Chapter 8
School Resources and Student Achievement: A Review of Cross-Country Economic Research

Eric A. Hanushek and Ludger Woessmann

Abstract How do school resources affect students' academic achievement? This chapter provides a survey of economists' work on the effect of expenditure and class size on student achievement using different international student achievement tests, with a particular focus on the use of quasi-experimental research methods to address challenges of the identification of causal effects. Overall, the international evidence provides little confidence that quantitative measures of expenditure and class size are a major driver of student achievement, across and within countries. The cross-country pattern suggests that class size is a relevant variable only in settings with low teacher quality. Among other school inputs, descriptive evidence suggests that measures of the quality of inputs and, in particular, teachers are more closely related to student outcomes.

8.1 Introduction

How do school resources affect students' academic achievement? A lot of work on this question has emerged since Jan-Eric Gustafsson (2003) reviewed the literature. In particular, much research has used data from international student achievement

This is a slightly extended version of 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.

E.A. Hanushek
Stanford University, NBER, and CESifo, Stanford, USA
e-mail: hanushek@stanford.edu

L. Woessmann (✉)
University of Munich, Ifo Institute, CESifo, and IZA, Munich, Germany
e-mail: woessmann@ifo.de

M. Rosén et al. (eds.), *Cognitive Abilities and Educational Outcomes*,
Methodology of Educational Measurement and Assessment,
DOI 10.1007/978-3-319-43473-5_8

tests to shed new light on the question. Much of this research, in particular from economists working in the field, has focused on challenges of the identification of causal effects by using quasi-experimental research methods. Some of the recent research is in line with the suggestion by Gustafsson (2007) that important analysis could come from changes in the performance of different countries over time. This chapter provides a survey of economists' work on the effect of expenditure and class size on student achievement using different international student achievement tests. Part of this research focusses on the challenge of overcoming possible bias in cross-country estimation, part on the identification of causal effects within countries.

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.

In the mid-1960s, international consortia started 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 (at each occasion) a common set of test questions in all participating countries. By 2016, 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 Literacy Study (PIRLS) testing primary-school reading performance on a five-year cycle since 2001.

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 school resources as one group of determinants of international educational achievement, covering both evidence across countries and evidence within different countries. For research on student background and institutional structures of the education system as two other groups of possible determinants, see Sects. 4.2 and 4.4 in Hanushek and Woessmann (2011a). For the second line of research, see Sect. 5 in Hanushek and Woessmann (2011a), as well as Hanushek and Woessmann (2015). Furthermore, Sects. 1–3 in Hanushek and Woessmann (2011a) provide a more detailed discussion of the unique advantages of and concerns with the use of cross-country data, a brief economic motivation to frame the discussions and an overview and critical assessment of the different available international datasets on educational achievement.

The cross-country comparative approach provides a number of unique advantages over national studies: It can exploit institutional variation that does not exist within countries; draw on much larger variation than is usually available within any country; reveal whether any result is country-specific or more general; test whether effects are systematically heterogeneous in different settings; circumvent selection issues that plague within-country identification by using system-level aggregated measures; and uncover general-equilibrium effects that often elude studies in a single country. The advantages come at the price of concerns about the limited number of country observations, the cross-sectional character of most available achievement data, and possible bias from unobserved country factors like culture.

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.

8.2 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 and 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 choices of parents, administrators, and schools that are correlated with the error term of the production function. 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. 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 well-equipped school may differ along both observable and unobservable dimensions from students taught in poorly equipped 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 input 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 average expenditure per student in a country 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 Heyneman and Loxley (1983) and Toma (1996). Early studies using the cross-country variation of international tests to estimate international education production functions 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 estimate extensive multivariate cross-country education production functions is Woessmann (2003). 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 8.1 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

[1]See Woessmann et al. (2009) for additional background and robustness analyses.

Table 8.1 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)

Table 8.1 (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	

Notes: 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 et al. (2009), who provide additional background details.

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 the 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 % of the achievement variation at the student level and for 87 % 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. Here, we focus on one specific group of determinants, namely school inputs. Sections 4.2 and 4.4 in Hanushek and Woessmann (2011a) discuss the literature on the other two groups of determinants—student and family background, as well as institutional structures of the education system—in greater detail.

[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 (2011a) for more extensive discussion.

8.3 Evidence on School Inputs Across Countries

We start with research that uses variation in school inputs across countries.[3] The studies reveal that in general, the cross-country association of student achievement with resources tends to be relatively weak.

When looking across countries, the most straightforward starting point is the simple association between the aggregate financial measure of average expenditure per student and average achievement. Figure 8.1 presents the international association between cumulative spending per student from age 6 to 15 and the average math achievement of 15-year-olds on the 2003 PISA test. Without considering the strong outliers of Mexico and Greece, there is no association between spending levels and average achievement across countries.[4] At the most basic level, countries

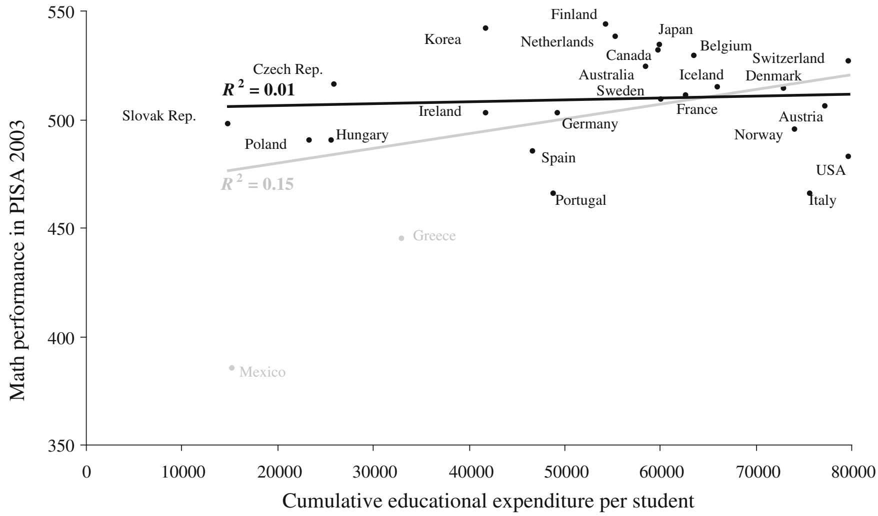

Fig. 8.1 Expenditure per student and student achievement across countries. *Notes* Association between average math achievement in PISA 2003 and cumulative expenditure on educational institutions per student between age 6 and 15, in US dollars, converted by purchasing power parities. *Dark line* regression line for full sample. *Light line* regression line omitting Mexico and Greece. *Source* Woessmann (2007)

[3]For a general overview of such studies see Table 2.6 in Hanushek and Woessmann (2011a).

[4]With the two outliers, there is a weak positive association as long as other effects are ignored. Taken literally, the full-sample association suggests that $60,000 per student in additional expenditure (a quadrupling of spending in the low spending countries) is associated with about a half standard deviation improvement in scores. However, once a country's GDP per capita is controlled for, the cross-country association between student achievement and expenditure loses statistical significance and even turns negative, suggesting that the bivariate association is driven by the omitted factor of average socio-economic status.

with high educational spending appear to perform at the same level as countries with low expenditures.

This picture has been evident in many other waves of the different international achievement tests (e.g., Woessmann (2002), Sect. 3.2, for the 1995 TIMSS test). Furthermore, in most cases the lack of a significant positive cross-country association between expenditure per student and educational achievement holds up when numerous other determining factors such as family background and school features (including instruction time) are accounted for in a regression framework. Hanushek and Kimko (2000) and Lee and Barro (2001) perform country-level regressions using different tests and Woessmann (2003) and Fuchs and Woessmann (2007) perform student-level microeconometric regressions using TIMSS 1995 and PISA 2000, respectively.

As discussed above, such cross-sectional analysis has to be interpreted cautiously, even when controlling for a large set of factors. There may be reverse causality, and unobserved country differences—e.g., cultural traits or institutional and political factors—may be correlated with both inputs and outcomes. As a first step to address such worries, one can look at within-country variation over time. By looking at changes in inputs and outcomes, one can rule out unobserved level effects. Thus, Gundlach et al. (2001) calculate changes in expenditure and achievement for individual OECD countries from 1970 t o 1994, and Gundlach and Woessmann (2001) for individual East Asian countries from 1980 to 1994.[5]

The results, depicted in Fig. 8.2, suggest that educational expenditure per student has increased substantially in real terms in all considered OECD countries between the early 1970s and the mid-1990s, and in all considered East Asian countries except the Philippines between the early 1980s and the mid-1990s.[6] Yet, comparing test scores over the same time intervals suggests that no substantial improvement in average student achievement has occurred in any of these countries. Combining the time-series evidence on resources and achievement, it is fair to conclude that substantial increases in real school expenditure per student did not lead to improvements in student outcomes in most of the sampled OECD and East Asian countries. In fact, the experience of many countries is much bleaker than what had been termed the “productivity collapse in schools” in the United States (Hanushek 1997).[7]

[5]Achievement data from the international tests at the two respective points in time are linked using U.S. longitudinal achievement data. Increases in educational expenditure are adjusted not only for average inflation, but also for the so-called “Baumol effect” of increasing costs in service sectors with constant productivity. Three different approaches of calculating price deflators for the schooling sector that account for this effect are averaged in the depiction of Fig. 2. For details, see Gundlach et al. (2001), Gundlach and Woessmann (2001), and Woessmann (2002), Sect. 3.3.

[6]Gundlach and Woessmann (2001) show that the resource expansion in the East Asian countries mostly results from government decisions to raise the number of teachers per student.

[7]One potential explanation for this bivariate longitudinal pattern might of course be that students’ family background might have deteriorated on average. Students may increasingly be lacking many of the basic capabilities required for a successful education and may thus be increasingly expensive to educate. Such effects may play a significant role in countries with a large inflow of

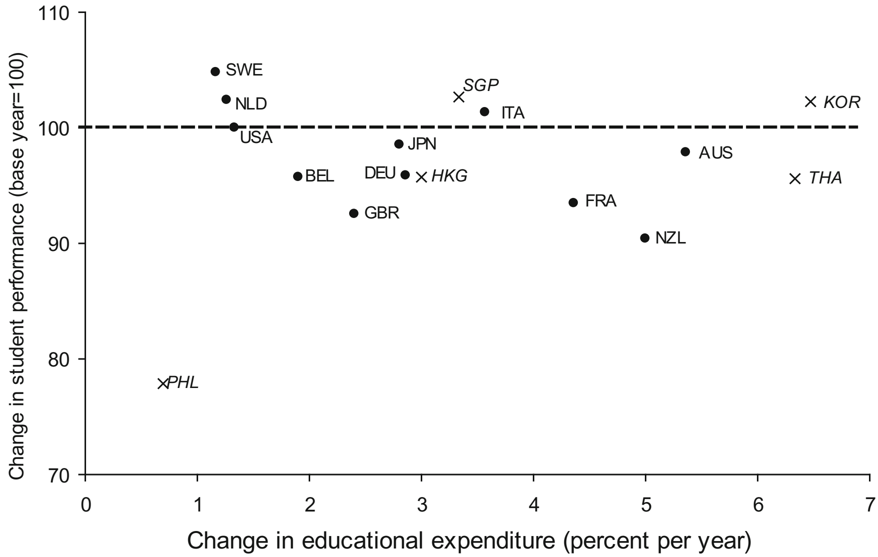

Fig. 8.2 Change in expenditure per student and in student achievement over time. *Notes* Data for OECD countries refer to 1970–1994, data for East Asian countries to 1980–1994. Change in student performance: students' average educational performance in math and science in 1994 relative to base year. Change in educational expenditure: average annual rate of change in real educational expenditure per student in percent. Country abbreviations: Australia (*AUS*), Belgium (*BEL*), France (*FRA*), Germany (*DEU*), Hong Kong (*HKG*), Italy (*ITA*), Japan (*JPN*), Netherlands (*NLD*), Philippines (*PHL*), Singapore (*SGP*), South Korea (*KOR*), Sweden (*SWE*), Thailand (*THA*), United Kingdom (*GBR*), United States (*USA*). *Source* Based on Gundlach et al. (2001) and Gundlach and Woessmann (2001)

More recently, the linking of the PISA tests over time allows for a direct comparison of spending changes to changes in achievement on psychometrically linked tests. As is directly obvious from Fig. 8.3, changes in PISA performance from 2000 to 2012 are not systematically related to concurrent changes in expenditure per student (Hanushek and Woessmann 2015). Countries with large spending increases do not show different achievement trends from countries that spend only little more. The coefficient estimate on expenditure in the simple underlying

(Footnote 7 continued)

immigrant students or with rising levels of poverty. But on average, parents in the considered countries have been enjoying higher incomes and better education over time, and the number of children per family has declined. Hence by the later periods, children may actually start schooling with better basic capabilities than before. These issues, however, await thorough econometric analysis.

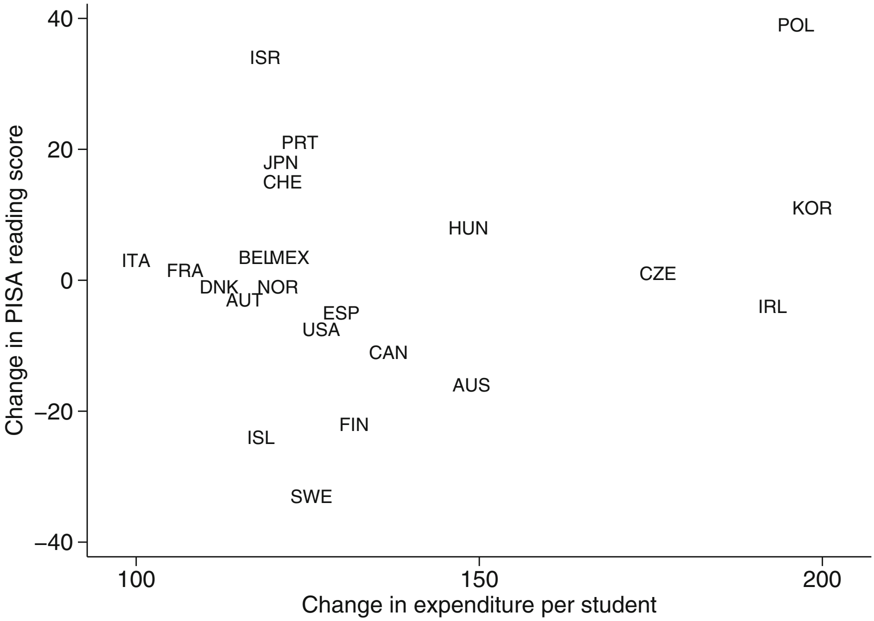

Fig. 8.3 Changes in educational spending and in student achievement across countries. *Notes* Scatter plot of the change in expenditure per student, 2000–2010 (constant prices, 2000 = 100) against change in PISA reading score, 2000–2012. *Source* Hanushek and Woessmann (2015) based on OECD data

first-differenced regression is insignificant, and without the apparent outlier Poland, the point estimate is negative.[8]

Apart from the aggregate expenditure measure, the cross-country variation has also been used to analyze specific resource inputs in cross-sectional analysis. Expenditure per student is an encompassing measure of school inputs which considers not only personnel costs but also material costs. But international comparisons of expenditure may be hampered by the problem of choosing an appropriate exchange rate (Fig. 8.1 uses conversion by purchasing power parities). Because personnel costs make up more than three quarters of total expenditure in nearly all countries, class size lends itself particularly well as a non-monetary input measure for international comparisons which determines a large part of total expenditure. However, using class size instead of expenditure per student yields the same general picture as in Fig. 8.1. Regression analyses that control for family background measures come to similar results. At the country level, Lee and Barro (2001) find a positive effect of smaller student-teacher ratios, but Hanushek and Kimko (2000)

[8]Similarly, using data from the first three PISA waves, the working-paper version of Hanushek and Woessmann (2011b) reports insignificant negative coefficient estimates on expenditure per student in first-differenced and fixed-effects models.

find no such relationship.[9] However, country-level analysis may suffer from aggregation bias (Hanushek et al. 1996), as Fertig and Wright (2005) show that the probability of finding statistically significant and correctly signed class-size effects increases with the level of aggregation. Student-level analyses that use data on the actual size of the class of the tested students, rather than ratios of teachers to students at some level, tend to find counterintuitive signs of the coefficient on class size that are often statistically significant (e.g., Woessmann 2003; Fuchs and Woessmann 2007; Table 8.1).

The latter studies also take indicators of the shortage of instructional material, usually reported by school principals, into account. Shortage of material tends to be negatively associated with student outcomes. Measures of instruction time also tend to be significantly related to achievement. By contrast, in multivariate analyses the availability of computers at school is not related to student outcomes, and intensive computer use is negatively related to test scores (Fuchs and Woessmann 2004).

In the student-level studies, measures of teacher education tend to show positive associations with student achievement in cross-country analyses. Drawing on information from teacher background questionnaires in TIMSS, Woessmann (2003) finds positive associations of student achievement with teacher experience and female gender and a negative one with teacher age. In their country-level analysis, Lee and Barro (2001) find a positive effect of teacher salary levels. Similarly, Woessmann (2005b) reports a significant positive coefficient on a country-level measure of teacher salary when added to an international student-level regression. Dolton and Marcenaro-Gutierrez (2011) pool country-level data from international tests in 1995–2006 to show that teacher salaries—both when measured in absolute terms and relative to wages in each country—are positively associated with student achievement, even after controlling for country fixed effects.

In sum, the general pattern of the cross-country analyses suggests that quantitative measures of school inputs such as expenditure and class size cannot account for the cross-country variation in educational achievement. By contrast, several studies tend to find positive associations of student achievement with the quality of instructional material and the quality of the teaching force. While these cross-country associations reveal to what extent different input factors can descriptively account for international differences in student achievement, studies that focus more closely on the identification of causal effects have reverted to using the within-country variation in resources and achievement. This literature is most advanced for the estimation of class-size effects. In the following, we discuss three approaches that have been suggested to estimate causal class-size effects on international data: a combination of school fixed effects with instrumental variables, a regression discontinuity approach that makes use of variation stemming from maximum class-size rules, and a subject fixed effects approach.

[9]Using country-level data for data envelopment analysis, Afonso and St. Aubyn (2006) find indications of substantial inefficiencies in the use of teachers per student in most countries.

8.4 Evidence on School Inputs Within Different Countries

The initial within-country studies have used conventional least-squares techniques to focus on developing countries and their comparison to developed countries, a particular advantage of using international data.[10] Relying on data from early international tests, Heyneman and Loxley (1983) suggested that school resources tend to be more closely related to student achievement in developing countries than in developed countries. Hanushek and Luque (2003) did not corroborate this conclusion using the more recent TIMSS data. Michaelowa (2001) uses the regional PASEC data to provide conventional evidence for five countries in Francophone Sub-Saharan Africa.[11]

The problem with such conventional estimates is that resources in general, and class sizes in particular, are not only a cause but also a consequence of student achievement or of unobserved factors related to student achievement. Many features may lead to the joint and simultaneous determination of class size and student achievement, making class size endogenous to student achievement. For example, schools may reduce class sizes for poorly performing students and policymakers may design compensatory funding schemes for schools with large shares of students from poor backgrounds (see West and Woessmann 2006 for international evidence). In both cases, class sizes are allocated in a compensatory manner, biasing the class-size coefficient upwards. In contrast, policymakers may also have high-performing students taught in special small classes to support elite performance. Likewise, parents who particularly care for the education of their children may both make residential choices to ensure that their children are taught in schools with relatively small classes and support their children in many other ways, leading them to be relatively high performers. In these cases, class sizes are allocated in a reinforcing manner, biasing the class-size coefficient downwards. In short, parents, teachers, schools, and administrators all make choices that might give rise to a non-causal association between class size and student achievement even after controlling extensively for family background. Conventional estimates of class-size effects may thus suffer from endogeneity bias, the direction of which is ambiguous a priori.

To identify causal class-size effects, two quasi-experimental strategies have been applied to the international test data (cf. Woessmann 2005b). The first quasi-experimental approach draws on exogenous variation in class size caused by natural fluctuations in the size of subsequent student cohorts of a school (similar to Hoxby 2000). In this case, the quasi-experiment results from the idea that natural fluctuations in student enrollment lead to variations in average class size in two

[10] See Table 2.7 in Hanushek and Woessmann (2011a) for an overview of within-country studies on school inputs.

[11] Using PIRLS data, Woessmann (2010) estimates a quasi-value-added model, controlling for retrospective information on pre-school performance, for primary-school students in two Latin American and several comparison countries.

adjacent grades in the same school. Natural birth fluctuations around the cut-off date that splits students into different grade levels occur randomly. Therefore, they lead to variation in class size that is driven neither by students' educational achievement nor by other features that might jointly affect class size and student achievement.

Woessmann and West (2006) develop a variant of this identification strategy that exploits specific features of the TIMSS database. The sampling design of the first TIMSS study, which tested a complete 7th-grade class and a complete 8th-grade class in each school, enables them to use only the variation between two adjacent grades in individual schools. This strategy aims to exclude biases from nonrandom between-school and within-school sorting through a combination of school fixed effects and instrumental variables using grade-average class sizes as instruments. The rationale of this approach is as follows. Any between-school sorting is eliminated in a first step by controlling for school fixed effects, restricting the analysis solely to variation within individual schools. Within schools, the allocation of students to different classes in a grade may also be non-random. Within-school sorting is filtered out in a second step by instrumenting actual class size by the average class size in the relevant grade in each school. Within-school variation in class size is thus used only insofar as it is related to variation in average class size between the 7th and 8th grade of a school. The identifying assumption is that such variation is not affected by student sorting but reflects random fluctuations in birth-cohort size between the two grades in the catchment area of each school. Thus, causal class-size effects are identified by relating differences in the relative achievement of students in 7th and 8th grade within individual schools to that part of the between-grade difference in class size in the school that reflects between-grade differences in average class size.

Figure 8.4 illustrates the basic intuition behind this identification strategy for the example of math achievement in Singapore. The top panel indicates that class-average test scores are *positively* associated with class size, as is the case in most countries—likely reflecting ability sorting of students between and within schools. The middle panel plots the achievement difference between the 7th-grade and 8th-grade class in each school against the same grade difference in class size, which is equivalent to including school fixed effects in a regression framework. Overcoming effects of between-school sorting by removing any difference in overall achievement levels between schools, the size of the positive correlation is reduced substantially, but remains statistically significant. The reduction suggests that poorly performing students tend to be sorted into schools with smaller classes in Singapore. The final step of the identification strategy, illustrated in the bottom panel, additionally eliminates any effects of within-school sorting by using only that part of the between-grade variation in actual class sizes that can be predicted by variation in grade-average class sizes. The picture suggests that class size has no causal effect on student achievement in math in Singapore. Rather, weaker students seem to be consistently placed in smaller classes, both between and within schools.

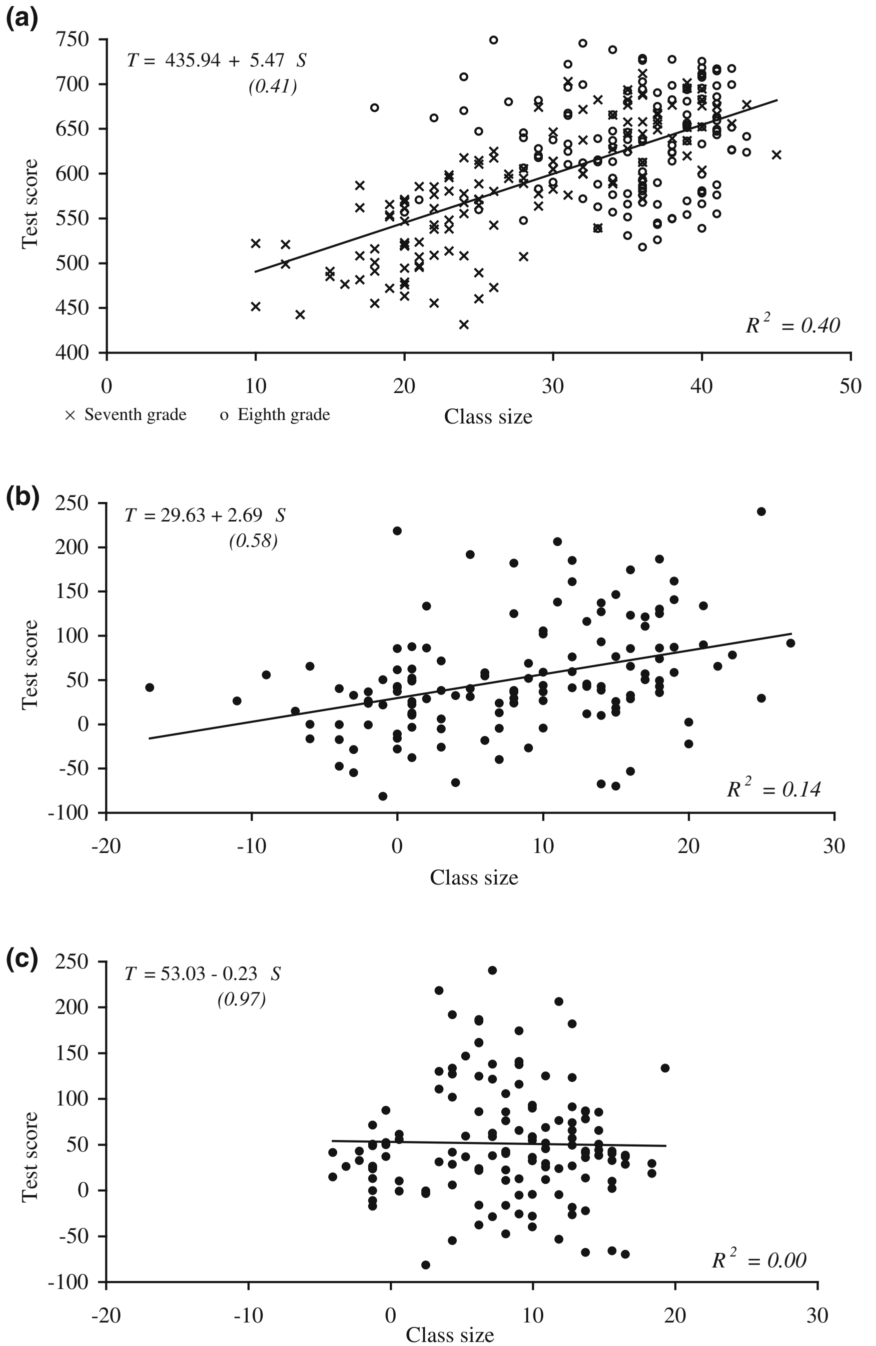

Fig. 8.4 Identifying class-size effects: Singapore as an illustrative example. *Source* Woessmann (2007) **a** All classes, **b** grade difference and **c** grade difference, instrumented

Woessmann and West (2006) implement this identification strategy in microeconometric estimations of education production functions for 11 countries around the world.[12] In line with Fig. 8.4, their results suggest that conventional estimates of class-size effects tend to be severely biased. They find sizable beneficial effects of smaller classes in Greece and Iceland, but reject the possibility of even small effects in four countries and of large beneficial effects in an additional four countries. Additional specification tests support the identifying assumption that students and teachers are not systematically sorted between grades within individual schools. There are no systematic differences at all in the observable characteristics of students or teachers between the two grades in schools in which one of the two adjacent grades has substantially larger average class sizes than the other; there are no systematic differences in the estimated class-size effects between expanding, stable, and contracting schools; and there are no systematic differences in the estimated class-size effects between countries where 7th grade is the first grade of a particular school and countries where it is not so that grade-average class sizes might have been adjusted based on schools' experience with the particular students.

The basic pattern of results is corroborated by a second quasi-experimental identification strategy based on rule-induced discontinuities. Following the study by Angrist and Lavy (1999) for Israel, Woessmann (2005b) exploits the fact that many countries have maximum class-size rules that induce a nonlinear association between the number of students in a grade of a school and average class size. In particular, the association has sharp discontinuities at multiples of the maximum class size that can be exploited to identify variation in class sizes that is exogenous to student achievement. The TIMSS data suggest that 10 West European school systems implement national maximum class-size rules reasonably strictly and with enough sharpness to enable an empirical implementation of this instrumental variable strategy.[13] In all 10 countries, results from identification by rule-induced discontinuities rule out the possibility of large causal class-size effects in lower secondary school. The only statistically significant, but small estimates are, again, in Iceland and, marginally, in Norway.

Woessmann (2005b) shows that these results are robust to several specification tests. Some models control for peer effects, in terms of the mean achievement and family background of each student's classmates, to exclude bias from peer sorting. Controlling for any continuous association between grade enrollment and student achievement by adding enrollment in the specific grade and its squared term as additional controls does not lead to substantive changes in results. When applying the specification to a discontinuity sample of students whose grade enrollment is within a margin of plus or minus 5 or 6 students of the rule-based discontinuities, so that identification does not come from observations far off the discontinuities, the

[12]Additional evidence based on the same identification strategy for countries in West Europe, East Europe, and East Asia is presented in Woessmann (2005b), Ammermueller et al. (2005), and Woessmann (2005a), respectively.

[13]The ten West European school systems that employ maximum class-size rules are: Denmark, France, Germany, Greece, Iceland, Ireland, Norway, Spain, Sweden, and Switzerland.

instrument gets weak in about half the countries, while results remain robust in the other half. Excluding especially large schools in each country (of a size three or four times the maximum class size) does not lead to a substantive change in results.[14]

However, as discussed by Woessmann (2005b), some reservations remain with this regression-discontinuity identification strategy (cf. also Urquiola and Verhoogen 2009). In particular, intentional exploitations of the rule by systematic between- and within-school choices might lead to remaining endogeneity in the rule discontinuity approach. Thus, it is possible that parents and schools "play the system": parents particularly keen to ensure low class sizes for their children may make their enrollment decisions—and school principals their acceptance decisions—on the basis of expected class size, and those decisions may be related to student achievement. Still, in the end both quasi-experimental identification strategies come to a very similar pattern of results. Moreover, the source of the potentially remaining biases differs in the two cases, adding confidence that any remaining bias in each strategy is of second-order magnitude.

Both identification strategies reach the conclusion that class size is not a major force in shaping achievement in lower secondary school in any of the countries considered. There is no single country for which any of the specifications showed a statistically significant and large class-size effect. In every case where one of the methods leads to a reasonably precise estimate, a large effect size can be ruled out with considerable statistical confidence. There is only one country, Iceland, where results create confidence that a causal class-size effects exists. However, in both specifications the estimates are relatively small and estimated precisely enough to reject the possibility of a large effect.

The unique value of cross-country research, however, lies in analyses of whether the cross-country differences in estimated class-size effects are systematically related to underlying features of the school systems. Such analyses can improve our understanding of the particular circumstances under which class sizes matter or not. Although causal class-size effects are small at best in all the countries considered, there are still differences across countries. The international evidence shows that the estimated effect size does not vary systematically for children from differing family backgrounds or for countries with different levels of average achievement, economic development, average class size, or educational spending (Woessmann and West 2006; Woessmann 2005b). But the existence of class-size effects is systematically associated with the salary and education level of the teaching force. In both studies, class-size effects were detected only in countries with relatively low teacher salaries and education. The pattern is similar within countries in which the education level of teachers varies. In these countries, the estimated class-size effect tends to be larger in classes that are taught by teachers with lower education. Interpreting average teacher salary and teacher education as proxies for average teacher quality, the results suggest that relatively capable teachers do as well when teaching large classes as when teaching small classes. By contrast, less capable

[14]The size of the induced discontinuity in class size is smaller when grade enrollment is larger.

teachers do not seem to be up to the job of teaching large classes, while doing reasonably well in small classes. Consequently, the pattern of international effect heterogeneity suggests that class-size effects occur only when the quality of the teaching force is relatively low.

A third approach to the identification of causal class-size effects tries to avoid bias from non-random sorting of students by using variation within individual students. If the same student is taught two different academic subjects in differently sized classes, the within-student between-subject variation can be used for identification (cf. Dee 2005; Dee and West 2011). The inclusion of student fixed effects, implemented by differencing across subjects, effectively excludes bias from subject-invariant student, family, and school characteristics, observable and unobservable. Unobserved characteristics that vary by subject and are correlated with class size, such as subject-specific fast-track or enrichment classes or teacher characteristics, could, however, still bias this research design. Altinok and Kingdon (2012) implement this identification strategy to estimate class-size effects in up to 45 countries using TIMSS 2003 data, which provide test scores in math and science for each student. Their results provide little support for class-size effects, with only few countries showing significant and sizeable positive effects of smaller classes. Analyzing the cross-country variation in class-size effects, they confirm that class-size effects are larger where teacher qualifications are lower, and also find indication of larger class-size effects in developing countries.

Beyond class-size effects, Ammermueller and Dolton (2006) use the same cross-subject identification strategy to estimate the effect of teacher-student gender interaction in England and the United States using TIMSS and PIRLS data. In most specifications (with the exception of one in England), they find little evidence of a significant effect of the interaction between student and teacher gender on student achievement. Schwerdt and Wuppermann (2011) use the same cross-subject identification with student fixed effects to identify the effects of teaching practices on TIMSS data in the United States. At a more descriptive level, Bratti et al. (2008) use the PISA data to estimate the association of student achievement with cooperative and competitive attitudes towards learning at the individual and school level.

8.5 Conclusions and Outlook

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.

All in all, the international evidence on the role of school inputs in educational production provides little confidence that quantitative measures of expenditure and class size are a major driver of student achievement, across and within countries. Studies using different methods to identify causal class-size effects consistently find no strong effects of class size in most countries. Among school inputs, descriptive evidence suggests that measures of the quality of inputs and, in particular, teachers are more closely related to student outcomes. However, research in this area awaits more work to identify the underlying causal links.[15]

A particular opportunity of the international research is that it can unveil whether certain effects differ systematically across countries. For example, the international pattern suggests that significant class-size effects are only present in systems with relatively low teacher quality. This result raises the cost-effectiveness question of whether student achievement is best served by reducing class size or by increasing the low teacher quality even in the countries where class-size effects are present.

Due to the limited role of differences in expenditures and class size in explaining cross-country achievement differences, it may be tempting to conclude that school systems do not matter so much for student achievement, after all. Nothing could be more wrong than that. Evidence that differences in teacher quality and instruction time do matter suggests that what matters is not so much the amount of inputs that school systems are endowed with, but rather how they use them. Correspondingly, international differences in institutional structures of school systems such as external exams, school autonomy, private competition, and tracking have been found to be able to account for a substantial part of the cross-country variation in student achievement (see Woessmann 2016 for a recent review).

As the economic literature on international evidence on educational achievement has emerged only relatively recently, there is obviously still considerable scope for future advances. 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, Falck and Woessmann (2013) attempt to derive measures of entrepreneurial intentions from the international background data, and Chap. 6 in Woessmann et al. (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 the international background questionnaires. As a more distant outlook, international testing of non-cognitive skills would be an obvious challenge.

[15]More recently, Hanushek et al. (2014) show effects of teacher cognitive skills on international differences in student achievement.

As more and more countries participate in the 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 inputs and start to analyze additional specific features beyond the broad concepts of input variables 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 20–30 years from now. The number of participating countries is as high as 52 in TIMSS 2011 and 65 in PISA 2012, and additional countries have signed up to participate in the most recent 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

Afonso, A., & St. Aubyn, M. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling, 23*(3), 476–491.

Altinok, N., & Kingdon, G. (2012). New evidence on class size effects: A pupil fixed effects approach. *Oxford Bulletin of Economics and Statistics, 74*(2), 203–234.

Ammermueller, A., & Dolton, P. (2006). *Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA*. ZEW discussion paper 06-060. Mannheim: Centre for European Economic Research.

Ammermueller, A., Heijke, H., & Woessmann, L. (2005). Schooling quality in Eastern Europe: Educational production during transition. *Economics of Education Review, 24*(5), 579–599.

Angrist, J. D., & Lavy, V. (1999). Using Maimondides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 114*(2), 533–575.

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

Bratti, M., Checchi, D., & Filippin, A. (2008). *Should you compete or cooperate with your schoolmates?* IZA discussion paper 3599. Bonn, Germany: Institute for the Study of Labor.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2), 158–165.

Dee, T. S., & West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis, 33*(1), 23–46.

Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy , 26*(65), 5–55.

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.

Fertig, M., & Wright, R. E. (2005). School quality, educational attainment and aggregation bias. *Economics Letters, 88*(1), 109–114.

Fuchs, T., & Woessmann, L. (2004). Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school. *Brussels Economic Review, 47*(3/4), 359–385.

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.

Gundlach, E., & Woessmann, L. (2001). The fading productivity of schooling in East Asia. *Journal of Asian Economics, 12*(3), 401–417.

Gundlach, E., Woessmann, L., & Gmelin, J. (2001). The decline of schooling productivity in OECD countries. *Economic Journal, 111*(471), C135–C147.

Gustafsson, J.-E. (2003). What do we know about effects of school resources on educational results? *Swedish Economic Policy Review, 10*(3), 77–110.

Gustafsson, J.-E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 37–63). Washington, DC: Brookings Institution Press.

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. (1997). The productivity collapse in schools. In W. J. Fowler Jr. (Ed.), *Developments in School Finance, 1996* (pp. 185–195). Washington, DC: National Center for Education Statistics.

Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics* (Vol. 4, pp. 2045–2141). Amsterdam, the Netherlands: 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, 104*, 212–232.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review, 22*(5), 481–502.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance*. NBER working paper no. 20727. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics, 78*(4), 611–627.

Hanushek, E. A., & Woessmann, L. (2011a). 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, 89–200). Amsterdam, the Netherlands: North Holland.

Hanushek, E. A., & Woessmann, L. (2011b). How much do educational outcomes matter in OECD countries? *Economic Policy, 26*(67), 427–491.

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. Cambridge, MA: MIT Press.

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. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics, 115*(3), 1239–1285.

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

Michaelowa, K. (2001). Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations. *World Development, 29* (10), 1699–1695.

Schwerdt, G., & Wuppermann, A. C. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review, 30*(2), 365–379.

Todd, P. E., & Wolpin, K.I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal, 113*(485), F3–F33.

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

Urquiola, M., & Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review, 99*(1), 179–215.

West, M. R., & Woessmann, L. (2006). Which school systems sort weaker students into smaller classes? International evidence. *European Journal of Political Economy, 22*(4), 944–968.

Woessmann, L. (2002). *Schooling and the quality of human capital*. Berlin, Germany: Springer.

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

Woessmann, L. (2005a). Educational production in East Asia: The impact of family background and schooling policies on student performance. *German Economic Review, 6*(3), 331–353.

Woessmann, L. (2005b). Educational production in Europe. *Economic Policy, 20*(43), 446–504.

Woessmann, L. (2007). International evidence on expenditure and class size: A review. In *Brookings papers on education policy 2006/2007* (pp. 245–272). Washington D.C.: Brookings.

Woessmann, L. (2010). Families, schools, and primary-school learning: Evidence for Argentina and Colombia in an international perspective. *Applied Economics, 42,*(21), 2645–2665.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives, 30*(3), 3–31.

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

Woessmann, L., & West, M. R. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review, 50*(3), 695–736.