**Kiel Institute of World Economics**
Duesternbrooker Weg 120
24105 Kiel (Germany)

**Kiel Working Paper No. 1051**

# New Evidence on the Missing Resource-Performance Link in Education

**by**

**Ludger Wößmann**

June 2001

# New Evidence on the Missing Resource-Performance Link in Education

**Abstract:**

New evidence confirms the conclusion of former surveys that the link between school resources and student performance is generally missing in educational production. While the conventional within-country cross-section evidence remains controversial, recent contributions which control for potential resource endogeneity suggest that returns to resource usage in education are decreasing. Resources may render positive effects at very low endowment levels prevailing in many developing countries, but their effect is weak to non-existent in advanced countries. The missing resource-performance link in education also shows up in international cross-section evidence and in within-country time-series evidence.

**Keywords:** education production functions, determinants of student performance, resource effects

**JEL Classification:** I2

**Ludger Wößmann**
Kiel Institute of World Economics
24100 Kiel, Germany
Phone: (+49) 431 8814-497
Fax: (+49) 431 8814-500
E-mail: woessmann@ifw.uni-kiel.de
Internet: www.uni-kiel.de/ifw/staff/woessman.htm

## 1 Resource Effects in Education: An Old Issue in the Literature

People's cognitive skills are of prime importance for their economic success. Evidence of substantial earnings returns to quantitative measures of education, like years of schooling, is abundant (cf. the recent surveys by Card (1999) and Ashenfelter et al. (1999)). But the earnings returns to qualitative measures of education, like scores on cognitive achievement tests, seem to be even higher (Boissiere et al. 1985, Bishop 1989, 1992) and increasing over time (Murnane et al. 1995). Even among school dropouts, i.e. students with the lowest educational quantity, there are substantial earnings returns to basic cognitive skills (Tyler et al. 2000).[1]

It is thus a crucial question how a high-quality education can be achieved. The issue of whether school resources render important effects on educational performance has occupied the literature for a long time. It goes at least as far back as the Coleman report (Coleman et al. 1966), and the literature was already voluminous at the time of Hanushek's (1986) well-known survey. The general finding of this literature was the surprising fact that "[t]here appears to be no strong or systematic relationship between school expenditure and student performance" (Hanushek 1986, p. 1162). This paper surveys the current state of this literature, which has advanced through a couple of outstanding new contributions, and adds two further strands of evidence on the relationship between school resources and educational performance.

The existing literature is nearly exclusively based on within-country cross-section evidence. The conventional approach is to estimate an education production function by regressing educational test score performance on measures of school resources and other covariates. It remains highly controversial what this vast literature of estimates from the United States tells us about the effect of school resources on educational performance (Section 2). The main problem raised against interpreting such estimates as causal effects is the potential endogeneity bias of the estimated resource effects. Several recent contributions to the within-country cross-section literature, most of which have been

[1] For further evidence on positive effects of cognitive test scores on individual labor-market performance, see Rivera-Batiz (1992), Neal and Johnson (1996), Currie and Thomas (1999), and the references in the latter. Also at the international level, the average performance on cognitive achievement tests has a strong and consistent influence on economic growth, and their explanatory power in growth regressions is superior to that of mere quantitative measures of education (Hanushek and Kimko 2000, Barro 2001).

published in leading economics journals, have tried to address the potential endogeneity of school resources to students' performance, either through instrumental variable estimation or through use of data from controlled experiments (Section 3). All in all, while cases exist where increased resource endowment led to superior student performance, most empirical evidence from the United States still seems to support the finding that an increase in the amount of school resources does not lead to an increase in educational performance. While equivalent evidence is basically non-existent for most other advanced countries, within-country cross-section research from developing countries shows some evidence of positive resource effects at very low levels of resource endowment (Section 4). Together, these findings suggest that schooling systems in advanced countries may be running into diminishing returns to resource usage at the resource levels currently reached.

In addition to the within-country cross-section evidence, this paper presents two further approaches to testing resource effects in educational production. International cross-section evidence shows that differences in the amount of inputs available in the schooling systems do not suffice as an explanation for the large international differences in students' educational performance levels (Section 5). Furthermore, within-countries time-series evidence reveals that a vast expansion of educational resources per student over the past 15 to 25 years did not lead to a considerable increase in average educational performance in most OECD and several East Asian countries (Section 6).

Thus, the main result which emerges from all three strands of evidence - within-country cross-section, international cross-section, and within-country time-series - is that merely increasing the resources available to the schooling system does *not* increase educational performance. A majority of schools does not seem to be making economically efficient use of given resources. To increase educational performance, education policy has to look for something else than mere resource enhancements.

## 2 Conventional US Within-Country Cross-Section Evidence

### *2.1 The Concept of the Education Production Function*

The approach of the education production function links the output of the education process, i.e. the educational achievement of students, to the inputs available in schools (cf. Hanushek 1986). Education production functions are usually estimated in the following form:

$$(1) \qquad t = B'\beta_1 + R'\beta_2 + \varepsilon \quad .$$

The output of educational production is usually measured by the test-score performance *t* of students on tests of cognitive achievement. Schooling inputs are encompassed in vectors of measures of students' family background characteristics *B* and of resource measures *R*. The $\beta$s are vectors of parameters, and $\varepsilon$ is an error term. The resource measures *R* can include variables such as expenditure per student, class size, teacher-student ratio, teacher characteristics, and endowment with teaching facilities. Since apart from the opportunity cost of students' time, teachers are the most important, and most costly, factor of production in schooling, much of the discussion of the impact of school resources on students' performance goes under the heading of the "class-size debate," which deals with the effect of differences in teacher-student ratios on performance.[2]

Given data on schooling output and schooling inputs, the education production function of equation (1) can be estimated, typically by single-equation regression, to yield evidence on the direction, magnitude, and significance of the input-output relationships. In this paper, the main focus is on the relationship between schooling resources and educational performance. In such studies, it is vital to control for students' family background, such as income and education of their parents. Otherwise, the estimates of resource effects may be biased because family background and schools' resource endowment may be correlated. E.g. in the United States, where most of the school funding comes from local property taxes, the resources available to a school district depend on the income of the local residents. The estimated resource effects would then not necessarily indicate a causal impact of resources but might instead just proxy for the impact of family background.

The econometric estimation of education production functions began in earnest with the Coleman report (Coleman et al. 1966), a congressionally mandated study by the Office of Education in the United States. Since then, a large amount of research has

[2] Note that class size is measured inversely to the teacher-student ratio. Thus, a positive effect of additional resources on student performance would be represented by an estimated *positive* coefficient on the teacher-student ratio, while being represented by an estimated *negative* coefficient on class size (or on the student-teacher ratio).

focused on the estimation of education production functions, especially in the United States. During the three decades of analysis until 1994, more than a hundred individual publications in books or academic journals have been counted in literature surveys. In recent years, several journals have dedicated whole issues to the topic of the effectiveness of resource use in education, among them the Journal of Economic Perspectives (1996), the Review of Economics and Statistics (1996), and the Economic Policy Review of the Federal Reserve Bank of New York (1998). The questions of how the econometric results in the literature can be summarized and what they tell us about policy implications initiated a controversial discussion in the literature, with Eric Hanushek and Alan Krueger as the main antagonists in recent debates.

### *2.2 Hanushek's Survey of the US Evidence*

Hanushek (1997, cf. also 1999a), updating his previous literature survey, summarizes the resource effects estimated by conventional within-country cross-section education production functions by tabulating them as statistically significantly positive, statistically significantly negative, and statistically insignificant. He collects all studies available through 1994 which meet three minimal publication, specification, and reporting criteria. First, they must be published in a book or journal to ensure minimal quality standards. Second, they must include some measure of family background in addition to at least one measure of school resources. And third, they must inform about the statistical significance of the estimated resource effects. Most studies measure educational output by student performance on standardized tests, while some studies use other measures such as dropout behavior or subsequent earnings in the labor market.

From all the publications which meet the three criteria for analytical design and reporting of results to be included in the summary, Hanushek takes each separate estimate of a resource effect from an education production function as a single independent observation. A typical publication is likely to contain several sets of estimates, derived from different output measures, different grade levels, and different sampling designs. Hence not all estimates can be considered as independent from each other. If several estimates in a publication are based on alternative specifications using the same sample and performance measure, only one of them was included.

Figure 1 reports the results of Hanushek's literature survey for the United States. The 90 included individual publications with US evidence (listed in Hanushek 1997) yielded 377 separate production function estimates. While many of the included studies were produced in the direct aftermath of the Coleman Report, half of the available studies have been published since 1985. Without going into detail of the different estimates and their relative merits with respect to suitability and data quality, it is apparent that for most resource measures only a small fraction of published estimates reports a statistically significant positive effect. For each of the resource categories, at least two thirds of the estimates are statistically insignificant, and the statistically significant positive estimates are usually nearly balanced by statistically significant negative estimates. That is, without even looking at the size of the estimated effects, it is fair to say that there is not much confidence in getting noticeable effects of schooling resources on educational performance. This leads Hanushek (1997, p. 148) to repeat the conclusion of his former survey that "[t]here is no strong or consistent relationship between school resources and student performance." Furthermore, there seem to be large differences in the relationship across classrooms and schools, and the distribution of underlying resource parameters suggests that while resources are used effectively in some circumstances, in most circumstances they are not.

### *2.3 Krueger's Reinterpretation of the US Evidence*

Krueger (1999b) criticizes Hanushek's (1997) method for summarizing the literature mainly on the basis of his weighting of estimates. As described above, Hanushek's rule for weighting estimates is to give equal weight to any production function estimate, regardless of whether there was only one estimate or several estimates within the publications where it is taken from. Hence more estimates would be taken from a publication which analyzes several subsamples of a larger data set than from a publication which uses only the complete sample of the larger data set. Therefore, Krueger argues that Hanushek's method gives larger weights to estimates based on smaller samples. Other things being equal, estimates based on smaller samples are likely to yield less significant results. Krueger (1999b) shows that publications included in Hanushek's summary with only one estimate tend to report far more statistically significant positive results than publications with many estimates. Thus, the prevalence of statistically insignificant estimates in Figure 1 may not be due to the lack of a systematic relationship between resources and student performance, but due to small sample sizes over-represented in Hanushek's averages.

In his reanalysis of Hanushek's results, Krueger (1999b) gives equal weight to each underlying publication, instead of each individual estimate extracted from the publications. His results, reported for the estimates of teacher-student ratios and expenditure per student, are replicated in Figure 2 ("Based on publications"). As expected, Krueger's method yields a larger proportion of statistically significant positive results and a smaller proportion of statistically insignificant results.

Based on a ratio of positive to negative results which gives equal weight to statistically significant estimates and statistically insignificant estimates, Krueger shows that the larger prevalence of positive results is unlikely to be caused by chance and therefore argues that class size is systematically related to student performance. However, the proportion of statistically significant positive results is still only 26 percent for teacher-student ratios and 38 percent for expenditure per student. Even under his reanalysis, the majority of results is either statistically insignificant or statistically significantly negative.[3]

In direct response to Krueger's criticism, Hanushek (2000) shows that Krueger's assumption that publications with fewer individual estimates will tend to have larger sample sizes is incorrect. In the publications included in the survey, there is no clear relationship between the number of estimates per publication and the sample size. The median and the average of the sample size is about the same in publications with one estimate and in publications with more than seven estimates, while publications with two to three estimates have the largest median sample size and publications with four to seven estimates have the largest average sample size. The simple correlation between sample size and number of estimates in the underlying publications is not statistically significantly different from zero. Since the number of estimates per publication is not

[3] Krueger (1999b) also reports results where publications are weighted by their number of citations in other work, which are quite similar to the results which give equal weight to each publication. Furthermore, he reports adjusted results based on regressions of the percent of estimates being positive or negative, and significant or insignificant, on the number of estimates used from each study. This method seems rather dubious since the assumption that these relationships are linear bears no theoretical foundation and can presumably be rejected empirically.

related to sample size, there is no obvious reason for Hanushek's weighting rule to be flawed. This fact basically erodes the foundation of Krueger's critique, since Hanushek's method does *not* give larger weights to estimates based on smaller samples. The prevalence of statistically insignificant resource estimates does not seem to stem from small sample size, but from a missing relationship between resources and student performance. Furthermore, the decision to include more than one estimate in a publication is generally based on sound econometric reasons, e.g. when the estimates come from different cities with differing standardized tests as measures of educational output.[4]

Hanushek (2000) argues that his general results may actually be biased in the direction of finding too many statistically significant positive effects rather than too few. First, there is considerable empirical evidence that a publication bias exists in the sense that studies with statistically significant estimates are more likely to be published.[5] Second, Hanushek et al. (1996) have shown that data aggregation to the state level leads to upward bias of estimated resource effects when differences in schooling policy across states are omitted in multiple state samples. When using only studies based on disaggregated data (at the classroom, school, district, or county level, but not at the state level), the proportion of statistically significant positive estimates shrinks ("Disaggregated data" in Figure 2). This may actually be an explanation for the findings of Krueger's (1999b) reanalysis, since 40 percent of the publications which contain only a single estimate use state aggregate data, compared to only 4 percent of all estimates. Thus, weighting by publications rather than by separate estimates puts heavy weight on low-quality estimates. Third, when the research question is the effect of current resources on student performance, studies which do not control for initial performance may be misleading, since education is a cumulative process and builds on knowledge

---

[4] Hanushek (2000) adds that Krueger's equal weighting of statistically significant estimates and statistically insignificant estimates in the calculation of his ratio of positive to negative results seems to violate the basic premise of his re-weighting. He also questions the assumption that citation weights give a more accurate estimate of the quality of the underlying estimates, since studies are often cited because they are innovative, controversial, or just diverging from the rest of the literature, neither of which is necessarily a sign of quality.

[5] E.g., in a recent review of estimates of the relationship between schooling and earnings, Ashenfelter et al. (1999) show that estimated rates of return to education suffer from considerable publication bias.

accumulated in earlier grades. Value-added models which control for students' performance at the beginning of a grade, usually by focusing on the growth in student performance over one grade, seem conceptually superior for estimating resource effects. While no value-added models are available for expenditure per student, only 12 percent of the estimates from value-added models for teacher-student ratios show a statistically significant positive effect ("Value-added models" in Figure 2), and only one of the available 24 estimates which come from value-added models with disaggregated data from within a single state shows a statistically significant positive sign.

In sum, Krueger's (1999b) reanalysis of the conventional US cross-section evidence seems to place much larger weight on low-quality and biased statistical estimates. And even then, still only 26 percent of the estimates for teacher-student ratios are statistically significant and positive. Based on conventional within-country education production functions, the econometric evidence for positive effects on students' educational performance through resource increases across the whole range of schools remains weak.

## 3 Resource Endogeneity, Instrumental Variables, and Experiments

### *3.1 The Unclear Direction of the Potential Bias Due to Resource Endogeneity*

The data used to estimate resource effects usually originate from surveys conducted for other purposes. Ideally, experimental or randomized data should be used to estimate education production functions. Conventional estimates of resource effects may thus be flawed because of endogeneity problems, i.e. because the resources devoted to the education process may not be exogenous to the performance of students. It is conceivable that the amount of resources depends on choices by politicians, administrators, teachers, and parents which may, among others, be based on the level of performance achieved by the students. In this case, conventional ordinary least squares (OLS) estimates of resource effects in equation (1) will produce biased estimates of the causal effects of resources on student performance.

The direction of the bias introduced by resource endogeneity is ambiguous (cf. Hoxby 2000a). On the one hand, if resource choices are compensatory in that poor student performance induces an increase in available resources, this will cause estimated resource effects to be biased downwards. The within-school allocation of resources, usually decided upon by school administrators and teachers, may be one potential source of compensatory variation in school resources. E.g., in his model of educational production, Lazear (1999, p. 2) argues that "class size is a choice variable and the optimal class size varies inversely with the attention span of students." If available resources indeed vary primarily because schools are adjusting resources in response to the behavior and performance of the students in the sense that poor performance induces increased resource endowment, then the classes which are larger and endowed with fewer resources will have the better students and, therefore, higher measured student performance. The ensuing negative relationship between resources and educational performance will not be an estimate of the causal effect of resources on student performance, but will mainly reflect the reverse causality running from student performance to allocated resources. Likewise, if policy makers design compensatory funding schemes which assign higher expenditure to weaker students, this will also bias the estimated resource effect downwards.

On the other hand, if resource choices are reinforcing in that better students receive more resources, this will cause estimated resource effects to be biased upwards. Both endogenous within-school resource allocations and between-school resource variations may actually be reinforcing. If schools respond more to the demands of parents who contribute more to their children's learning, and if these parental contributions are not sufficiently controlled for, the positive correlation between resource allocation and parental contributions will cause the estimated resource effects to be biased upwards. Likewise, policy makers may introduce reinforcing policies, e.g. by endowing better students with special facilities in order to support elite students.

In addition, residential choices by parents may be based on the amount of resources available in a school district and may thus also trigger compensatory or reinforcing resource variations. Indeed, Hoxby (2000a, p. 1245) suggests that school choice through parents' residential choice is "probably the single largest source of variation in school inputs" in the United States. If parents move to school districts with high schooling expenditure, the problem for empirical research is that their residential decision may be based on their children's ability or prior educational performance, either in a reinforcing or in a compensatory manner. Thus, residential choice is likely to generate resource endogeneity, and the resulting bias can go either way.

In sum, the direction of any potential bias introduced by resource endogeneity is unclear. Two possibilities exist to answer the questions of whether endogenous resource assignments indeed cause biased findings, and in which direction a potential bias may work. One possibility is to use instrumental variables to account for potential endogeneity of resource endowments (Section 3.2). The other possibility is to perform experiments which randomly assign different amounts of resources to different students (Section 3.3).

### *3.2 Instrumental Variable Estimates*

The instrumental variable (IV) estimator is the most common econometric method used to solve a potential endogeneity problem. If there is resource endogeneity, the problem with conventional OLS estimates of equation (1) is that the resource measures $R$ will not be independent of the error term $\varepsilon$. With the IV estimator, an instrument which is correlated with the resource effect but uncorrelated with $\varepsilon$ is used to extract information only from that part of the variation in the resource variable which is exogenous to the measure of educational output. In a first stage, the resource variables $R$ are regressed on the instrumental variables $I$ ($\gamma$ is a vector of estimated parameters and $\mu$ is an error term):

$$R = I'\gamma + \mu \quad . \tag{2}$$

Given that the vector of instruments $I$ is indeed independent of the error term $\varepsilon$ of equation (1), the predicted values $\hat{R}$ of equation (2) will also be independent of $\varepsilon$, thus representing resource variations which are truly exogenous to conditional student performance. Therefore, in a second stage, the predicted values $\hat{R}$ are used instead of $R$ in equation (1) to estimate the causal effect of resources on student performance. Quite often, the two-stage least squares (2SLS) approach of IV estimation is employed, where the list of instruments $I$ in equation (2) entails not only the additional variable(s) chosen as special instrument(s) for the resource variable, but also all other exogenous variables of equation (1).

Several recent studies have used IV estimates to account for potential resource endogeneity. Akerhielm (1995) applies 2SLS estimation to account for the endogeneity of class size due to non-random allocation of students to different classes. She uses the average class size for a given subject in a student's school and total student enrollment in the school at the specific grade level as instruments for class size. The average class size in the school is meant to be exogenous to an individual student's performance while being strongly correlated with the student's actual class size, thereby bypassing the within-school non-random student allocation problem by excluding the within-school source of variation in class size. Total grade enrollment in the school is also meant to be exogenous to an individual student's performance but is meant to identify the fact that larger schools may be more able and therefore more likely to offer specialized small classes for lower ability students.

Akerhielm (1995) finds for a nationally representative sample of students from the United States that the OLS estimates show a positive relation between class size and student performance which is statistically significant in three out of four subjects. However, her IV estimates show a negative relation between class size and performance which is statistically significant in science and history while being statistically insignificant in mathematics and English. Hence, her findings seem to confirm positive resource effects, at least in some subjects. However, all the estimated effects are very small in size. Furthermore, it is unclear whether her IV approach yields truly exogenous variations in class size because the between-school variation in class size may be influenced by between-school sorting due to residential choice by parents, which is not instrumented for. Additionally, enrollment, which is used as an instrument, may be associated with student performance because better schools grow larger when parents choose to enroll their children there and because socio-economic status may be inversely related to population density, so that the IV estimate of the class size effect may be contaminated by effects not caused by changes in class size.

Boozer and Rouse (2001) use state regulations on special education to instrument for class size. They argue that actual class size will be correlated with state regulations on maximum class sizes for several types of special needs classes to the extent that schools base the entire structure of their class sizes on such state policy. Using the same source of data used in Akerhielm's (1995) study, most of the IV estimates of their different specifications yield a negative effect of class size on student performance and are usually reported to be statistically significant.[6] The choice of the instruments may be flawed, however, since state regulations will probably reflect the preferences of the residents of the state towards education rather than maximum class sizes which are exogenously determined. That is, the instrument may be both endogenous and a proxy for parental preferences. Overidentification tests reject the null hypothesis of consistent IV estimates for most of their results.

Sander (1999) uses demographic characteristics of a school district, namely the percentage of the population of a school district which is of school age, to instrument for either expenditure per student or for teachers' salary in two specifications of an education production function. He argues that an increase in the school-aged part of a district's population reduces expenditure per student. No instrument is applied to the class size variable. Sander (1999) estimates a 2SLS regression and finds that expenditure per student and average teachers' salary have modest positive effects on eighth-grade mathematics test scores of students in Illinois which are statistically significant at the 10 percent level, while neither has a statistically significant effect on mathematics performance in third grade. Class size estimates are statistically insignificant in all cases. Sander's (1999) study suffers from the use of aggregated data, and doubts about the validity of the choice of instruments may be due.

Looking at evidence outside the United States, Angrist and Lavy (1999) exploit a potentially exogenous source of variation in the class size of Israeli primary schools which originates from an administrative rule. Maimonides, a twelfth century rabbinic scholar, proposed the rule of a maximum class size of 40. This rule on maximum class size establishes a nonlinear and nonmonotonic function between total grade enrollment in a school and class size in Israeli schools today which is presumably exogenous to student performance. Therefore, the class size predicted by enrollment through the nonlinear class-size function derived from Maimonides' rule can be used as an instrument for the actual class size. At the same time, any potential linear or monotonic effect of total enrollment on student performance can be controlled for in the estimation.

[6] However, the reported significance levels should be seriously biased upwards since they do not take into account a school component of the error term due to the stratified sampling structure. Therefore, Boozer and Rouse's (2001) significance statements have to be viewed with great caution.

Angrist and Lavy (1999) find statistically significant positive IV estimates of effects of smaller classes for most of their specifications on mathematics and reading achievement in fifth grade and (albeit smaller) for reading in fourth grade, while the estimates for mathematics in fourth grade and for mathematics and reading in third grade are statistically insignificant and sometimes negative. While being superior to Akerhielm's (1995) study in controlling for smooth effects of total school enrollment on test scores, Angrist and Lavy's (1999) study may still suffer from endogeneity due to residential choice. Furthermore, Israel has relatively large class sizes and the class-size variation analyzed in this study lies in the range between 20 and 40 students in a class, so that the relevance of the findings for general class-size reductions in developed countries is in doubt.

Hoxby (2000a) uses two identification strategies based on the natural longitudinal variation in the population of school-aged children to instrument for class size effects. In the first strategy, exogenous variation in class size is generated by the fact that student cohorts vary in size due to the natural randomness in the timing of births. Using long panels of data on enrollment and kindergarten cohorts in each district, the random part of population changes can be isolated separately for each grade in each school and used as an instrument for class size. The second strategy is similar to the one pursued by Angrist and Lavy (1999) in that discontinuous changes in class size due to explicit class-size rules are exploited. Events are analyzed where a small change in enrollment triggers a maximum or minimum class-size rule in a specific school district and thereby changes the number of classes in a grade in a school. Based on a panel data set for Connecticut school districts, Hoxby (2000a) finds that there is no statistically significant relationship between class size and scores on mathematics, reading, and writing tests in fourth and sixth grade with any of the two identification strategies.

The lack of any statistically significant effect of class size reductions on student achievement in Hoxby's (2000a) study is especially noteworthy since the coefficients are very precisely estimated, that is, their estimated standard errors are very small. This virtually rules out the possibility that the statistical insignificance might be due to a problem of small sample sizes or other causes of imprecise estimates. Instead, it clearly points to the fact that there is no genuine link between resources and educational performance.

Summing up, just like the conventional OLS estimates, the studies which try to account for possible resource endogeneity by IV estimation present no clear evidence that changes in resource endowments have positive effects on student performance. All those studies which find positive resource effects in some situations also report insignificant or even negative estimates for other situations. Furthermore, there is still some doubt with most of the studies whether truly exogenous resource variations are identified. The two most rigorous studies find some evidence of positive effects of large-scale class-size reductions in Israel (Angrist and Lavy 1999), but no evidence at all of effects of class-size reductions on students' educational performance in the United States (Hoxby 2000a). That is, any potential biases introduced by resource endogeneity, which might hypothetically be directed either upwards or downwards (cf. Section 3.1), seem to be canceling out at least in developed countries.

### *3.3 Experimental Estimates: Tennessee's Project STAR*

While the studies summarized in the preceding Sections rely on econometric methods to separate the different factors influencing the educational performance of students, an alternative and potentially superior method is to carry out a controlled experiment. In such an experiment, students are randomly assigned to classes with different resource endowments. Given the randomness of the assignment, the resource endowment of a student should be independent of his or her educational performance, so that an experiment opens the possibility of estimating the true relationship between resource expenditure and educational performance.

The only large-scale random assignment experiment on educational resource endowment ever conducted in the United States is the Tennessee Student/Teacher Achievement Ratio experiment, usually referred to as Project STAR. In the mid-1980s, Project STAR was pursued as a longitudinal study which randomly assigned students and teachers to classes of different size from kindergarten through third grade. Specifically, the design initially included three treatments: small classes of 13-17 students, regular classes of 22-25 students, and regular classes of this larger size with a full-time teacher's aide. Schools large enough to have at least one of each class-size treatment were solicited for participation, and random assignment took place within schools. After the initial assignment, the design called for students to remain in the same class-size treatment group for the full four years. Between 6000 and 7000 students in initially 79 schools (subsequently falling to 75) were involved in the study each year. After third grade, the experiment ended, with all students being assigned to regular classes.

The results of the Project STAR experiment can in principle be easily derived by a comparison of treatment and control groups. With random assignment, a simple comparison of the mean achievements between students in small and large classes should provide an unbiased estimate of the effect of class size on achievement. As publicized through an internal team of researchers and reviewed in Mosteller (1995), students in smaller classes in the first year of their education (kindergarten) scored statistically significantly higher on standardized tests in mathematics, reading, and word recognition than students in regular-size classes, but the achievement advantage of small classes increased only slightly and statistically insignificantly in the next three years of schooling. The class size effects were larger for Black students and for students on free lunch. Students in regular classes with a teacher's aide did not perform significantly better than those in regular classes without a teacher's aide. Further findings on resource effects, seldom referred to in the discussion of Project STAR, is that teachers' experience and teachers' having a master's degree did not have a statistically significant effect on student achievement.

As is self-evident, the validity of experimental results depends crucially on the appropriate implementation of the test design of the experiment. In the actual implementation of Project STAR, considerable uncertainty about the results is introduced by several deviations from an ideal experimental design. While the initial examination of the Project STAR results by the internal researchers paid little attention to these potential threats to the validity of the experiment, Krueger (1999a) pursues a reanalysis of the Project STAR data in which he tries to address these issues. First, since randomization was done within schools and participating schools were self-selected, the exogeneity of the class-size assignment may only be valid within schools. Krueger adjusts for school effects by including separate dummy variables for each school. Still, the self-selection of schools may mean that schools with an efficient input-output relationship may be over-represented in the Project STAR sample because these may show a higher willingness and motivation to participate in the experiment.

Second, students in regular-sized classes were re-randomized between classes with and without aides after kindergarten, thereby changing the classmates of students in regular-size classes. Third, many new students entered the program in first grade, mainly because kindergarten attendance was not mandatory in Tennessee at the time of the study. While their allocation to treatment groups was random, any difference in prior treatment is unknown and thus cannot be controlled for. Fourth, between each grade, about 10 percent of the students switched between small and regular classes. These nonrandom transitions seem to have been mainly triggered by behavioral problems and parental complaints. Fifth, actual class size varied more then intended - from 11 to 20 students in small classes and from 15 to 30 students in regular classes -, presumably due to natural family relocation. To address the problems of class switching and class-size variability within a given type of assignment, Krueger uses the initial random assignment to a class type as an instrument for actual class size in a 2SLS approach. Thereby, only that part of class-size variation which is due to the initial assignment to the different class types is used in the estimation of test-score effects.

Sixth, less than half of the initial experimental group of students who entered the experiment in kindergarten remained in the experiment for all four grades. Some of these losses seem to have been caused by nonrandom switching to other schools after the class-type assignments had been made public. Krueger tries to adjust for nonrandom attrition in a crude way by imputing test scores for students who exited the sample through assigning the student's most recent test score. However, withdrawal rates prior to the start of kindergarten, which seem to have been larger for students assigned to regular-size classes, cannot be adjusted for.[7] All in all, Krueger (1999a) finds that none of the adjustments which he could make for deviations from an ideal experimental design overturn the main findings of Project STAR.

A seventh problem, which is imminent to any experiment, is that the actors in an experiment are aware of it (Hoxby 2000a). An experiment usually alters the prevailing incentive conditions, thereby potentially also altering the participants' behavior. E.g., the schools participating in Project STAR may realize that the experiment potentially affects the amount of resources available in the future and that if the experiment fails to show

[7] Krueger (1999a) presents an upper bound estimate of the possible effect of differential withdrawal before the start of kindergarten and shows that it seems to be relatively minor.

beneficial class-size effects, broad-based class-size reductions will never be enacted. Given that schools tend to have a self-interest in broad-based class-size reductions, they face incentives in the experiment which the fully enacted policy would not generate. Furthermore, teachers and students may respond to the fact that they are part of an experiment by temporarily increasing their productivity when they are being evaluated (so-called Hawthorne effects).[8]

An eighth limitation of the Project STAR experiment is that no pretest scores are available for participating students before they entered the experiment. As a consequence, it is impossible to assess the quality of student randomization for the initial experimental sample and for the subsequent additions to it (Hanushek 1999b). The estimated class-size effects would actually also be consistent with an initial assignment of somewhat better students to small kindergarten classes. Additionally, few information is available on the assignment of teachers.

Finally, it should be noted that the beneficial effect of being in a small class is largely restricted to the first year a student attends a small class. The interpretation of this finding depends on the underlying learning model with which one assesses it (Hanushek 1999a). Given that education is a cumulative process, one would expect the differences in performance to become wider through the grades as students in small classes continue to get more resources. But independent of the learning model employed, the empirical finding implies that additional resources in grades after kindergarten or possibly first grade do not yield significant achievement benefits. Even more, in their long-term follow-up of the students who participated in Project STAR, Krueger and Whitmore (2001) find that the benefit from being assigned to a small class from kindergarten through third grade appears to have declined to between one half and one quarter of the initial test-score advantage after students were returned to regular-size classes in fourth grade. While having been assigned to a small class increases the likelihood of taking a college-entrance exam by the end of high school, students who attended small classes do not seem to have scored significantly better on these exams.

Most of the discussed flaws in the experimental design of Project STAR can only poorly be addressed, while others cannot be addressed at all since they are imminent to

---

8 Krueger (1999a) tries to partially check for Hawthorne effects by looking just at the effects of class-size variations among students assigned to regular-size classes.

explicit experiments. Thus, while explicit experiments have clear advantages over non-experimental estimates, there are also a few disadvantages of explicit experiments. Furthermore, explicit experiments are rare, and up to now, there is only one experiment pursued in the US state of Tennessee to draw conclusions from.

However, it can be argued that "[o]ne well-designed experiment should trump a phalanx of poorly controlled, imprecise observational studies based on uncertain statistical specifications" (Krueger 1999a, p. 528). It is a pity, then, that the design and the implementation of the Project STAR experiment carried so many uncertainties that there is considerable doubt whether Project STAR can produce unbiased results. In addition, the estimated effects appear relatively small, given the magnitude of the class-size reductions of about eight students (Hanushek 1999b). Project STAR provides no results for smaller reductions in class size than the one-third reductions it pursued or for reductions in later grades than the very first ones. Moreover, the addition of a teacher's aide as another kind of class size reduction proved to be without any effect on student performance, as did the experience and education of the teacher as further resource variables.

Summarizing the within-country cross-section evidence on the effects of school resources on educational performance in developed countries, neither conventional estimates of education production functions nor studies which acknowledge the potential endogeneity of schooling resources show clear evidence that changes in resource endowments have sizable positive effects on educational performance. Results based on the controlled experiment of Project STAR point to relatively minor effects of class size on educational performance and suffer from methodological ambiguities. Hence evidence based on IV estimation may actually provide the most reliable evidence up to now. But these studies are certainly not the last word on the issue.

In the class-size debate, there seem to be two sides which stand heavily opposed to each other, one arguing that there are positive resource effects, the other arguing that if anything, estimated resource effects are negative. In a sense, each of the sides tries to show that the evidence presented by the other side is dubious. By being largely successful in this attempt, both sides together might lead an independent observer of the debate to conclude that the evidence presented is indeed dubious. Put differently, it seems fair to conclude that there is no clear evidence that a broad-scale increase in resource expenditure has a significant impact on student performance.

## 4 Within-Country Cross-Section Evidence from Developing Countries

Most of the available evidence on the link between resources and educational performance within a country comes from the United States. Unfortunately, no such encompassing evidence is available for other developed countries. Actually, Psacharopoulos (2000) identifies a big gap between the United States and Europe in terms of applied research in the economics of education in general. He ventures the statement that "[d]ue to the active role of international organizations, […] more research has been done on the economics of education in developing countries relative to Europe" (Psacharopoulos 2000, p. 93).

While existent, educational research in developing countries tends to be much less extensive and less rigorous than that for the United States. Still, evidence from developing countries might be especially illuminating since schools in developing countries on average work on much lower expenditure levels than schools in developed countries. If the education production function exhibits decreasing returns to resource usage, estimates of resource effects from countries further left on the production function (developing countries) should be more likely to yield positive resource effects than estimates from countries further right (such as the United States). As the Israeli evidence in Angrist and Lavy (1999) has shown, large-scale class-size reductions starting from a high class size may induce positive effects on student performance, while class-size variations at relatively low class sizes as already reached in developed countries do not seem to show effects on student performance.

### *4.1 Conventional Within-Country Cross-Section Evidence from Developing Countries*

Harbison and Hanushek (1992; cf. also Hanushek 1995) summarize 96 conventional estimates of education production functions from developing countries in a way similar to the US evidence presented in Section 2.2.[9] As the findings depicted in Figure 3 show, there are obviously more statistically significant estimates of resource effects on student

[9] It should be noted that not all of the underlying studies on which the 96 estimates are based seem to satisfy the publication criteria mentioned in Section 2.2.

performance in developing countries than in the United States. While there are several potential explanations for this - e.g., there might be less rigorous controlling for family background -, the differences in resource availability are large within developing countries, and most schools in developing countries work on much lower expenditure levels than schools in the United States. Therefore, the relatively more positive findings from developing countries lend support to the suggestion that schools in developed countries may be running into decreasing returns to the use of resources. That is, while there may be many circumstances in developing countries where resources lead to superior student performance, the marginal return to an increase in expenditure in developed countries seems to be close to zero.

Still, it should be noted that even in developing countries, only about one quarter of the estimates of teacher-student ratios are statistically significantly positive, and these are matched by an equal number of estimates with a statistically significant negative sign. For none of the classroom resources, two thirds of the estimates are statistically significantly positive. Therefore, a systematic relationship between resources and student performance is also in doubt in developing countries.

### *4.2 Quasi-Experimental Evidence from Apartheid South Africa*

Just like the conventional within-country estimates for the United States, the developing-country studies summarized in Figure 3 might again be biased through resource endogeneity. A recent study from a developing country by Case and Deaton (1999) tries to address this problem by looking at a specific setting where variations in resource endowment can be presumed to be largely exogenous. They use data from South Africa immediately before the end of the apartheid regime. Under apartheid, resource decisions for most Black schools were made centrally by White-controlled entities over which Blacks had no control, and Black households were severely limited in their residential choice.[10] Therefore, there is presumably little endogeneity in the resource allocation in the sense that Black parents or administrators could deliberately change resource endowments. Furthermore, the system generated marked disparities in class sizes, with average class sizes varying across districts between 20 and more than

---

[10] As is standard in the literature, the terms Black and White stand for the racial classification introduced under apartheid; following Case and Deaton (1999), I use capitals to register the specialized use.

80 students per class. This setting specific to South Africa during apartheid can be viewed as a quasi-experiment where a substantial variation in resource endowments existed which could not endogenously be altered by parents or the administration. Thus, similar to the studies reported in Section 3, Case and Deaton's (1999) study focuses on exogenous resource variations.

They find statistically significant positive effects of smaller classes for Black students' achievement in numeracy tests, while the effects on literacy test scores are statistically insignificant, very small, and sometimes even negative. They also analyze the effects of the availability of specific facilities as further resource variables. For Black students, the presence of a secondary school library has a statistically significant positive effect on literacy scores, while the effect on numeracy scores is statistically insignificant. Likewise, the effects of primary school libraries and secondary school laboratories are statistically insignificant in both subjects.

For White students, no statistically significant effects of class size or other resource variables could be observed.[11] The fact that statistically significant effects are confined to Black students is consistent with the view that at the low class sizes which characterize schooling for non-Blacks in South Africa, reductions in class size have little or no effect on student performance. That is, Case and Deaton's (1999) findings point to potential positive resource effects only at very low levels of resource endowment.

## 5 International Cross-Section Evidence

While the within-country cross-section evidence on performance effects of resources might be flawed because of the endogeneity problem, the following two Sections present new evidence based on different resource variations. This Section looks at differences in school resources and educational performance *across countries*, which are not plagued by the within-country cross-section endogeneity problem. This is because there is no mechanism which would move resources across countries in response to differences in the educational performance of students. While resources may vary due to

---

[11] The statistical insignificance of the effects on Whites' test score achievements might be caused by the small sample size for the sample of White test score results. However, in addition to the effects of resources on test scores, Case and Deaton (1999) also analyze the effects of resources on educational attainment and enrollment, where the sample size is large also for White students. There, as well, they tend to find positive estimates for Black schools, while not for White schools.

adjustments made by policy makers, school administrators, and parents in response to students' performance within each country (cf. Section 3.1), no country will transfer sizable amounts of its educational resources to another country because students there show inferior performance. Nor is there any large-scale residential choice across countries due to international differences in educational spending.

### *5.1 Evidence from Previous Cross-Country Tests of Student Performance*

Two studies by Hanushek and Kimko (2000) and Lee and Barro (2001) have examined international student achievement tests, albeit excluding the newest and most extensive one, namely the Third International Mathematics and Science Study which was conducted in 1994/95. Both studies estimate the effect of school resources on student performance in country-level education production functions which control for measures of family background. As their measure of educational performance, Hanushek and Kimko (2000) use the normalized country test scores on six separate international tests. They find that international variations in schooling resources do not have strong effects on international differences in test score performance. The estimated effects of resource measures on educational performance are actually mostly significant with the wrong sign, and sometimes statistically insignificant. This finding holds regardless of the specific measure of schooling resources used, be it expenditure per student, teacher-student ratios, or a variety of other measures.

Lee and Barro (2001) pool evidence from thirteen subtests of international achievement tests, disaggregated by subject, year, and age of students. They find a statistically insignificant negative estimate for expenditure per student, while the effect of teacher-student ratios is statistically significantly positive. However, Lee and Barro's (2001) results are based on data of limited quality, as they use the average teacher-student ratio in primary schools of a country as resource variable, while their test scores mainly reflect performance in secondary education. They further report a positive effect of average teacher salary which is statistically insignificant in their more stringent specification and a statistically insignificant effect of the length of school days.

In sum, there is not much evidence in the cross-country data which would suggest that international differences in educational performance are closely related to differences in the amount of schooling resources available. Additionally, it should be noted that the explanatory power of the models estimated in both of the mentioned studies is very low: Hanushek and Kimko (2000) report an adjusted $R^2$, i.e. the proportion of the variation in test scores explained by all the explanatory variables together, of between 0.19 and 0.26 for different estimations, and the average $R^2$ of Lee and Barro's (2001) panel estimation is 0.23. This suggests that cross-country differences in the performance levels of students are at best only weakly related to measures of family background and schooling resources.

*5.2 International Variation in School Resources and TIMSS Results*

The previous studies did not consider the latest, largest, and most coherent cross-country student achievement study, the Third International Mathematics and Science Study (TIMSS). Figure 4 shows the relationship between educational expenditure per student and TIMSS results in secondary school (seventh and eighth grade). Again, cross-country differences in expenditure per student do not seem to help in understanding cross-country differences in educational performance. Schooling resources and student performance appear to vary independently from each other. The simple correlation between expenditure per student and average TIMSS test scores in the middle school years across the 39 education systems with available data is a mere 0.16.

In order to compare educational expenditure across countries, expenditure in local currencies (available from UNESCO Statistical Yearbooks) had to be converted to a common scale. For Figure 4, purchasing power parity (PPP) conversion factors (from World Bank 1999) were used to transform local currency units into international dollars. This procedure would lead to unbiased international results if the deflator for educational spending was equal to the deflator for the broad basket of goods and services entering the PPP calculations. However, schooling is a labor-intensive service sector without substantial technological progress, while other sectors of the economy exhibit substantial capital deepening and technological advances during economic development. Therefore, the schooling sector will face the cost disease of services, i.e. increasing relative costs with rising levels of economic development, because in an efficient labor market teachers' salaries will rise in line with increasing labor productivity in the other sectors of the economy (Baumol 1967). As a consequence, education should be cheaper in poor countries than in rich countries. The variation on the horizontal axis for an unbiased measure of the relative resource endowments in different countries will thus probably be smaller than shown in Figure 4, because rich countries' educational resources are overvalued with a PPP conversion factor relative to poor countries' educational resources. However, this does not change the basic finding of Figure 4 that educational resources and student performance are unrelated across countries.

One possible way to account for this cost-disease effect is to deflate educational expenditure of each country by the respective economy-wide labor productivity, because the latter can serve as a proxy for the wage level in the economy and thus for the wage level in the education sector. When calculating real educational expenditure per student by dividing the nominal expenditure per student through the average labor productivity in the economy rather than using a PPP conversion factor, the conclusion that there is no relationship between expenditure per student and educational performance remains unchanged. The correlation between the measure of educational expenditure per student deflated by economy-wide labor productivity and average TIMSS test scores is 0.17 in secondary schools.

While educational expenditure has the advantage of being the most encompassing measure of resource availability, average class size has the advantage of being a resource measure which is directly comparable across countries. However, international variations in the quality of teachers are not captured by this measure, while they should be reflected in educational expenditure through differential spending on teachers. Relating TIMSS test results to average class size gives a picture very similar to Figure 4. The simple correlation between class size and TIMSS test scores in secondary schools is -0.21 (based on 34 countries).[12] That is, there is also virtually no relationship between resources and performance across countries based on the class-size measure.[13]

---

[12] Some caution has to be given to the TIMSS measure of average class size, however, because the fraction of classes for which class size was not reported is fairly high in some countries.

[13] Similar results apply for TIMSS results in the primary school years (third and fourth grade), where data is available for only 25 countries. The simple correlation between PPP-adjusted expenditure per student and average primary-school TIMSS test scores is 0.13, and it is 0.02 when the labor-productivity-deflated expenditure measure is employed. Using class size as the resource measure, the correlation coefficient is 0.06 (based on 23 countries); i.e., in primary schools, smaller classes are slightly negatively related to student performance.

Table 1 reports the results of estimations of parsimoniously specified country-level education production functions for TIMSS mathematics and science performance in secondary education. After controlling for parents' educational background and three schooling policy variables - indicators of the extent of school autonomy in supply choice, of teacher unions' influence, and of the scrutiny of assessment -, average class size is statistically significantly positively related to both performance measures, suggesting that larger rather than smaller classes go hand in hand with better student performance.[14] If expenditure per student is used as the resource measure, the effect is statistically insignificant and negative both in mathematics and in science, regardless of the choice of the PPP-adjusted or the labor-productivity-deflated expenditure measure.

In sum, the international evidence on resource effects based on TIMSS corroborates the finding that a strong and systematic relationship between increased educational expenditure and higher student performance is lacking across countries. Neither in the previous cross-country tests of student performance nor in TIMSS are better test results of students related to higher expenditure levels on schooling. The large international differences in student performance levels in mathematics and science are a fact, and their occurrence cannot be explained by differences in the amounts of resources used.

## 6 Within-Country Time-Series Evidence

A second new piece of evidence on the relationship between educational resource endowment and students' educational performance is to look at time series of resource and performance data. Within-country time-series data can reveal whether changes in school resources over time have an effect on student performance. Gundlach et al. (2001) and Gundlach and Wößmann (2001) derive measures of educational expenditure and performance which can be traced over time for several OECD and East Asian countries.

### *6.1 Changes in Spending on School Resources over Time*

Changes in the resource endowment of the schooling systems are measured as real changes in schooling expenditure per student. Using data on schooling expenditure and

---

14 For definitions of and greater detail on the data, as well as for discussion of the results on the non-resource variables, see Wößmann (2001).

students from various issues of the UNESCO Statistical Yearbook, changes in schooling expenditure per student in nominal terms can readily be calculated. One major obstacle to comparing schooling expenditure over time is that price deflators for the schooling sector are needed to separate changes in real resources from price changes. However, specific price deflators for the schooling sector are not readily available in the official statistics, and economy-wide price deflators may not be directly applicable to the schooling sector because of the "cost disease" of services (cf. Section 5.2). To allow for the calculation of changes in real schooling expenditure per student over time, price deflators for the schooling sector have to be derived.

In the derivation of a price deflator for the schooling sector, it is important to note that differences in productivity growth across the sectors of an economy will result in changes in relative prices. Since schooling as a labor-intensive service sector may be assumed to exhibit constant or only relatively slowly growing productivity, differential productivity growth across sectors poses a special problem for time-series research on schooling resources. As shown in Wößmann (2001), this reasoning of Baumol's (1967) cost-disease model can be used to derive price deflators for the schooling sector which can be calculated on the basis of data on the price deflator for the GDP, on changes in economy-wide labor productivity, and on price deflators for other service sectors with presumably constant productivity.

By deflating the nominal changes in educational expenditure per student by these price deflators for the schooling sector, average annual growth rates of real school resources per student can be calculated for a sample of OECD countries in 1970-1994 and for a sample of East Asian countries in 1980-1994. This data is plotted on the horizontal axis of Figure 5. In the OECD sample, real schooling expenditure per student have risen by between 1.2 percentage points per annum in Sweden and 5.4 percentage points per annum in Australia.[15] That is, real resources per student available in the schooling sector have expanded substantially. Likewise, in the sampled East Asian countries real expenditure on schooling per student also increased by an order of magnitude in all cases except for the Philippines.

---

[15] The reported values are simple averages of three estimates which use price deflators for schooling calculated on the basis of three different methods.

These figures show that the amount of resources per student available in the schooling sectors of many OECD and East Asian countries increased substantially, and that there seem to be large differences in the change in schooling resources per student across countries. For example, the average annual increase in real schooling expenditure per student in Australia of 5.4 percent means that real resources per student rose by 250 percent in Australia over the 24 years from 1970-1994. Real schooling expenditure per student nearly tripled in France, about doubled in Germany and Japan, and increased by about 40 percent in the United States. Four of the five sampled East Asian countries experienced similar increases in real resources per student over a time span of only 14 years. Real schooling expenditure per student increased by about 150 percent in South Korea and Thailand from 1980-1994, and by more than 50 percent in Hong Kong and Singapore. Exceptions with only slight increases in real school resources are the Philippines, where real educational expenditure per student stayed largely constant in 1980-1994, and Sweden and the Netherlands, where the increase was relatively small in 1970-1994.[16]

*6.2 Changes in Student Performance over Time*

Real schooling expenditure per student increased substantially in most of the sampled OECD and East Asian countries. The question of interest is now whether this vast expansion of schooling resources per student led to an improvement in students' educational performance. To answer this question, Gundlach et al. (2001) and Gundlach and Wößmann (2001) derive a measure of the change in student performance in the sampled countries over the sampled time periods. The problem with measuring student performance over time is that consistent time-series data on the cognitive achievement of students are not available from within-country sources for nearly all of the countries. Only in the United States, intertemporally comparable tests of the cognitive achievement of students in mathematics and science were administered which show that the average performance of US students by and large did not change over the period 1970-1994 (US Department of Education 1997).

16 Gundlach et al. (2001) and Gundlach and Wößmann (2001) show that shifts in the structure of educational spending from primary towards secondary education cannot account for the large increases in schooling expenditure per student. Furthermore, the OECD estimates of the increase in real schooling expenditure actually seem to be biased downward due to a structural break in the education data series.

The constant performance of US students can be used as an intertemporal benchmark to compare other countries' performance over time in international student achievement tests which have been performed at different points in time by the International Association for the Evaluation of Educational Achievement (IEA).[17] While a direct comparison of the results at the different points in time is impossible because the test design was not held constant, it is possible to reformat the level and the distribution of test scores in previous cross-country tests in order to derive a measure of the cognitive achievement of students in mathematics and science in OECD and East Asian countries which can be traced over time relative to the constant performance of US students.

As the results depicted on the vertical axis of Figure 5 show, the performance of students in mathematics and science did not change much within the sample of OECD countries in 1970-1994.[18] No OECD country achieved a sizable increase in the average performance of its students in 1970-1994. While there may have been a slight increase in the average cognitive achievement of students in the Netherlands and in Sweden, and probably constant performance in Italy, all other countries in the OECD sample seem to have faced a decline in student achievement in mathematics and science. Similarly, no East Asian country achieved a major increase in average student performance in 1980-1994. The performance of South Korean and Singaporean students slightly increased, while the performance of students in Hong Kong and Thailand slightly decreased, and student performance in the Philippines seems to have deteriorated substantially. Converting the index depicted in Figure 5 into annual changes reveals that in no country did the index of student performance increase by more than 0.2 percentage points per annum.

Bringing the time-series resource and performance evidence together, it is fair to conclude that substantial increases in real schooling expenditure per student did not lead to improvements in student performance in most of the sampled OECD and East Asian countries. In the OECD sample, only the moderate increases in real per-student expenditure on schooling in Sweden and the Netherlands were accompanied by an

---

[17] The IEA's Third International Mathematics and Science Study (TIMSS) is the latest of these cross-country studies (cf. Section 5.2).

[18] The reported index of student performance is a simple average of three different methods based on three sets of hypotheses on the evolution of the level and distribution of international test scores in mathematics and science.

increase in the measure of student performance. Other OECD countries in the sample experienced substantial increases in real resource endowment per student, but with the exception of Italy, the average performance of their students in cognitive achievement tests declined. That is, in the vast majority of countries, the large increases in educational expenditure per student did not trigger an increase in student performance. Likewise, in East Asia, real resources per student available to the schooling systems in Hong Kong, Singapore, South Korea, and Thailand expanded substantially, while the average performance of students remained largely constant. As shown in Gundlach and Wößmann (2001), the resource expansion in these countries seems to result from government decisions to increase the amount of schooling inputs by raising the number of teachers per student. The exception among the East Asian countries is the Philippines, where the teacher-student ratio actually decreased. But the effectiveness of resource use in schooling also declined in the Philippines because the performance of students deteriorated substantially while the overall per-student resource level increased slightly.

The time-series evidence from many countries suggests that a positive effect of increased school resources on the educational performance of students is missing also over time. The schooling sectors in most OECD and East Asian countries have seen a substantial increase in available resources over the last 15 to 25 years, while sizable improvements in the educational performance level of students can hardly be detected.

## 7 The Missing Resource-Performance Relation in Education

The general conclusion which emerges from the different strands of evidence on school resources and students' educational performance is that there is no clear relationship between the two. The link between resources and performance in education is still missing. There may be some controversy in the literature on within-country cross-section evidence about whether the absence of any statistically significant positive input-output relation can be taken at face value, and there certainly are circumstances where expenditure does matter for student performance, especially in developing countries where schools have to work at very low levels of resource endowment. However, a uniform across-the-board expansion of the amount of resources available to the schooling system does not seem to render significant positive effects on student performance at the expenditure levels currently prevailing in developed countries.[19] Furthermore, it is obvious that international differences in educational resources do not suffice as an explanation for the large international differences in student performance levels in mathematics and science. And the aggregate within-country time-series evidence shows that large-scale increases in real expenditure per student over the last 15 to 25 years did not trigger advancements in students' educational performance in most OECD countries and in selected East Asian countries.

This range of evidence implies that just providing more resources is unlikely to improve student performance if future actions of schools follow their past behavior. While some schools seem to make good use of additional resources, others do not. In total, schools do not seem to be economically efficient in transforming resources into educational performance. In the schooling systems as presently constituted in most countries, and at the resource levels currently reached in the developed world, a general increase in school resources does not seem to promise significant positive improvements in student performance.

Furthermore, it must be stressed that resource policies are costly. Any potential benefit of increased educational spending has to be valued against its costs. The critical question is whether the future productivity of a student with its ensuing benefits is increased enough to warrant the cost of a resource policy. For policy decisions, it is essential to see whether the benefits of any initiative measure up to its costs. The evidence currently available suggests that this is not the case for increased educational spending per student because the costs outweigh even the most favorable estimates of returns by far (Heckman 2000).

It seems that in order to improve student performance, the structure of decision making and the incentives within the schooling sector may have to be changed rather than the level of available resources. The most important and most promising way forward for future research in this area therefore seems to be to look for evidence on the effect of the institutional set-up of the schooling system, since this set-up generates the incentives which drive the behavior of the different groups of agents involved in

---

[19] Some studies which have not been surveyed here focus on other measures of output than cognitive achievement, like future labor market performance (cf. Card and Krueger 1992, Heckman et al. 1996, Case and Yogo 1999), again with inconclusive results on resource effects.

educational production. Why are there differences in the effectiveness of resource use between schools? What is it that makes some schools use their resources efficiently and others not? How do the incentives faced by the different actors in education differ among schools, and how does this affect student performance?

Some interesting contributions in this direction already exist. E.g., evidence by Bishop (1997, 1999) suggests that curriculum-based external exit examination systems have a positive impact on student performance. Hoxby presents evidence suggesting that competition from private schools and competition among public schools exert positive effects on students' educational performance (Hoxby 1994, 2000b), while a huge influence of teacher unions seems to affect educational production negatively (Hoxby 1996). Wößmann (2000) shows that, among other institutional features of the schooling system, the distribution of responsibilities in budgetary, curricular, and personnel affairs between schools and agencies of external control strongly impacts on student performance. More evidence on such institutional effects and their incorporation into a unifying framework seem necessary to allow a more thorough understanding of educational production and ultimately to make sense of the missing link between resources and performance in education.

## References


Akerhielm, Karen (1995). Does Class Size Matter? *Economics of Education Review* 14(3): 229-241.

Angrist, Joshua D., Victor Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114(2): 533-575.

Ashenfelter, Orley, Colm Harmon, Hessel Oosterbeek (1999). A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias. *Labour Economics* 6(4): 453-470.

Barro, Robert J. (2001). Human Capital and Growth. *American Economic Review, Papers and Proceedings* 91(2): 12-17.

Baumol, William J. (1967). Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review* 57(3): 415-426.

Bishop, John H. (1989). Is the Test Score Decline Responsible for the Productivity Growth Decline? *American Economic Review* 79(1): 178-197.

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

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

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

Boissiere, M., J.B. Knight, R.H. Sabot (1985). Earnings, Schooling, Ability, and Cognitive Skills. *American Economic Review* 73(5): 1016-1030.

Boozer, Michael, Cecilia Rouse (2001). Intraschool Variation in Class Size: Patterns and Implications. *Journal of Urban Economics* 49: forthcoming (*NBER Working Paper* 5144, June 1995).

Card, David (1999). The Causal Effect of Education on Earnings. In: Orley Ashenfelter, David Card (eds.). *Handbook of Labor Economics, Volume 3A*. Elsevier, Amsterdam et al.: 1801-1863.

Card, David, Alan B. Krueger (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy* 100(1): 1-40.

Case, Anne, Angus Deaton (1999). School Inputs and Educational Outcomes in South Africa. *Quarterly Journal of Economics* 114(3): 1047-1084.

Case, Anne, Motohiro Yogo (1999). Does School Quality Matter? Returns to Education and the Characteristics of Schools in South Africa. *NBER Working Paper* 7399, October.

Coleman, James S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld, R.L. York (1966). *Equality of Educational Opportunity: Summary Report*. US Government Printing Office, Washington, D.C.

Currie, Janet, Duncan Thomas (1999). Early Test Scores, Socioeconomic Status and Future Outcomes. *NBER Working Paper* 6943, February.

Federal Reserve Bank of New York (1998). Excellence in Education: Views on Improving American Education. Federal Reserve Bank of New York, *Economic Policy Review* 4(1): 1-129.

Gundlach, Erich, Ludger Wößmann (2001). The Fading Productivity of Schooling in East Asia. *Journal of Asian Economics* 12: forthcoming (*Kiel Working Paper* 945, September 1999).

Gundlach, Erich, Ludger Wößmann, Jens Gmelin (2001). The Decline of Schooling Productivity in OECD Countries. *Economic Journal* 111(471): C135-C147.

Hanushek, Eric A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* 24(3): 1141-1177.

Hanushek, Eric A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer* 10(2): 227-246.

Hanushek, Eric A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis* 19(2): 141-164.

Hanushek, Eric A. (1999a). The Evidence on Class Size. In: Susan E. Mayer, Paul E. Peterson (eds.). *Earning and Learning: How Schools Matter*. Brookings Institution Press, Washington, D.C. and Russell Sage Foundation, New York, N.Y.: 131-168.

Hanushek, Eric A. (1999b). Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects. *Educational Evaluation and Policy Analysis* 21(2): 143-163.

Hanushek, Eric A. (2000). Evidence, Politics, and the Class Size Debate. Stanford University, Mimeo, October.

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

Hanushek, Eric A., Steven G. Rivkin, Lori L. Taylor (1996). Aggregation and the Estimated Effects of School Resources. *Review of Economics and Statistics* 78(4): 611-627.

Harbison, Ralph W., Eric A. Hanushek (1992). *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. Published for the World Bank, Oxford University Press, Oxford et al.

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

Heckman, James J., Anne Layne-Farrar, Petra Todd (1996). Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings. *Review of Economics and Statistics* 78(4): 562-610.

Hoxby, Caroline M. (1994). Do Private Schools Provide Competition for Public Schools? *NBER Working Paper* 4978, December.

Hoxby, Caroline M. (1996). How Teachers' Unions Affect Education Production. *Quarterly Journal of Economics* 111(3): 671-718.

Hoxby, Caroline M. (2000a). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics* 115(4): 1239-1285.

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

International Association for the Evaluation of Educational Achievement (IEA) (1998). *Third International Mathematics and Science Study: International Achievement Reports.* Available at: http://timss.bc.edu/timss1995i/Achievement.html.

Journal of Economic Perspectives (1996). Symposium on Primary and Secondary Education. *Journal of Economic Perspectives* 10(4): 3-72.

Krueger, Alan B. (1999a). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114(2): 497-532.

Krueger, Alan B. (1999b). An Economist's View of Class Size Research. Princeton University, Mimeo, December.

Krueger, Alan B., Diane M. Whitmore (2001). The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *Economic Journal* 111(468): 1-28.

Lazear, Edward P. (1999). Educational Production. *NBER Working Paper* 7349, September.

Lee, Jong-Wha, Robert J. Barro (2001). Schooling Quality in a Cross Section of Countries. *Economica* 68: forthcoming (*NBER Working Paper* 6198, September 1997).

Mosteller, Frederick (1995). The Tennessee Study of Class Size in the Early School Grades. *The Future of Children* 5(2): 113-127.

Murnane, Richard J., John B. Willett, Frank Levy (1995). The Growing Importance of Cognitive Skills in Wage Determination. *Review of Economics and Statistics* 77(2): 251-266.

Neal, Derek A., William R. Johnson (1996). The Role of Premarket Forces in Black-White Wage Differences. *Journal of Political Economy* 104(5): 869-895.

Psacharopoulos, George (2000). Economics of Education à la Euro. *European Journal of Education* 35(1): 81-95.

Review of Economics and Statistics (1996). Symposium on School Quality and Educational Outcomes. *Review of Economics and Statistics* 78(4): 559-691.

Rivera-Batiz, Francisco L. (1992). Quantitative Literacy and the Likelihood of Employment Among Young Adults in the United States. *Journal of Human Resources* 27(2): 313-328.

Sander, William (1999). Endogenous Expenditures and Student Achievement. *Economics Letters* 64(2): 223-231.

Tyler, John H., Richard J. Murnane, John B. Willett (2000). Do the Cognitive Skills of School Dropouts Matter in the Labor Market? *Journal of Human Resources* 35(4): 748-754.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (var. iss.). *Statistical Yearbook.* UNESCO, Paris.

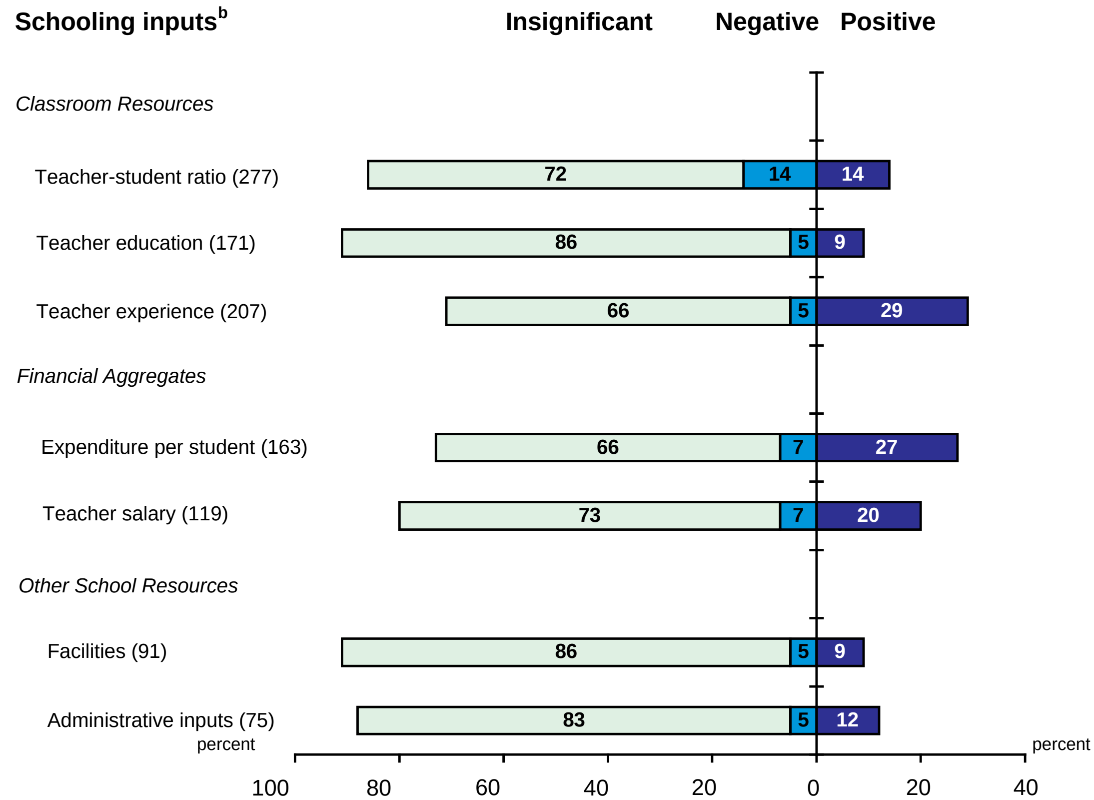
US Department of Education, National Center for Educational Statistics (1997). *The Condition of Education 1997*. US Government Printing Office, Washington, D.C.

World Bank (1999). *World Development Indicators CD-Rom*. Washington, D.C.

Wößmann, Ludger (2000). Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. *Kiel Working Paper* 983, December.

Wößmann, Ludger (2001). *Schooling and the Quality of Human Capital*. Springer: forthcoming.

**Figure 1: Estimated Effects of School Resources: United States**[a]

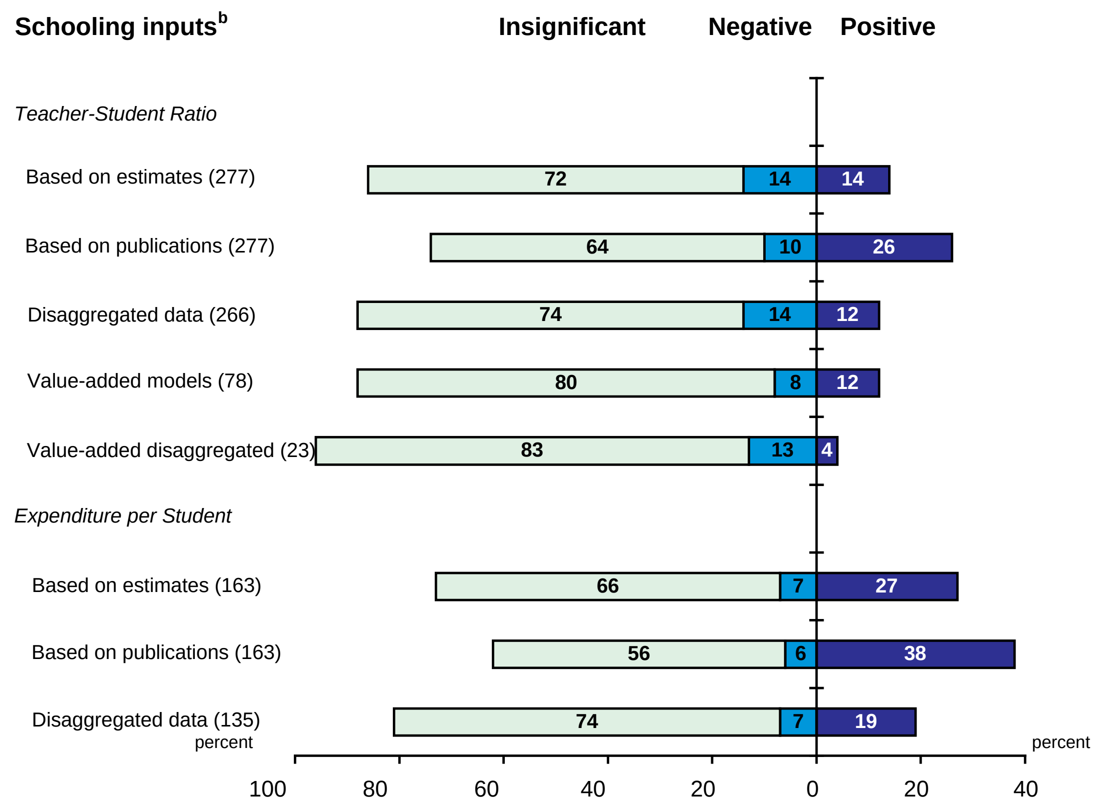



[a] Summary of selected regression coefficients on various resource measures estimated by education production functions (dependent variable: student achievement). Percentage of statistically insignificant (light grey), statistically significant negative (dark grey), and statistically significant positive (black) regression coefficients, respectively.

[b] Number of estimates in parenthesis.

Source: Adapted from Hanushek (1997, 2000).

**Figure 2: A Further Look at the US Evidence**[a]

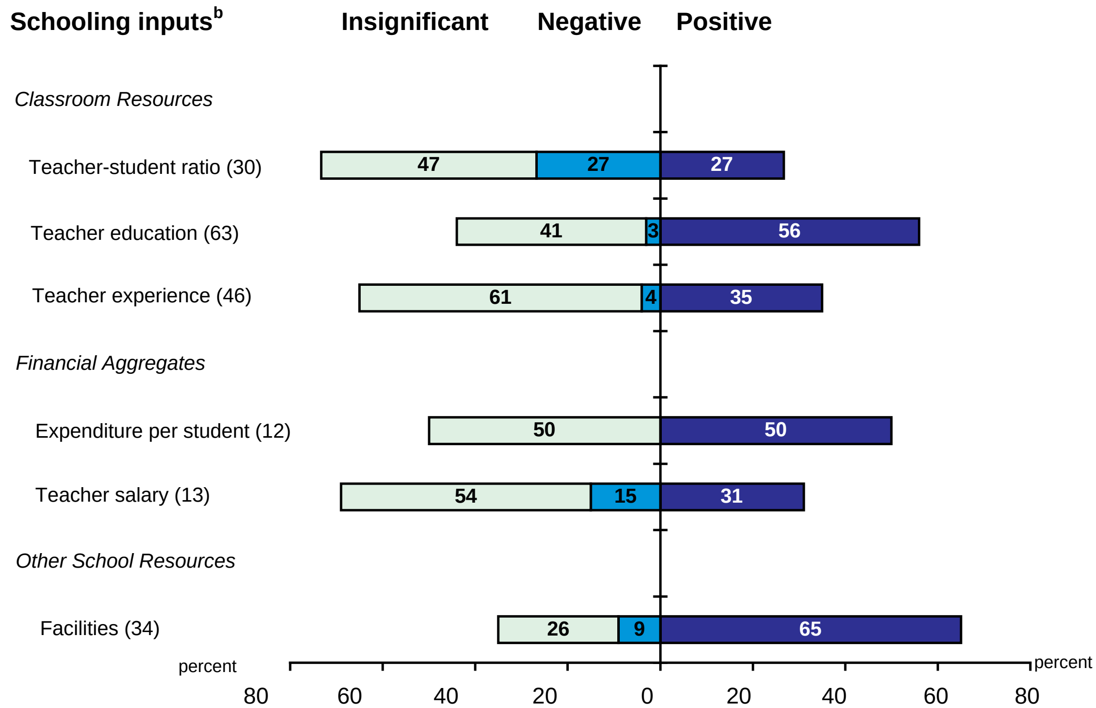



[a] Summary of selected regression coefficients on various resource measures estimated by education production functions (dependent variable: student achievement). Percentage of statistically insignificant (light grey), statistically significant negative (dark grey), and statistically significant positive (black) regression coefficients, respectively.

[b] Number of estimates in parenthesis.

Sources: Adapted from Krueger (1999b), Hanushek (2000), and own calculations.

**Figure 3: Estimated Effects of School Resources: Developing Countries**[a]

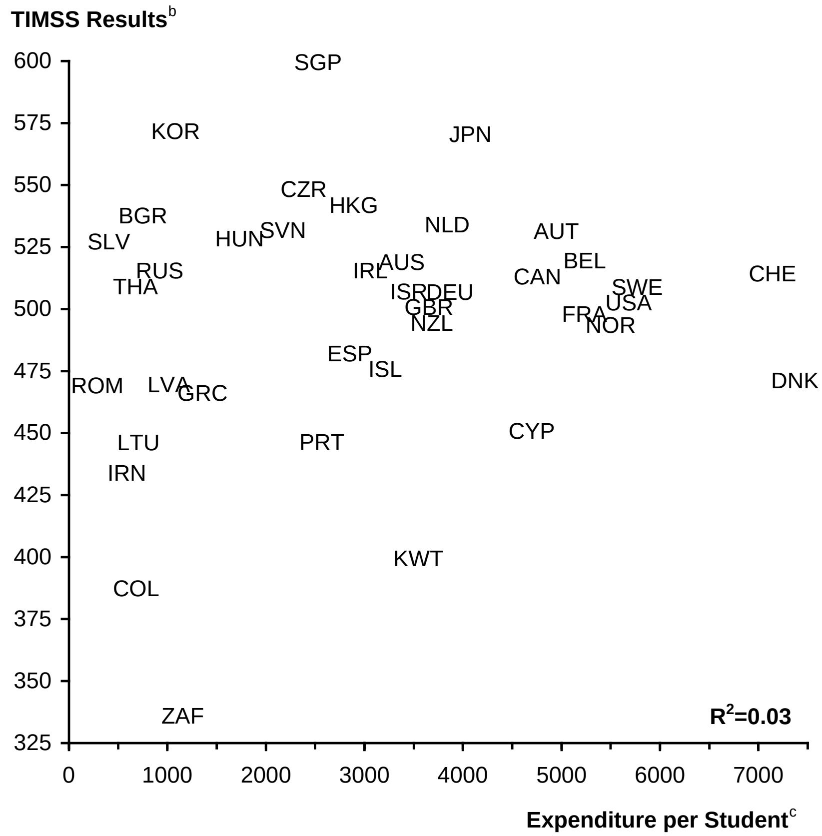



[a] Summary of selected regression coefficients on various resource measures estimated by education production functions (dependent variable: student achievement). Percentage of statistically insignificant (light grey), statistically significant negative (dark grey), and statistically significant positive (black) regression coefficients, respectively.

[b] Number of estimates in parenthesis.

Source: Adapted from Harbison and Hanushek (1992).

**Figure 4: Expenditure per Student and Educational Performance: The TIMSS Cross-Country Evidence in Secondary Schools**[a]



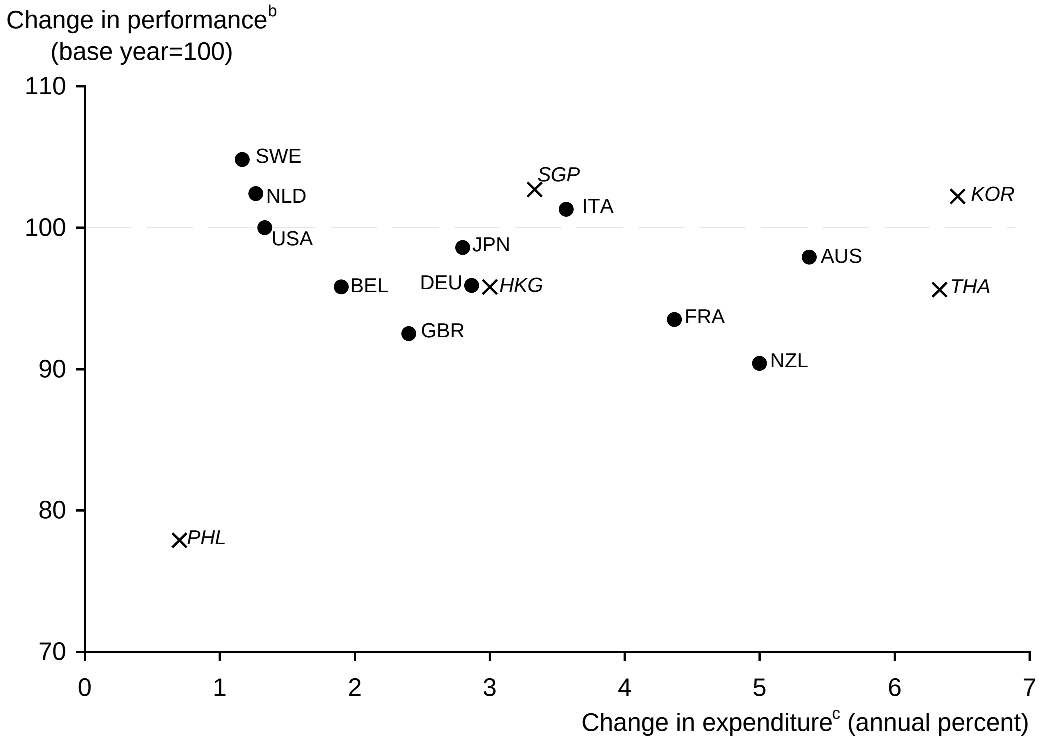

[a] The country abbreviations stand for: Australia (AUS), Austria (AUT), Belgium (BEL), Bulgaria (BGR), Canada (CAN), Colombia (COL), Cyprus (CYP), Czech Republic (CZR), Denmark (DNK), France (FRA), Germany (DEU), Greece (GRC), Hong Kong (HKG), Hungary (HUN), Iceland (ISL), Iran (IRN), Ireland (IRL), Israel (ISR), Japan (JPN), Kuwait (KWT), Latvia (LVA), Lithuania (LTU), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (PRT), Romania (ROM), Russian Federation (RUS), Singapore (SGP), Slovak Republic (SLV), Slovenia (SVN), South Africa (ZAF), South Korea (KOR), Spain (ESP), Sweden (SWE), Switzerland (CHE), Thailand (THA), United Kingdom (GBR), United States (USA).

[b] Average of TIMSS international mathematics and science scores in seventh and eighth grade.

[c] Expenditure per student at the secondary level, in international dollars, 1994.

Sources: Own calculations based on IEA (1998), UNESCO (var. iss.), and World Bank (1999).

**Figure 5: Changes in School Expenditure and Student Performance in OECD Countries (1970–1994) and in East Asian Countries (1980–1994)[a]**




[a] The country abbreviations stand for: 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).

[b] Change in students' average educational performance in mathematics and science.

[c] Real change in school expenditure per student.

Sources: Based on Gundlach et al. (2001), Gundlach and Wößmann (2001), and Wößmann (2001).

## Table 1: Country-Level Education Production Functions

Dependent variable: TIMSS international mathematics/science test score.
Standard errors in parentheses.

| **I. Mathematics** | Coefficient | Standard Error | Standardized Coefficient |
|---|---|---|---|
| Constant | 144.424† | (63.234) | |
| Class size | 3.873* | (0.720) | 0.524 |
| Parents' education: finished secondary | 176.271* | (23.308) | 0.725 |
| Parents' education: beyond secondary | 91.367* | (22.557) | 0.429 |
| School autonomy in supply choice | 98.464‡ | (49.381) | 0.170 |
| Teacher unions' influence | -467.790* | (90.000) | -0.455 |
| Scrutiny of assessment | 26.911* | (6.610) | 0.355 |
| Observations | 39 | | |
| F | 19.31 | | |
| $R^2$ (adj.) | 0.74 | | |

| **II. Science** | Coefficient | Standard Error | Standardized Coefficient |
|---|---|---|---|
| Constant | 234.843* | (62.028) | |
| Class size | 2.557* | (0.856) | 0.370 |
| Parents' education: finished secondary | 136.873* | (23.448) | 0.705 |
| Parents' education: beyond secondary | 60.989† | (23.174) | 0.359 |
| School autonomy in supply choice | 106.307† | (48.440) | 0.230 |
| Teacher unions' influence | -369.327* | (92.859) | -0.450 |
| Scrutiny of assessment | 12.428‡ | (6.963) | 0.197 |
| Observations | 39 | | |
| F | 10.49 | | |
| $R^2$ (adj.) | 0.60 | | |

* Significant at the 1 percent level.
† Significant at the 5 percent level.
‡ Significant at the 10 percent level.