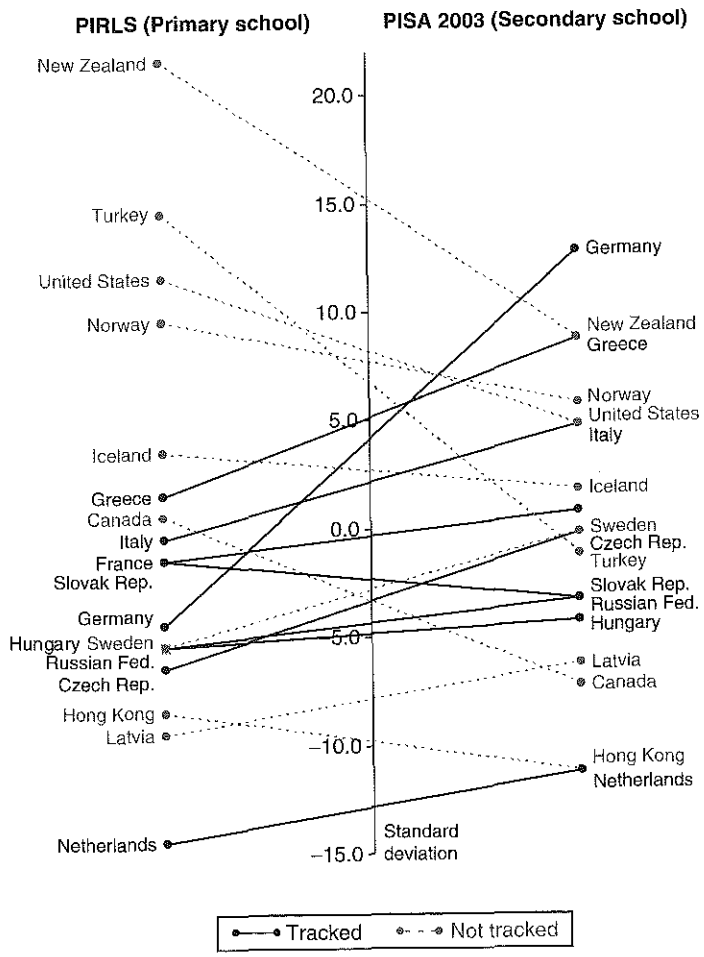
Combining German state-level data with data for OECD countries, Woessmann (2010b) shows that the association of private school shares with student achievement is not statistically different between the sample of German states and the sample of OECD countries. The result suggests that the international finding is not driven by major cultural differences between countries.

But there are additional challenges to causal identification of the effect of private competition. Omitted variables may be correlated with both the extent of private schooling and student achievement, such as factors related to the demand for private schooling or institutional or policy factors that affect its supply. Moreover, even well-controlled comparisons of countries or regions with small and large private sectors will be biased to the extent that low-quality public schools increase demand for private schooling as a substitute. To address these concerns, West and Woessmann (2010) develop an instrumental variable identification that exploits the fact that resistance of the Catholic Church to state schooling emerging in the 19th century has repercussions for the size of the private school sector today. This historical source of variation can be used as a natural experiment to identify exogenous variation in private school competition. The instrumental-variable specification uses the share of Catholics in 1900 – interacted with an indicator for Catholicism not being the state religion, as Catholics had no need to opt out of the state school system if the Church could control it – as an instrument for current private-school shares. The historical nature of the instrument allows controlling directly for any effect that the current Catholic share has on student achievement.

Estimating richly controlled student-level international education production functions on the PISA 2003 data, West and Woessmann (2010) confirm a significant positive effect of the share of privately operated schools on student achievement in math, science, and reading. The fact that the current share of Catholics, a control in some of their model, is negatively related to student achievement suggests that distinctive cultural features of traditionally Catholic countries are unlikely to be driving the results. The validity of the identification is additionally corroborated by the fact that Catholic shares are historically related to lower literacy and lower GDP per capita. To account for other possible channels through which the historical prevalence of Catholicism might be related to student outcomes today, the models also control for current GDP per capita and educational spending per student. Additional specification tests show that other current outcomes that might be conceived to be related to historical Catholicism, such as the decentralization of school policy decision-making, public social spending, and income inequality, are in fact uncorrelated with historical Catholic shares. West and Woessmann (2010) also show that much of the positive effect of private school shares accrues to students in public schools. This suggests that the overall effect is not simply due to privately operated schools being more effective, but rather reflects general-equilibrium effects of private competition. Finally, private competition is also found to reduce educational expenditure per student in the system, so that the better educational outcomes are obtained at lower cost.

As the overview in Table 4 shows, a topic that emerged only relatively recently in international literature is the question to what extent institutional features of the school systems can account for differences in the equity (rather than level) of student achievement across countries. A consistent pattern in this literature is that shares of privately operated schools and shares of public funding are not only associated with higher levels of student achievement but also with a reduced dependence of student achievement on socio-economic background. This has been shown both for the books-at-home indicator of family background in TIMSS and TIMSS-Repeat (Schuetz, Ursprung, & Woessmann, 2008) and for an index of socio-economic back-

Figure 3: Educational inequality in primary and secondary school



Note. Dispersion of student achievement measured as standard deviation of test scores in primary school (PIRLS) in comparison to lower-secondary school (PISA 2003), in both cases measured as difference to the international mean of national standard deviations in each test. The lines indicate the change in performance dispersion from primary to lower-secondary school. The dark solid lines indicate school systems that track their students into different school types before the age of 16, while the light dashed lines refer to those school systems that do not track their students by this age.

Source: Based on Hanushek and Woessmann (2006).

ground in PISA 2003 (Woessmann, Luedemann, Schuetz, & West, 2009, chapter 7).[6] In addition, Woessmann, Luedemann, Schuetz, and West (2009) find that a higher difference between private and public schools in the share of government funding is negatively associated not only with average student achievement, but also with equality of educational opportunity.

6 Ammermueller (2005) finds a negative association of the share of private schools with his measure of equality of opportunity, but this may be due to the fact that the model does not control for public versus private funding of schools.

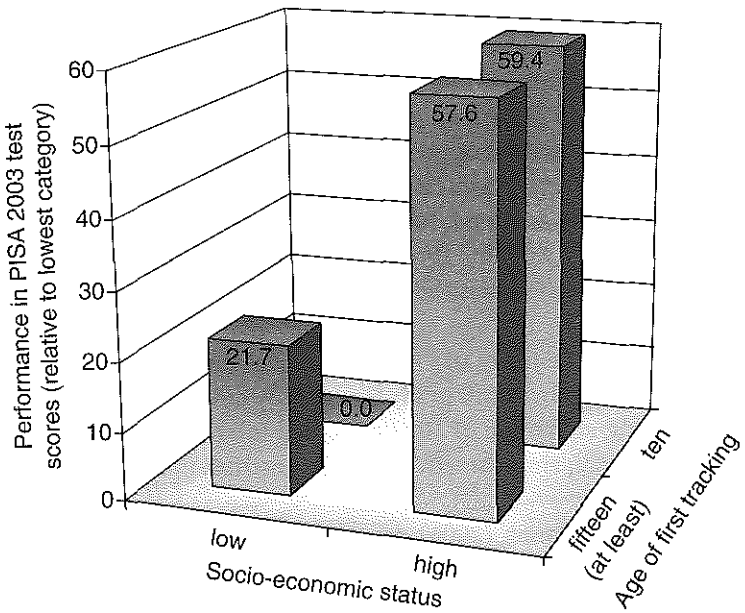
Tracking

Another institutional feature of school systems that has been discussed mostly in terms of the equity of student outcomes is tracking. Here, tracking is meant to refer to the placement of students into different school types, hierarchically structured by performance. Such school placement policies are variously called tracking, streaming, ability grouping, or selective (as opposed to comprehensive) schooling. From a theoretical viewpoint, the effects of educational tracking are controversial: Depending on the nature of peer effects assumed, homogeneous classes may contribute to optimal learning situations for all students through focused curricula and adequate progress, or weaker groups may be systematically disadvantaged if they are separated early on.[7] Countries differ widely in the age at which they first track children into different types of schools. In the majority of OECD countries, tracking takes place at the age of 15 or 16, with no tracking until grade 9 or 10. By contrast, some countries undertake the first tracking at the age of 10. Again, this international variation lends itself particularly well to analyzing the effects of the institutional feature of tracking (cf. Woessmann (2009a)).

Hanushek and Woessmann (2006) develop an international differences-in-differences approach to identify the causal effect of early tracking in a cross-country setting (cf. Table 4). The basic idea starts with the fact that, in *all* countries, students are taught in a uniform school type for the first four years of schooling. Therefore, a comparison of the *change* in educational inequality between 4th grade and the end of lower-secondary school between countries with and without early tracking can provide information on possible impacts of tracking. The analysis takes out the general level of inequality and considers only the change in inequality that occurs after 4th grade to determine the effect of early tracking. This method basically involves an investigation of the relationship depicted in Figure 3. The figure shows the inequality in reading achievement in 4th grade (in PIRLS) and at age 15 (in PISA 2003) for all countries that participated in both studies, measuring educational inequality by the standard deviation in student test scores. The essence of the analysis is to compare the *change* in inequality that occurs from primary to lower-secondary school between countries with and without educational tracking during this period. When looking at the change between the achievement dispersion in PIRLS and PISA, that part of the inequality measured at the end of lower-secondary school that already existed in 4th grade is eliminated. The change is indicated by the lines that connect the two points of each country. For countries with early tracking, solid connecting lines are used, while dashed lines indicate countries without early tracking. It is clearly visible that nearly all dark solid lines point upwards whereas nearly all light dashed lines point downwards: In countries with early tracking, inequality increases systematically, whereas it decreases in countries without tracking.

7 Here we concentrate entirely on tracking that occurs between schools, i.e., where children are sorted into separate schools. Many countries of the world, including the U.S., pursue tracking within schools but not generally across schools. For more on within-school tracking, see Betts (2010).

Figure 4: Tracking and socio-economic status in PISA



Note. Performance difference between the four categories relative to the lowest category which is set equal to zero. Based on a cross-country student-level multiple regression using the PISA 2003 micro database that extensively controls for family background, school inputs, and other institutional features. Low and high socio-economic status correspond to the 25th and 75th percentile of the PISA ESCS index, respectively.
Source: Based on Woessmann, Luedemann, Schuetz, and West (2009).

Hanushek and Woessmann (2006) confirm this graphic depiction in country-level econometric estimates based on a differences-in-differences approach: The difference between countries with and without early tracking is investigated in terms of the difference in inequality between primary and lower-secondary school. The results show that early tracking systematically increases the inequality of student achievement. In total, their analyses take into account eight pairs of tests in primary and secondary schools, combining a total of 176 country observations. In contrast to the results on inequality, the results on achievement levels are less clear. But there is little evidence that early tracking increases the achievement level. To the contrary, in the most comprehensive model there is a marginally significant negative effect of early tracking on the average achievement level. When evaluating achievement at different percentiles of the performance distribution, not even for the best 5 percent of students is there a positive effect of early tracking.

While this investigation considers the dispersion of student achievement, Schuetz, Ursprung, and Woessmann (2008) investigate the more direct measure of inequality of opportunity outlined above: the extent to which individual student achievement depends on the family background of the student. At a more descriptive level, the

effect of early tracking on equity is identified by the interaction of the country-level measure of early tracking with the student-level measure of family background in a student-level model with country fixed effects. Their measure of inequality of opportunity is found to be significantly smaller, the later the tracking age of students. If tracking is postponed by four years, for example, the impact of family background on student achievement is smaller by one quarter of the entire impact of the family background averaged across the OECD countries. In a model without country fixed effects, the association between early tracking and the average achievement level is statistically insignificant and negative.

The same association between tracking and equality of opportunity is found in a related study using PISA 2003 data (Woessmann, Luedemann, Schuetz, and West (2009), chapter 7). Using the Index of Economic, Social, and Cultural Status (ESCS) provided by the PISA study as an alternative measure for family background, the qualitative results are the same: The association between test scores and family background is significantly smaller, the higher the age of first tracking. This association is depicted in Figure 4: In countries with earlier tracking, the achievement difference between children with different socio-economic backgrounds is considerably larger. As the figure reveals, this effect arises primarily from the fact that children with low socio-economic status in countries without early tracking perform considerably better. At the same time, children from families with a relatively high socio-economic status perform at approximately the same level. Accordingly, the effect of later tracking on the average achievement level is again positive, albeit not statistically significant.

In terms of the equity effects, Ammermueller (2005) reports similar results for the number of school types (rather than the age of first tracking) based on the international PIRLS and PISA data. Waldinger (2006) uses a combination of the approach of Hanushek and Woessmann (2006) and Schuetz, Ursprung, and Woessmann (2008) and tends to find statistically insignificant results, but this may be largely due to limited degrees of freedom in samples of only 8–14 countries and a less informative tracking measure. Brunello and Checchi (2007) use the international approach described here for results beyond school age, finding that tracking increases the effect of family background on earnings in the labor market. Using a similar approach, Schneeweis (2010) finds some indication that an index of between-school social segregation, presumably partly caused by early tracking, is positively associated with the achievement gap between migrants and natives across countries.

Using system-level data, Woessmann (2010b) pools German states (most of which track after 4th grade, but some of which track after 6th grade) with OECD countries in a sample of 42–54 observations. Results indicate that the negative association between early tracking and the measure of inequality of educational opportunity (the slope of the socio-economic gradient) is statistically indistinguishable between the sample of German states and the sample of OECD countries. This shows that the cross-country association cannot be accounted for by such country-level omitted factors as differences in culture, language, or legal background.

Pre-primary Education System

The fact that student achievement is strongly associated with family background is suggestive of the idea that learning in the formative years before formal schooling is important for ultimate academic achievement. Consequently, student achievement towards the end of compulsory school is not only related to features of the school system, but also to pre-school education (see the conceptual discussion in Cunha, Heckman, Lochner, & Masterov, 2006 and Blau & Currie, 2006 for a general review of the empirical literature). In line with this reasoning, Schuetz, Ursprung, and Woessmann (2008) find a positive association of student achievement in 8th grade (in TIMSS and TIMSS-Repeat) with the usual duration of the pre-school cycle in a country.

Schuetz (2009) uses the individual-level information on pre-primary attendance available in PISA 2003 to show that pre-primary attendance is positively associated with achievement at age 15 in most countries. She goes on to exploit the fact that the size of this association varies substantially across countries. Estimating a cross-country student-level specification with country fixed effects and interactions between individual pre-primary attendance and country-level indicators of the quality of pre-primary education, she finds that the achievement gap between students who did and did not attend pre-primary education is positively associated with country-level measures of per-student spending in pre-primary education, of the share of pre-primary institutions being privately operated, and of the training level and relative pay of pre-primary teachers. Thus, indicators of institutions and structural quality of pre-primary education systems can account for variation in the estimated coefficients on pre-primary attendance across countries. While the study is descriptive in the sense that individual pre-primary attendance may not be conditionally random, the estimators of interest are unbiased by selection decisions as long as the selection process is the same in all countries. In this sense, under the assumption that that enrollment in pre-primary education follows the same rules in all countries, interacted specification can be interpreted as an international differences-in differences approach.

In terms of equity, using the interacted specification described above, Schuetz, Ursprung, and Woessmann (2008) show that the association between socio-economic background and 8th-grade student achievement is negatively related to the duration of pre-primary education in a country. Furthermore, the thus measured equality of educational opportunity follows an inverted U-shaped relationship with pre-school enrollment: The association between the books-at-home measure of socio-economic background and TIMSS student achievement increases up to a pre-primary enrollment rate of 60 percent in a country and decrease beyond that threshold. This result pattern may indicate that, initially, children who are otherwise advantaged attend pre-primary education. Only once most of a country's children attend pre-primary institutions does pre-primary attendance increase equality of educational opportunity for children from lower socio-economic backgrounds. Using a similar approach to focus on equality of educational opportunity between native and migrant children, Schneeweis (2010) finds that the migrant achievement gap is negatively associated with pre-primary enrollment across countries.

Conclusions and Outlook

The economic literature on international educational achievement has expanded our understanding of the determinants and economic consequences of international educational achievement tremendously. Considering that, with few antecedents, this literature is only a decade old, it has clearly covered a lot of ground and made remarkable progress. In doing so, it has exploited the possibilities opened up by the international data to raise fundamental questions, ones not amenable to any simple within-country analysis. For example, this work highlights the crucial role of educational achievement in understanding the vast international differences in economic well-being. It also begins to suggest some key factors that account for the immense international differences in educational achievement. At the same time, by the very nature of the limited degrees of freedom in cross-country identification, it can mostly reveal broad patterns. At the very least, a lot of the details of specific implementation issues related to any policy application obviously must be left for national approaches.

The economic literature on determinants of international differences in educational achievement has applied two main approaches. The first approach exploits the cross-country variation for identification of cross-country associations. The second approach estimates the same association within different countries in order to enhance understanding of whether a factor's importance differs systematically in different settings. Part of the existing work is descriptive in nature, estimating the association of student achievement with certain factors after controlling for the rich set of possible inputs into educational production available in the international background data. But quasi-experimental work has been developed to identify some of the underlying causal mechanisms both in the cross-country and in the within-country approach.

On family background and school inputs, the international results tend to mirror the existing national evidence on educational production. Many dimensions of students' family background are important factors for their educational achievement. At the same time, it is hard to find evidence of substantial positive effects of most resource inputs, in particular class sizes and expenditure levels. Among school inputs, there is somewhat more indication of positive effects of measures capturing teacher quality, such as (in an international setting) teacher education. A particular opportunity of the international research is that it can unveil whether certain effects differ systematically across countries, such as class-size effects or the equality of educational achievement for students with different family or migration backgrounds.

The second particular opportunity of the international research is the substantial institutional variation that exists across countries. The international evidence on education production functions suggests that schools matter for student outcomes,

but not so much in terms of traditional inputs. Instead, the impact of schools comes through teacher quality and institutional structures that determine incentives. Institutional features of school systems can account for a substantial part of the cross-country variation in student achievement. In the school system, institutions that tend to be associated with higher achievement levels include accountability measures like external exit exams, school autonomy in process and personnel decisions (if combined with accountability), private-school competition, and public financing. Later tracking, public funding, and private operation are systematically related to the equality of student outcomes. While some of the evidence is descriptive, convincing causal identification has been developed that supports the results on external exit exams, private-school competition, and tracking. Also beyond the school system, institutions of the pre-school and post-school education systems are related to international differences in educational achievement. In particular, more extensive pre-primary education systems relate to more equalized student achievement for children from different family and migration backgrounds, and measures of pre-primary quality tend to be related to the size of the pre-primary effect.

Given its infancy, there is obviously still considerable scope for future advances in the economic literature on international evidence on educational achievement. A topic unexplored by economists is the international tests in non-traditional subjects, such as foreign languages, civic education, and information technology. More generally, some of the rich background information contained in the international studies could be explored further, and part of it may provide information on relevant non-cognitive skills. For example, recent work by Falck and Woessmann (2013) attempts to derive measures of entrepreneurial intentions from the international background data, and chapter 6 in Woessmann, Luedemann, Schuetz, and West (2009) explores such measures of non-cognitive outcomes as student morale and commitment, non-disruptive behavior, disciplinary climate, and tardiness. Further information on non-cognitive skills may be derived from international background questionnaires. As a more distant outlook, international testing of non-cognitive skills would be an obvious challenge.

As more and more countries participate in international tests, the opportunities grow for future research on the determinants of international educational achievement. With the additional variation, the international research will be able to draw on more experience with different institutions and start to analyze additional specific features beyond the broad concepts of institutional structures analyzed so far. There is also considerable scope for future research to advance identification in quasi-experimental research settings. Furthermore, as more regular tests with reasonable comparability over time become available, a panel structure of international tests emerges that provides longitudinal information within countries. This will allow future research to exploit educational reforms in different countries over time (see Hanushek et al., 2013 for a recent first example). A limiting factor remains the lack of individual-level panel data in the international tests.

In the more distant future, it is tempting to envision what research will be able to do with the sort of achievement data that will be available in 20 to 30 years from

now. The number of participating countries is as high as 63 in TIMSS 2011 and 64 in PISA 2012, and additional countries are planning to participate in future cycles. With these sets of comparable achievement data for extensive samples of countries being linked to subsequent economic growth, and with the emerging long panels of regular achievement data for large samples of countries, the outlook for future research in the economics of international differences in educational achievement is clearly bright.

References

Ammermueller, A. (2013). Institutional features of schooling systems and educational inequality: Cross-country evidence from PIRLS and PISA. *German Economic Review, 14* (2), 190–213. doi: 10.1111/j.1468-0475.2012.00565.x.

Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 341–381). Amsterdam: North Holland.

Bishop, J. H. (1995). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Educational Research, 23* (8), 653–752.

Bishop, J. H. (1997). The effect of national standards and curriculum-based examinations on achievement. *American Economic Review, 87* (2), 260–264.

Bishop, J. H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review, 6* (2), 349–398.

Bishop, J. H. (2006). Drinking from the fountain of knowledge: Student incentive to study and learn – Externalities, information problems, and peer pressure. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 909–944). Amsterdam: North Holland.

Bishop, J. H., Moriarty, J. Y., & Mane, F. (2000). Diplomas for learning, not seat time: The impact of New York Regents examinations. *Economics of Education Review, 19* (4), 333–349.

Blau, D. M. & Currie, J. (2006). Pre-school, day care, and after-school care: Who's minding the kids? In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education*, (Vol. 2, pp. 1163–1278). Amsterdam: North-Holland.

Brunello, G. & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy, 22* (52), 781–861.

Corten, R. & Dronkers, J. (2006). School achievement of pupils from the lower strata in public, private government-dependent and private government-independent schools: A cross-national test of the Coleman-Hoffer thesis. *Educational Research and Evaluation, 12* (2), 179–208.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 697–812). Amsterdam: Elsevier.

Dronkers, J. & Robert, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools: A cross-national analysis. *Educational Policy, 22* (4), 541–577.

Falck, O. & Woessmann, L. (2013). School competition and students' entrepreneurial intentions: International evidence using historical Catholic roots of private schooling. *Small Business Economics, 40* (2), 459–478. doi: 10.1007/s11187-011-9390-z.

Figlio, D. & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 383–421). Amsterdam: North Holland.

Foshay, A. W. (1962). The background and the procedures of the twelve-country study. In A. W. Foshay, R. L. Thorndike, F. Hotyat, D. A. Pidgeon & D. A. Walker (Eds.), *Educational achievement of thirteen-year-olds in twelve countries: Results of an international research project, 1959–61*. Hamburg: Unesco Institute for Education.

Fuchs, T. & Woessmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics, 32* (2–3), 433–462.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14* (3), 351–388.

Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics*, (Vol. 4, pp. 2045–2141). Amsterdam: North Holland.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review, 90* (5), 1184–1208.

Hanushek, E. A., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, forthcoming.

Hanushek, E. A. & Raymond, M. E. (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, E. A. & Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal, 116* (510), C63-C76.

Hanushek, E. A. & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 89–200). Amsterdam: North Holland.

Heyneman, S. P. & Loxley, W. (1983). The effect of primary school quality on academic achievement across twenty-nine high and low income countries. *American Journal of Sociology, 88* (6), 1162–1194.

Hoxby, C. M. (2003). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review, 10* (3), 9–65.

Jürges, H. & Schneider, K. (2010). Central exit examinations increase performance... but take the fun out of mathematics. *Journal of Population Economics, 23* (2), 497–517.

Jürges, H. Schneider, K., & Büchel, F. (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.

Lee, J.-W. & Barro, R. J. (2001). Schooling quality in a cross-section of countries. *Economica, 68* (272), 465–488.

Rouse, C. E. & Barrow, L. (2009). School vouchers and student achievement: Recent evidence and remaining questions. *Annual Review of Economics, 1*, 17–42.

Schneeweis, N. (2011). Educational institutions and the integration of migrants. *Journal of Population Economics, 24* (4), 1281–1308.

Schuetz, G. (2009). Does the quality of pre-primary education pay off in secondary school? An international comparison using PISA 2003 *Ifo Working Paper*. Munich: ifo Institute for Economic Research at the University of Munich.

Schuetz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos, 61* (2), 279–308.

Sprietsma, M. (2008). Regional school choice and school selectivity: How do they relate to student performance? Evidence from PISA 2003. *European Journal of Comparative Economics, 5* (2), 133–156.

Todd, P. E. & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal, 113* (485), F3-33.

Toma, E. F. (1996). Public funding and private schooling across countries. *Journal of Law and Economics, 39* (1), 121–148.

Toma, E. F. (2005). Private schools in a global world: 2004 presidential address. *Southern Economic Journal, 71* (4), 693–704.

Vandenberghe, V. & Robin, S. (2004). Evaluating the effectiveness of private education across countries: A comparison of methods. *Labour Economics, 11* (4), 487–506.

Waldinger, F. (2006). *Does tracking affect the importance of family background on students' test scores?* London: London School of Economics.

West, M. R. & Woessmann, L. (2010). 'Every Catholic child in a Catholic school': Historical resistance to state schooling, contemporary private competition and student achievement across countries. *Economic Journal, 120* (546), F229–F255.

Woessmann, L. (2001). Why students in some countries do better: International evidence on the importance of education policy. *Education Matters, 1* (2), 67–74.

Woessmann, L. (2003a). Central exit exams and student achievement: International evidence. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 292–323). Washington, D.C.: Brookings Institution Press.

Woessmann, L. (2003b). Schooling resources, educational institutions, and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics, 65* (2), 117–170.

Woessmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics, 13* (2), 143–169.

Woessmann, L. (2007). International evidence on school competition, autonomy and accountability: A review. *Peabody Journal of Education, 82* (2–3), 473–497.

Woessmann, L. (2009a). International evidence on school tracking: A review. *CESifo DICE Report – Journal for Institutional Comparisons, 7* (1), 26–34.

Woessmann, L. (2009b). Public-private partnerships and student achievement: A cross-country analysis. In R. Chakrabarti & P. E. Peterson (Eds.), *School choice international: Exploring public-private partnerships* (pp. 13–45). Cambridge, MA: MIT Press.

Woessmann, L. (2010). Institutional determinants of school efficiency and equity: German states as a microcosm for OECD countries. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics, 230* (2), 234–270.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.