Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Economics of Education Review 24 (2005) 579–599
Economics of Education Review
www.elsevier.com/locate/econedurev

Schooling quality in Eastern Europe: Educational production during transition

Andreas Ammermüller[a,*], Hans Heijke[b], Ludger Wößmann[c]

[a]Centre for European Economic Research, P.O. Box 103443, 68034 Mannheim, Germany
[b]Research Centre for Education and the Labour Market and Maastricht University, P.O. Box 616, 6200MD Maastricht, The Netherlands
[c]Ifo Institute for Economic Research at the University of Munich and CESifo, Poschingerstr. 5, 81679 Munich, Germany

Received 20 April 2003; accepted 21 August 2004

Abstract

This paper uses student-level Third International Mathematics and Science Study (TIMSS) data to analyze the determinants of schooling quality for seven Eastern European transition countries by estimating educational production functions. The results show substantial effects of student background on educational performance and a much lower impact of resources and the institutional setting. Two different groups of countries emerge. For the first group that features high mean test scores and has progressed far in transition, large effects of family background on student performance and a higher spread of test scores illustrate the similarity to Western European schooling systems, the performance of which it surpasses. Schools of the second group produce instead a denser distribution of educational achievement, characteristic of communist societies.

JEL Classification: I21; P36

Keywords: Educational economics; Input–output analysis; Productivity

1. Introduction

This paper analyzes and compares the production of schooling quality in seven Eastern European transition countries striving for EU accession. A main focus of the progress in the transition countries is on reforms of institutions and a changing structure of society. The supposedly egalitarian societies in which mainly the party rank defined the social position are replaced by a new social distinction based on occupation and income. Education is the major vehicle through which the societal changes take place. The function of education shifts from keeping the social consensus of a classless society to allocating its individual members to economic roles and positions, allowing for greater differentiation by increased educational choice (Heyneman, 1997). The readiness for and speed of transition depends therefore essentially on reforms in education, which prepare people for their new roles in society. Reforms include the decentralization of the educational system, which might increase its effectiveness by increasing its responsiveness to market forces at the local and national level. A greater choice of different types of institutions for students and an increased influence of parents on their children's education might as well result in higher effectiveness in the new economic terms (Heyneman,

*Corresponding author. Tel.: +49 621 1235284; fax: +49 621 1235225.
E-mail address: ammermuller@zew.de (A. Ammermüller).

0272-7757/$ - see front matter
doi:10.1016/j.econedurev.2004.08.010

1997). The introduction of decentralization and free choice may also serve to convey the values of democracy and the market system to the population (EBRD, 2000).

Furthermore, the market economies demand new abilities of students, like managing skills and high flexibility that were not fostered under the communist regime (Berryman, 2000). In the phase of transition, where old practices become obsolete and new opportunities arise quickly, allocative skills like the ability to take appropriate decisions, which constitute an important effect of schooling (Schultz, 1975), are rewarded greatly and affect the income distribution. Consistent with this reasoning, rates of return to education have been found to increase during the transition from communism to a market system (Newell & Reilly, 1999; Boeri & Terrell, 2002). For example, the returns to higher education more than doubled over the past 10 years in the Czech Republic, as did their spread across occupations (Klazar, Sedmihradský & Vančurová, 2001). The quality and variance of educational achievement may thus have an even bigger impact on the societal structure in transition countries than in advanced countries. When the Eastern European countries gain access to the EU and the labor markets become integrated in the coming years, they need to compete with the Western European labor force. A well-educated work force is hence imperative for a successful integration.

The challenges are accompanied by new threats to the formerly high-performing education systems. Due to economic recessions in the early phase of transition, the level of finance for schools is hard to maintain. This problem relates particularly to the countries whose setbacks in the first years of transition were greater and who continue to struggle in their reform progress.

The available resources, the institutional setting of schools and especially their usage depict the quality of the schooling systems. The main goal of this paper is to examine the impact of these factors, as well as of student characteristics and family backgrounds, on the performance of individual students by estimating educational production functions for seven Eastern European countries. While the former factors are determined by school policy, the latter display the ability of schools to diminish the impact of the environment surrounding students. By estimating production functions of many Eastern European schooling systems for the first time, the paper intends to elaborate on the determinants of schooling quality during the phase of transition. In addition, it contributes to the widely discussed topic of the effect of resources. We also compare the characteristics and the results of the production functions for the Eastern European schooling systems to those of a sample of EU member countries.

Educational production functions relate an outcome of education like educational achievement to various inputs. In this study, standardized test scores from the Third International Mathematics and Science Study (TIMSS), which are comparable across countries, are related to variables drawn from background questionnaires on student background, resources employed and the institutional setting (cf. Wößmann, 2003). These unique micro-level data on test scores and background information are available on over 42,000 students in the 7th and 8th grade of the seven countries, who took the math and science tests in 1994/95. As the only measurable outcome taken into account is educational achievement, the analysis is restricted to the cognitive dimension of schooling. Other educational outcomes like civic values conveyed to students have to be neglected in the data analysis.

The main finding of the paper is that a distinction can be made between two groups of accession countries with respect to the characteristics of the schooling systems, which constitute a decisive factor during the transition process. One group has moved decisively towards the features of Western European countries while the other cannot display successful results of transition yet. The more advanced group, consisting of the Czech and Slovak Republics, Hungary and Slovenia, outperforms most EU countries and has many traits similar to the Western schooling systems. The schooling systems of the less advanced group, including Lithuania, Latvia and Romania, still feature characteristics of communist times. Further findings of the paper are the relative importance of student background for explaining test scores, the ambiguous impact of resources, and the limited but existing role of the institutional setting in understanding within-country variations in test scores.

The remainder of the paper is structured as follows. The second section describes the seven transition economies and characterizes the development of their schooling systems in the 1990s. This qualitative review facilitates an assessment of the data and results and elaborates on particular characteristics that are not implied in the variables. The third section describes the TIMS study, the data for the Eastern European countries, and the model used for the estimation of the production functions, including discussions of the advantages and limitations of the cross-sectional data and methods employed. The fourth section presents and discusses the results for the Eastern European countries. The fifth section compares them to a sample of Western European countries. Finally, the sixth section concludes with a summary of the findings and an assessment of their contribution and relevance in the context of transition economies.

2. Review of the schooling systems in Eastern Europe

Seven countries that belong to the group of transition countries and are EU accession countries participated in

Table 1
Transitional progress and education systems in the Eastern European countries

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Population in '000[a]	10,264	10,106	1930	5415	3611	2385	22,364
GDP/capita US$ 1994[b]	3977	4052	7231	2721	1143	1442	1323
GDP/capita US$ 2000[b]	4797	4552	9073	3556	3064	3019	1644
Estimated GDP level in 2000, 1989 = 100[b]	98	104	114	103	65	64	77
Ed. exp. per student US$ in 1994[c]	671	840	1492	319	195	283	115
Student pop. in '000[d]	1146	1360	189	804	512	299	2461
Enrollment in 1994[e]	99.5%	99.1%	96.7%	97.0%	92.2%	89.0%	91.4%
No. of school types for lower secondary[f]	2	2	1	4	5	1	1
Average attendance rate in pre-primary[f]	86%	86%	59%	82%	40%	23%	65%

[a]Mid 2001 estimates, CIA Factbook Country Profiles, http://www.cia.gov/cia/publications/factbook/.
[b]EBRD Transition Report 2001, Country Assessments and Table A3.1.
[c]Total educational expenditure in 1994 US$ divided by population of age 3–24 in 1994, UNESCO Statistical Yearbook 1997 and Berryman (2000).
[d]No. of children of compulsory school age in 2000, Eurybase, http://www.eurydice.org.
[e]Enrollment rates for basic education, ages 6/7–14/15 in 1994, Berryman (2000), Table A6.
[f]School types in 1994, attendance rate age 3–5 for 2000. See Eurybase.

TIMSS 1995. They include the Central and Eastern European countries Czech Republic, Hungary, Romania, Slovakia and Slovenia and the Baltic States Latvia and Lithuania. Table 1 presents information on the transition progress and level of development for each country. The figures on GDP indicate the exceptional status of Slovenia, featuring almost twice the Czech and Hungarian level of GDP per capita. Slovakia is slowly catching up to the three frontrunners. The Baltic States and Romania, the biggest country and laggard in recent years, belong to the lowest developed countries of the sample. According to the similarities in the countries' development, which might also affect the transitional states of the schooling systems, two groups of countries are formed. The countries Czech Republic (CSK), Hungary (HUN), Slovakia (SLV) and Slovenia (SVN) will be referred to as the first group from now on. The second group comprises the remaining three countries Latvia (LVA), Lithuania (LTU) and Romania (ROM). The grouping is also motivated later on based on the results.

With the retreat of the communist ideology in Eastern Europe in 1990, quick and fundamental reforms of the educational systems followed. Administration of schooling was decentralized, leaving multiple authorities a say on education, including parents and the church. Additionally, national schools were established, especially in the Baltic States, to foster national culture and language. Moreover, the heterogeneity of schools increased, changing from a system of only one basic school to more specialized institutions like the gymnazium or technical and vocational schools and leading as well to the development of private schools (cf. Filer & Münich, 2000). For lower secondary education, where TIMSS took place, the choice of different types of schools is limited to one form in Latvia, Romania and Slovenia, two in the Czech Republic and Hungary, four in Slovenia and five in Lithuania. Policies also aimed at decentralization and more heterogeneity of the system within the single school types, by setting up special ability classes for both low- and high-performing students.

3. Data and regression technique

3.1. The TIMSS data

TIMSS was conducted in 1994/95 by the International Association for the Evaluation of Educational Achievement. Over 40 countries worldwide participated in TIMSS, making it the largest and most complex achievement study ever conducted (Gonzalez & Smith, 1997). This paper considers only the sample Population II with students from the middle school years. It comprises students from two adjacent grades who have the largest proportion of 13-year-old students. They correspond to the 7th and 8th grade in lower secondary schools in the seven countries considered here. For the analysis of Eastern European countries, data for over 42,000 students from more than 1000 schools are available.[1]

The students took standardized achievement tests in both mathematics and science. The results were scaled according to an international test score with a mean of

[1]For more detailed information on TIMSS, see the internet homepage http://timss.bc.edu.

500 and an international standard deviation of 100. In addition, each student, his teacher and principal had to complete a questionnaire giving background information on students, the community, resources of the school including teacher characteristics and the institutional setting. All available data for an individual student were merged in one database (Wößmann, 2003) together with his sampling weight. Table A1 provides data on the participation at student, class and school level and the ratio of sampled students. Schools in geographically remote regions, extremely small schools and schools for students with special needs were excluded from the study, as were disabled students in regular schools. This might have led to a bias of the sampled students, especially when many disordered but not disabled students have been placed into special schools. However, all other students could be sampled and the exclusion rate was not to exceed 10%.

As the questionnaires handed to students and teachers comprise a multitude of questions, the problem of missing data is inevitable. In order to prevent a selection bias by ignoring all observations with incomplete data and to keep the sample size high, missing data are imputed. A set of fundamental variables that are available for the greatest part of the students is selected among the explaining variables. In cases where these variables are missing, the average at the lowest level available is taken as an approximation, meaning first the class average, then school average or finally country average. Each of the other explaining variables is then regressed on this set of variables, and missing values in these other variables are substituted by the predicted values from this regression. In the case of qualitative data, the prediction is conducted on the basis of probit and ordered probit models.[2] For the purposes of this paper, the data imputation is conducted separately within each country.[3]

The dataset that is built on TIMSS offers the unique opportunity to analyze and compare the educational systems in the seven transition countries. With the exception of Hungary and Poland, the former communist countries have not previously participated in an international cross-country study on student achievement. The dataset allows for the first estimation of educational production functions for this large group of Eastern European countries. Moreover, the quality of the available data with the immense background information on various inputs and the quantity, with data available for between 4976 and 7471 students in the individual countries, allow for a very thorough analysis. The sampling design to test two adjacent grades of each school also permits to estimate the between-grade variation, as will be discussed later on in the identification strategy.

An obvious limitation of each cross-sectional study is that data are only available for one point in time. It is therefore impossible to control for prior educational achievement and to consider the value added of a school year to students separately. Instead, the data compare the level of student achievement. Further possible limitations are missing variables for state or regional factors. Omitting them in the regressions might lead to a bias of coefficients. Aggregation of data above the state level is likely to exacerbate this problem (Hanushek, Rivkin, & Taylor, 1996), which does not apply to the micro-level database used here, though.

3.2. Description of country data

This section sketches the different schooling systems by considering the most noticeable mean values of the explanatory variables given in the data and used in the estimated models. The definitions and ranges of value of all 25 variables are given in Table A2, while Table A3 displays their mean values and standard deviations.

A distinction between the two groups is evident in the mean test scores that the students achieved in TIMSS. The four countries of the first group all reach mean scores in both math and science that lie above the international mean of 500. They even accomplish higher scores than most Western European nations, including Denmark, France and Germany. Czech students performed best among the participating transition countries, with average scores of 544 in math and 553 in science. The countries of the second group instead all scored below the international mean of 500, with Lithuania being the worst performer with 454 points in math, superior only to Portugal in Europe. The spread of the test scores is lowest in the low-scoring Baltic States and highest for the Czech Republic, Hungary and also for Romania.

Overall, the parents of the tested students are well-educated, with the minor exception of Romania. The average class size varies within the two groups of transition countries, being lowest in the Baltic States with around 22 students per class and highest in Slovakia and Romania with an average class size of over 26. The share of female teachers is around 80% in the transition countries.

The separation of the countries holds also for the descriptive statistics of other variables. The second group has a higher share of students from separated families, of parents belonging to the lowest educational group and a lower average education, except for Lithuania. The second group suffers more from a shortage of materials in schools and has fewer well-educated teachers. The Czech Republic and Slovakia are

[2] For more details on the imputation method, see Wößmann (2003).

[3] A table specifying the share of missing values for each variable is available from the authors upon request.

instead well endowed with materials and a well-educated teaching staff. The schooling systems are most decentralized in Hungary and Slovenia and still very centralized in Romania.

3.3. Regression models and techniques

3.3.1. The basic specification

To estimate educational production functions for the individual countries at the student level, the following general model will be employed:

$$T_{ics} = \beta_0 + B_{ics}\beta_1 + R_{cs}\beta_2 + I_{cs}\beta_3 + v_s + \varepsilon_{ics}, \quad (1)$$

where T is the math or science test score of student i in class c at school s, B is a set of background variables on the student and her family, R are measures of resources used and teacher characteristics, I is a set of variables reflecting the institutional setting and v and ε are error terms at the school and student level, respectively.[4]

If the assumption held that any factors which are not controlled for by the included explanatory variables and which might therefore enter the error term are not systematically related to the included explanatory variables, least-squares estimation of model (1) would yield estimates of the influence of the different explanatory variables on student performance. The specification measures effects on student performance in levels, rather than in the value added to performance from 1 year to the next. Given that background hardly changes over time, it affects student performance over several years, so that the relevant performance measure is indeed one of levels rather than the value added in a single year. Background can also be reasonably viewed as exogenous to student performance (cf. Wößmann, 2003 for a more detailed discussion).

Furthermore, because the basic school, which includes grades 1–9, still attracts the majority of students in many Eastern European countries, even many resource and institutional features affect student performance over several years. Also, as pointed out by Krueger (1999), the effect of class size may exert itself mainly in the first year in which a student is placed in a smaller class, so that value-added estimations in later years may miss the main effect, while level estimations do not.[5] On the other hand, employing cross-sectional data is a clear limitation because it provides only imperfect measures of the real inputs into the cumulative process of education (Hanushek, 2003). Therefore, employing cross-sectional data to estimate the level model (1) might limit causal interpretation of the obtained estimates.[6] Estimates based on the specific model should rather be interpreted cautiously in terms of descriptive conditional correlations.

The survey design of TIMSS demands specific regression techniques for the estimation of the educational production functions. The sampling design of TIMSS contains both varying sampling probabilities for students from different schools and clustered data. Giving different weights to students who had different sampling probabilities allows obtaining nationally representative coefficient estimates. This is done by applying weighted least squares (WLS) as a regression technique for all regressions performed with the data (cf. Wooldridge, 2001).

The second issue of clustered data is more troublesome. In each country, participating schools were chosen in a first step, and then the classes which took the standardized tests within each school in a second step. Therefore, the primary sampling units (PSU) are not the individual observations, the students, but instead their schools. The problem arises that the observations within the cluster of a school are not independent as they share common characteristics, which cannot be totally controlled for. The error term of the regression may therefore be more complex than assumed by conventional least-squares methods, comprising besides an individual component also class and school elements. Ignoring these latter parts can lead to spurious regression results, as the supposedly independent observations depend on each other.

The method of clustering-robust linear regression (CRLR) offers a solution to the obstacle. It allows for any dependence of observations within the PSU,

[4] The information on individual student test scores and background variables provided by the TIMSS micro-level database allows for a more precise estimation of coefficients and less bias than does aggregated data used for most estimations (Card & Krueger, 1996). Although also using the TIMSS database, Hanushek and Luque (2003) do not employ it at the student level but aggregate it to the classroom level instead.

[5] However, observational data on class sizes may be endogenous to student performance, a point we discuss further in Section 3.3.2.

[6] As a robustness check, we also estimated an approximated value-added model, where we use the mean 7th-grade score in the respective school as a proxy for the lagged performance of the 8th-grade students in this school. In addition to several obvious shortcomings of this approximation, the main problem with this specification is that it would only be appropriate if students in the two grades do not differ systematically in unobserved characteristics. However, Kane and Staiger (2001) have shown that much of the performance difference between the students of two consecutive grades in the same school is actually noise (sampling variation and variation due to non-persistent sources). Despite this shortcoming, the mean 7th-grade score enters statistically significantly positive as an explanatory variable for 8th-grade performance in all countries and subjects. The general pattern of our results and all the main findings are strongly corroborated by this alternative specification. Detailed results are available from the authors upon request.

demanding only independence across PSU. Ignoring the cluster design and treating dependent observations as independent by using standard formulas would result in standard errors that are too small (White, 1980). There is no impact on the coefficients of the parameters, however. Therefore, the standard errors estimated by OLS, or in our case WLS, need to be corrected. The clustering-robust variance–covariance matrix of the coefficient estimates $\widehat{V}(\beta)$ can be approached by

$$\widehat{V}(\widehat{\beta}) = (X'WX)^{-1}\left(\sum_{s} X_s' W_s e_s e_s' W_s X_s\right)(X'WX)^{-1}, \tag{2}$$

where $\widehat{\beta}$ represents the coefficients β_0 to β_3 from model (1), the matrix X all explanatory variables, W is the weight matrix and $e_s e_s'$ are the cluster matrices of the WLS residuals from each cluster e_s (White, 1980, p. 821; see also Deaton, 1997, pp. 73–78).

This formula offers a consistent estimate of the variance–covariance matrix of the WLS estimator, even if the error variances differ across clusters and arbitrary correlation patterns exist within clusters. A supposition is a fixed cluster size as the sample size increases, which is fulfilled because the number of students tested in each school is independent of the overall number of students in the sample. For the estimation of the educational production functions of the individual countries, model (1) will be estimated by CRLR. Hence, the WLS coefficients and clustering-robust estimated standard errors will be presented for the production functions. Whenever merged data of several countries are analyzed, a dummy for each but one country is included in the regressions. This allows for a correlation of error terms within countries, which is likely.

3.3.2. Estimating the effect of resources

While the impact of background and institutional measures on students' educational achievement can reasonably be estimated by the CRLR-level model (1), the impact of a school's resource endowment and its correct identification are hotly disputed topics in the literature on educational production (cf. Hanushek, 2003; Krueger, 2003). The causality of the resource–performance link is ambiguous since the supposedly exogenous resource variables might be influenced by actual performance of the students and might thus be endogenous (cf. Hoxby, 2000). Estimating the effect of resources, especially of class size, on student achievement is therefore burdensome. Inasmuch as the TIMSS data come from an observational study and not an experiment, the coefficient may be biased by the non-random allocation of students to different class sizes, both between schools and within schools. Controlling for these biases is essential in order to obtain a consistent estimate for the effect of resources.

Several mechanisms are imaginable that lead to the non-random allocation of students between schools, depending on the schooling system. Parents may either move to districts that offer smaller classes to their children, or the local school administration might put low-performing students into schools with smaller classes, especially when various types of schools are available. In both cases, between-school sorting takes place and biases the estimator of the class size effect.

One strategy to eliminate all variation between schools is to control for school fixed effects (SFE). For its implementation, a dummy variable D for each school is included in model (1), leading to the model

$$T_{ics} = \alpha D_s + \beta_0 + B_{ics}\beta_1 + R_{cs}\beta_2 + \varepsilon_{ics}. \tag{3}$$

The institutional variables I that are mostly determined at the school level, are not included in this model because the inclusion of the school dummies removes all possible variation between them. This model is referred to as SFE model.[7]

Having controlled for between-school variation and having only between-grade variation left, a potential bias may still stem from within-school sorting. In order to account for the non-random allocation of students to different class sizes within a school, the technique of instrumental variables (IV) is used. Akerhielm (1995) instrumented actual class size with average class size for a given subject in the school and student enrollment at the given grade. The legitimacy of using student enrollment as an instrument is questionable, however, as overall school enrollment may also exert an impact on student achievement (Summers & Wolfe, 1975; Angrist & Lavy, 1999).

In this analysis, actual class size is instrumented by average class size at the grade level. It is highly correlated with actual class and, by assumption, not with the error term. It affects student achievement only indirectly through the impact on actual class size. For the regression, a two-stage least-squares estimation procedure is used. Actual class size S is regressed on average class size at the grade level A, the other exogenous variables Ψ and the school dummies D:

$$S_c = \alpha_1 A_c + \alpha_2 \Psi_{ics} + \alpha_3 D_s + \varepsilon_{ics}. \tag{4}$$

The predicted value $\widehat{S}_c = S_c - \varepsilon_{ics}$ consists of the non-random part S_c and the random part ε_{ics}. Using only the systematic part of S_c, no correlation will exist between $\widehat{S}_c$ and ε_{ics}, allowing the second stage of the regression to produce a consistent estimator for class size:

$$\mathrm{T}_{ics} = \gamma_1 \widehat{S}_c + \gamma_2 \Psi_{ics} + \gamma_3 D_s + \varepsilon_{ics}. \tag{5}$$

[7]The validity of random effects models is rejected by tests showing that the hypothesis of no correlated effects can be rejected at the 1% significance level in each country.

Having controlled for between- and within-school sorting effects, the coefficient γ_1 should then be a consistent estimate of the relationship between class size and student achievement. Model (5) combines the SFE model and the IV technique and is referred to as SFE+IV.[8] The only variation left to explain is within-school between-grade variation.[9] Therefore, comparable data are needed for at least two grades, if possible adjacent ones, for each school to implement this estimation strategy. Fortunately, the TIMSS data fulfill this requirement and thus the regression strategy can be implemented.

4. Results

The results of the estimation of educational production functions for the seven transition countries are discussed for each category of explanatory variables. In addition, a closer look is paid to possible interaction effects and to the effect of resources that are measured in the form of class size.

4.1. Results of the educational production functions

The production functions have all been estimated by CRLR (cf. Section 3.3.1) using model (1). First, the test scores are regressed only on the student background variables, then the resource and the institutional variables are added to the production function, respectively.

4.1.1. Student background

The variables of the student background in Table 2 feature the largest and most significant coefficients of the production function, which is in line with the results from other estimations in the literature.[10] The estimates in Table 2 do not yet control for resource and institutional measures, in order to obtain estimates of the total impact of family background on student performance, including any effect that might work through families' differential access to schools.

The results show large and statistically significant performance differences by grade, age, gender, immigration status, parental education, and the number of books at home in basically all countries and in both subjects. The coefficients from the regressions for the science test score are very close to the math estimates but on average slightly lower. This suggests that student characteristics and background have less effect on the science test score, so that it depends on a student's 'quality' to a lesser degree. The only coefficient that is consistently larger in absolute terms for science is the one on students' sex. The negative effect of being a female student is even larger in science than in math.[11]

Across the different countries, a pattern of the magnitude of the coefficients is apparent. In CSK and HUN, the coefficients always have the greatest values in absolute terms. In ROM and LVA instead, the values are the lowest whenever they are statistically significant, except for the community location. In general, the countries belonging to the first group have higher coefficients than the countries of the second group. This pattern is especially clear for the variables concerning the students' family background.

Table 3 shows the *F*-statistics from a test on equal student background coefficients for all combinations of countries. For the majority of combinations, the difference in student background coefficients is statistically significant at the 1% level, justifying our presentation of results by individual countries. Only for countries of the first group, four of the six possible combinations have no significantly different coefficients at the 5% significance level for science and one combination for math. Taking an *F*-statistic of 4 as a cutoff value, the separation of the countries into the two groups emerges quite clearly. All combinations of countries of the first group can then be grouped together and all of the second group as well, except for the combination LTU-ROM. There are also few combinations of countries between the groups who have an *F*-statistic below 4. However, these combinations differ between the subjects, while the within group combinations with a low *F*-statistic are identical for both subjects.

The lower coefficients in the second group of countries imply that background differences between students affect the test scores less in this group. This suggests that in countries, where the reform process commenced later and where, at least in the Baltic States, the Russian grip over the country was strongest, a major aim of schooling was the homogenous performance of students. The outcomes are relatively low mean scores, lower variations in the test scores in these three countries as well as lower returns to individual characteristics in schools. The extreme two cases are ROM for a system that seems hardly unchanged from communist times and CSK with great returns to individual background features.

[8]See Wößmann and West (2002) for details on the SFE+IV method.

[9]This variation is similar to the one identified by Hoxby (2000), who exploits random variations over time in student enrollments to identify exogenous variation in class sizes.

[10]The marginal effects confirm the great effect of the student background variables relative to other variables.

[11]For a detailed discussion of gender differences in TIMSS, see Mullis, Martin, Fierros, Goldberg, and Stemler (2000).

Table 2
Student background and educational performance

	CSK		HUN		SVN		SLV		LTU		LVA		ROM	
	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.
Upper grade	70.92* (5.92)	57.02* (5.54)	63.49* (3.09)	56.78* (3.44)	63.05* (4.90)	41.64* (4.18)	61.06* (4.19)	54.32* (4.08)	67.25* (4.44)	86.56* (4.56)	48.51* (4.24)	61.90* (4.25)	33.90* (3.93)	40.06* (5.11)
Age	−33.41* (2.83)	−18.30* (2.82)	−31.63* (2.11)	−23.51* (2.04)	−22.46* (3.31)	−13.20* (3.57)	−20.73* (2.74)	−18.77* (2.80)	−23.97* (2.56)	−16.63* (2.91)	−16.87* (2.51)	−11.96* (2.31)	−7.85* (2.48)	−6.73** (3.30)
Female	−13.34* (3.25)	−25.70* (2.72)	−7.00* (2.36)	−21.47* (2.28)	−7.10* (2.42)	−22.13* (2.64)	−5.99** (2.39)	−19.34* (2.45)	4.36*** (2.67)	−9.72* (2.64)	−6.92* (2.41)	−14.20* (2.42)	−2.55 (2.55)	−9.01* (2.92)
Immigrant	−22.07* (8.06)	−14.45*** (7.49)	−7.26 (8.36)	−18.46*** (10.00)	1.87 (4.18)	−12.86* (3.68)	−1.65 (8.19)	−11.64 (8.46)	−8.03 (9.64)	−14.02*** (9.33)	−14.29*** (7.39)	−22.36** (9.03)	16.09** (6.36)	8.46 (7.93)
Living with both parents	14.92* (4.83)	9.39* (3.55)	7.80*** (4.09)	9.06** (4.15)	2.36 (3.87)	−.04 (3.70)	1.65 (2.74)	−4.42*** (2.63)	1.25 (3.20)	4.98 (3.81)	−.71 (2.82)	−2.69 (3.09)	−3.52 (3.62)	−3.91 (4.51)
Parents' education														
Finished secondary	24.25* (3.51)	14.82* (3.43)	7.39*** (4.03)	6.54 (4.58)	23.99* (2.90)	16.40* (3.17)	15.43* (3.25)	12.32* (4.01)	11.02*** (6.10)	4.92 (5.77)	−.47 (4.31)	2.43 (3.83)	1.93 (5.40)	−1.02 (7.19)
Finished university	47.17* (4.68)	33.59* (4.67)	47.16* (6.14)	37.18* (5.76)	49.34* (4.22)	36.08* (4.71)	40.11* (4.51)	35.56* (4.68)	31.61* (6.59)	23.10* (6.55)	29.16* (5.32)	25.13* (4.52)	21.39* (7.03)	20.85** (9.02)
Books at home in ln	22.59* (1.89)	20.17* (1.85)	22.61* (1.49)	19.98* (1.51)	18.57* (1.58)	17.42* (1.58)	21.54* (1.33)	19.14* (1.40)	17.92* (1.16)	13.36* (1.44)	15.39* (1.74)	10.05* (1.63)	11.97* (1.72)	12.68* (2.05)
Community location														
Close to center	5.75 (7.93)	1.81 (5.14)	8.90** (4.45)	10.29* (3.75)	−2.78 (3.88)	−1.47 (3.32)	−.18 (5.70)	−.90 (5.36)	7.34*** (4.32)	2.71 (4.58)	4.69 (4.92)	1.00 (4.22)	19.98* (6.78)	16.38** (8.31)
Cons	832.19* (41.99)	671.24* (39.08)	796.56* (31.38)	724.60* (29.55)	706.08* (46.73)	633.38* (50.86)	671.68* (38.09)	673.21* (39.78)	647.59* (34.38)	556.97* (41.41)	603.36* (37.85)	544.20* (33.03)	506.50* (36.17)	494.08* (48.99)

Significance levels (based on robust standard errors): *1%, **5%, ***10%.
Separate least-squares regressions within each country and subject, weighted by students' sampling probabilities. Dependent variable: TIMSS math/science test score. Clustering-robust standard errors in parentheses.

Table 3
F-statistics on equal student background coefficients

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
CSK	—	1.44	1.32	1.67***	9.08*	6.82*	6.72*
HUN	2.54*	—	3.19*	1.90**	11.49*	8.26*	6.03*
SVN	3.19*	3.99*	—	.94	11.13*	5.15*	4.44*
SLV	3.63*	2.85*	.72	—	8.31*	4.43*	3.65*
LTU	5.75*	4.98*	2.26**	2.06**	—	3.29*	6.98*
LVA	7.49*	5.24*	3.81*	2.77*	2.89*	—	2.84*
ROM	14.77*	13.15*	8.47*	7.18*	6.62*	3.08*	—

Note: Statistics below the diagonal are for math, those above the diagonal are for science.
Prob. $F>0$: *1%, **5%, and ***10%.

4.1.2. Resources, teacher characteristics and institutional setting

The coefficients of the category of resource and teacher inputs are shown in Table 4 and indicate the relationships between the differences in school endowments and teaching staff and their students' test scores. The estimated relationships are useful for policy implications, as resources are allocated to schools by policy makers. The evidence on class size is examined in detail in Section 4.3 below. The specific needs of schools are reflected in the variable 'Great shortage of materials', as compared to some or no shortage. It is mostly negative though seldom significant. The characteristics of teachers give only small insights into a further explanation of students' test scores.

Most of the variables that describe the institutional setting at the school level through stating the degree of responsibility and autonomy that school heads and teachers have are statistically insignificantly related to student performance. Table 5 presents the results. When teachers have a strong influence on the curriculum, the direction of the effect seems to depend on whether they act individually or collectively. In the former case, the coefficient is mostly positive, though only once statistically significant. The coefficient for a collective influence is instead negative in the majority of cases but never statistically significant. Individual or class teachers have an informational advantage and do not act as an interest group, which is the case for collective teachers' influence or that of teacher unions.

The results for the influence of resources, teacher characteristics and the institutional setting feature far less statistically significant coefficients, which are of a lower magnitude than those for the first category of student background variables. Moreover, the effects of greater endowments of schools are somewhat ambiguous. Most unexpected signs of effects can be attributed to the unusual distribution within countries, or are discussed more extensively later on in the case of class size. However, there is no consistent picture that clearly indicates the merits for students from greater resources or better staff. A proper endowment with materials, teachers' experience and their educational level of a Master's degree still seem to be related to higher test scores of students in some cases. Possible beneficial side effects of the variables that are not grasped directly by the variable itself are examined later on in Section 4.4 with the help of interaction variables.

The limited number of schools, around 150 per country, allows only for little measurable variation within a country and leads to low degrees of freedom when all 23 explaining variables are included in the regression. This concerns the variables on school responsibility and teacher influence on the curriculum, which are measured at the school level. For the other variables, which are measured at the class level, the degrees of freedom should suffice but the effects are not clear-cut either. It seems that differences in the institutional setting are mainly relevant for understanding the cross-country variation in student performance (Wößmann, 2003), whereas the descriptive statistics in Table A3 show that there is little variation in the institutional setting within most Eastern European countries which could lead to a variation in test scores. Still, if we compare the optimal institutional setting with the least favorable one, taking into account only coefficients that are statistically significant at least at the 10% level, students score on average around 50 points higher in CSK for math and around 20 points for science. A difference of 50 points is half of the international standard deviation of test scores and around one-tenth of the mean score. Hence, institutions do matter in some countries, and their setting should not be neglected, especially since a modification might be achievable at lower cost than an increase in resources.

4.1.3. Explanatory power of the three categories of variables

The statistical significance and magnitude of the coefficients from the three categories of variables are

Table 4
Resources, teacher characteristics, and educational performance

	CSK		HUN		SVN		SLV		LTU		LVA		ROM	
	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.
Class size	2.75* (.88)	1.58** (.68)	.40 (.40)	.26 (.35)	.35 (.60)	−.47 (.56)	.12 (.50)	.84 (.72)	1.83* (.53)	−.25 (.45)	−.31 (.31)	−.43*** (.25)	1.49* (.52)	.68 (.71)
Great short. of materials	−9.50*** (5.02)	−7.08 (6.47)	6.54 (5.00)	6.65 (4.26)	−.06 (5.04)	−6.63 (4.41)	−7.29 (5.72)	2.09 (6.56)	−9.87*** (5.17)	−8.93*** (5.07)	5.04 (5.27)	9.07** (4.46)	−6.28 (9.02)	−10.51 (9.42)
Teacher characterist.														
Teacher is female	−5.02 (6.49)	−3.03 (6.52)	−4.72 (5.67)	2.72 (3.79)	−.03 (4.98)	1.32 (4.37)	4.06 (6.88)	6.85 (5.45)	5.12 (6.03)	6.75 (5.74)	−3.33 (8.17)	−1.68 (4.64)	8.65 (7.20)	3.35 (8.05)
Teacher exper. in ln	1.34 (4.11)	−4.98 (3.05)	5.23** (2.29)	1.00 (2.25)	2.65 (3.36)	.78 (2.26)	7.51*** (3.96)	1.46 (2.79)	2.03 (2.69)	.34 (2.31)	1.69 (2.55)	−1.20 (2.05)	8.91 (5.49)	9.78** (4.43)
Teacher education														
BA or equiv.	No observ.	No observ.	−9.30 (6.52)	−2.36 (5.14)	3.69 (7.10)	9.46** (4.61)	No observ.	No variation	4.82 (8.76)	−18.95*** (10.04)	5.39 (10.51)	6.80 (4.57)	7.33 (6.44)	11.24 (7.24)
MA/Ph.D.	17.90 (13.93)	18.41** (7.92)	20.09 (15.00)	−29.50** (14.27)	No observ.	No observ.	−12.01*** (6.21)	No variation	5.84 (9.80)	−7.59 (10.64)	−16.44 (20.16)	−5.97 (14.03)	51.49* (9.13)	32.92* (12.48)

Significance levels (based on robust standard errors): *1%, **5%, and ***10%.
Separate least-squares regressions within each country and subject, weighted by students' sampling probabilities. Dependent variable: TIMSS math/science test score. Controlling for all student background variables reported in Table 2. Clustering-robust standard errors in parentheses.

reflected in their contribution to the explanatory power of the models. Table 6 displays the R^2 for the regressions including all variables and the percentage decrease in R^2 when categories of variables are excluded from the regression. When the student background variables are excluded, the R^2 drops by over 75% in all countries except for ROM. Institutional and resource variables contribute only little to the share of the explained variance in test scores. The highest R^2 for the entire model is reached in LTU, where a quarter of the variation of the science test scores can be explained by the production function. ROM instead features the lowest values in both math and science, which is only slightly above 11%. The values differ greatly between countries, which might either suggest that the model used to estimate the production function suits certain countries better or that the quality of the data is lower in others.

4.1.4. *Sensitivity analysis of the imputation technique*

The estimations of the production functions have also been performed with the original values only. For each variable included in model (1), all imputed values (cf. Section 3.1) have been dropped. These 'robust' estimates are not affected by the method of imputation. This sensitivity analysis of the results examines whether changes in the data, in this case the introduction of changes through the imputation of values, alter the outcome of the regressions, in which case inferences from these data would seem fragile (Mukherjee, White, & Wuyts, 1998).

In general, there are no great distinctions between the two differently estimated coefficients for each variable. No statistically significant variable changes its sign. However, some statistically insignificant variables change signs and for others the statistical significance level changes.

Given that the estimates are so close to each other, the imputation technique should not have led to a bias of the data and the inferences are not fragile. Having the full dataset available for the estimations is of great advantage because the higher amount of observations allows for a better explanation of the variation. This is shown by the decreasing standard errors. Basing the estimation of the effect of school or class differences on even fewer observations might have led to even weaker inferences. Thus, the imputation of missing values was a worthwhile step in the estimation of the educational production functions.

4.2. *Interaction effects*

This section considers possible interaction effects between variables of the educational production function. The interaction terms have been included separately at the end of model (1) and indicate any further

Table 5
Institutional settings and educational performance

	CSK		HUN		SVN		SLV		LTU		LVA		ROM	
	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.	Math	Sci.
School responsibilities														
Autonomy	7.80 (4.97)	3.02 (3.40)	10.58*** (5.76)	6.26 (4.02)	−3.09 (6.38)	−3.85 (5.26)	3.12 (3.25)	2.13 (3.01)	.77 (4.27)	−1.72 (4.77)	−6.43 (6.58)	−1.65 (4.73)	−3.45 (3.53)	−1.00 (4.51)
Determining teacher's salary	8.12 (6.95)	5.92 (5.36)	−5.72 (6.14)	−7.24*** (4.11)	3.28 (4.25)	.65 (3.98)	−6.52 (6.89)	−3.66 (8.51)	−10.87** (5.05)	−.37 (6.52)	−1.73 (5.41)	−7.79 (4.92)	−2.56 (10.65)	−6.23 (10.65)
Strong infl. on curriculum														
Teachers individually	14.32*** (7.88)	3.62 (5.88)	7.36 (6.24)	4.86 (5.64)	−.46 (6.93)	4.75 (6.40)	−3.12 (9.26)	−2.11 (7.96)	6.23 (4.67)	6.44 (4.90)	−3.04 (4.79)	−.96 (4.13)	10.37 (11.24)	15.70 (10.47)
Teachers collectively	−7.92 (7.28)	−2.72 (5.08)	1.14 (5.52)	1.60 (5.08)	6.31 (8.91)	−1.10 (7.95)	2.86 (11.56)	.70 (9.94)	−1.59 (4.49)	−1.55 (5.20)	3.14 (5.20)	3.00 (4.33)	−6.53 (17.57)	−3.08 (19.86)
Class teacher has strong influence on														
Supplies or subject matter	−.90 (8.80)	−5.33 (4.39)	2.47 (4.55)	5.57 (3.48)	16.12* (6.03)	4.28 (4.89)	−.51 (6.62)	−.32 (5.26)	14.43 (10.24)	−10.67 (6.65)	−1.12 (8.08)	−3.09 (5.15)	−8.03 (8.96)	−12.25 (9.75)
Kind supplies/ textbooks	4.90 (5.17)	4.19 (4.70)	2.99 (5.35)	−6.67*** (3.73)	−4.74 (3.76)	−5.07 (3.69)	7.33 (6.27)	−2.07 (4.99)	−2.72 (7.57)	6.77 (7.16)	1.33 (5.56)	−.49 (4.52)	19.41** (8.28)	−11.69 (8.99)
Additional														
Homework	19.23*** (9.86)	−6.62 (10.15)	3.22 (3.57)	−1.47 (2.68)	1.68 (2.02)	−4.82 (6.56)	2.07 (3.13)	9.26 (10.40)	−2.66 (2.12)	5.43 (3.41)	3.03 (2.75)	7.29** (2.82)	1.93 (1.29)	10.07** (4.18)
Unint. parents limit teaching	−19.79* (6.62)	−19.11** (9.47)	−8.40 (6.65)	1.84 (5.71)	.58 (5.19)	−7.43** (3.71)	−1.38 (8.01)	−11.82 (7.45)	2.64 (6.63)	−7.69 (9.61)	−9.26*** (4.99)	4.40 (5.62)	−.63 (5.05)	5.73 (6.29)

Significance levels (based on robust standard errors): *1%, **5%, and ***10%.
Separate least-squares regressions within each country and subject, weighted by students' sampling probabilities. Dependent variable: TIMSS math/science test score. Controlling for all student background variables reported in Table 2 and for all resource variables reported in Table 4. Clustering-robust standard errors in parentheses.

Table 6
R^2 and percentage decrease in R^2 when categories of variables are excluded

Excl. category	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Math	.1996	.2208	.1770	.1476	.2379	.1282	.1253
Background (%)	79.71	85.24	88.70	91.67	76.33	91.42	50.84
Resources (%)	3.26	2.54	.34	2.03	3.61	2.42	14.76
Institutions (%)	11.02	1.77	2.66	1.42	2.31	3.35	7.58
Science	.1615	.1761	.1321	.1284	.2505	.1443	.1111
Background (%)	85.26	89.44	89.48	94.47	89.66	79.42	59.59
Resources (%)	3.34	1.02	2.35	1.71	3.07	3.26	15.30
Institutions (%)	4.27	1.76	1.89	1.56	1.80	2.70	14.58

effects a variable may have in connection with other variables.[12]

In both math and science, students seem to perform better under a teacher of the same sex. This is true for both boys and girls. The additional effect of teacher experience of teachers who hold a Master degree on students varies across countries. It is twice statistically significantly negative in ROM, due to the extremely high coefficient of the Master level. In HUN, the effect is statistically significantly positive for science, where the coefficient of a Master degree is statistically significantly negative, and negative for math, where the Master coefficient is positive. The interaction effects seem thus to offset some of the counterintuitive effects in the production function. The effect of class size does not consistently depend on the experience of teachers or on their educational level. When teachers hold a Master degree, class size exerts an additional significantly negative effect on student performance in three cases and a significantly positive effect in two cases.

When the students' parents have more than 200 books, the positive coefficient of class size is intensified in five countries. The positive interaction term could support the hypothesis of Lazear (2001) that the optimal class size for well-behaved students is larger than that for more disruptive students.

4.3. Analyses of the class size effect

The class size measure available in the TIMSS database is superior to most other studies, which often only have data on pupil–teacher ratios at the school level, in that each math and science teacher reports the size of her specific math and science class. Therefore, the TIMSS class size measure measures class size correctly even if students change classes between subjects. However, TIMSS does not report whether there are additional teacher aides in a classroom.[13] In order to give more scrutiny to the possible endogeneity bias in the least-squares estimation of resource effects, we analyze the class size effect in greater detail. We first look at class size effects in different segments of class sizes to see whether this is indicative of sorting of students into differently sized classes, and then we implement the model combining SFE and IV derived in Section 3.3.2 to eliminate any effects of between- and within-school sorting from the estimate of the class size effect.

4.3.1. Class size effects for class segments

In each country, the class size effect is estimated separately for three segments of class sizes, the lower, middle and upper segment. The class sizes that are included in the segments are chosen such that the difference between the number of students in the segments is minimized for each country. Thus, the segments can cover different class sizes across countries. In CSK (HUN) for example, the lower segment comprises classes up to a class size of 23 (21), the middle one classes between 24 (22) and 26 (25) and the upper classes with more than 26 (25) students. In order to avoid further bias, all explanatory variables of the production function are included in the estimation and control for other effects. As the number of classes whose size is estimated is greatly reduced in the regressions, standard errors of the class size coefficient increase for the individual categories.

It is enticing to compare the overall coefficients of class size to the coefficients from the segments. If the coefficients for class size of the segmented student population were mainly consistent and pointed in the same direction as the overall coefficient, this would give

[12]Detailed results are available from the authors upon request.

[13]This may not be a serious problem, though, because teacher aides are rare in the sampled countries and because teacher aides have been found to have negligible effects on student performance even in studies that find significant class size effects (Krueger, 1999).

Table 7
Class size coefficients for class segments

	CSK	HUN	SVN	SLV	LTU	LVA	ROM	EAST
All students								
Math class	2.18** (.92)	.50 (.38)	.44 (.59)	.23 (.51)	1.62* (.53)	−.36 (.31)	1.37* (.47)	.80* (.20)
Science class	1.22*** (.69)	.24 (.34)	−.50 (.54)	.82 (.67)	−.20 (.44)	−.43*** (.26)	.67 (.69)	.37*** (.20)
Lower segment								
Math class	.23 (1.38)	−.19 (.83)	−.30 (1.82)	.04 (1.12)	1.53 (1.21)	−1.36 (.83)	−.98 (1.18)	.24 (.40)
Science class	−.50 (.92)	−.17 (.83)	−3.04** (1.25)	.90 (1.02)	1.46 (1.23)	−1.56 (.96)	−1.46 (1.49)	−.46 (.41)
Middle segment								
Math class	16.52* (5.02)	5.35*** (2.61)	10.15* (3.68)	−6.53 (4.47)	−.45 (7.33)	−1.94 (2.19)	−1.10 (2.89)	3.78* (1.42)
Science class	−5.00 (5.92)	−3.62 (2.64)	7.75*** (4.28)	−5.61 (3.64)	4.02 (6.49)	−2.28 (4.03)	−1.40 (3.13)	2.65 (1.79)
Upper segment								
Math class	−.06 (2.89)	2.83 (2.03)	3.43 (2.07)	1.40 (3.17)	4.36* (1.51)	−2.06** (.78)	4.68* (1.44)	1.55* (.64)
Science class	4.78** (2.18)	2.88** (1.27)	−.86 (1.73)	.98 (2.50)	−1.30* (.41)	−1.11* (.31)	4.78* (1.72)	.57 (.39)
Mean math class	25.35	22.41	24.67	26.09	20.86	21.65	26.67	23.95
Mean science class	25.52	22.14	24.42	26.54	21.63	23.74	26.23	24.32

Significance levels (based on robust standard errors): *1%, **5%, and ***10%.
Separate least-squares regressions within each country and subject, weighted by students' sampling probabilities. Dependent variable: TIMSS math/science test score. Controlling for all student background variables reported in Table 2, for all other resource variables reported in Table 4, and for all institutional variables reported in Table 5. Clustering-robust standard errors in parentheses.

evidence for no selection bias. Then the effect would be identical for different ranges of class sizes. However, if there were greatly varying coefficients and especially statistically significantly negative ones among them, this would point to an outside involvement like the non-random allocation of students. If this were the case and the low-performing students were put into smaller classes, then the effect for a limited range within which no selection takes place might still be negative. If the segment coefficients were negative but the overall coefficient positive, a selection between segments would seem likely. A positive coefficient would indicate that selection takes place within the considered range instead.

The results shown in Table 7 cannot reveal any clear evidence on whether between- or within-school selection takes place, though. The results from the segmented class size estimations are not very consistent. For each country, both positive and negative coefficients are reported, with the exception of LVA with only negative coefficients. There are 11 statistically significantly positive coefficients versus four statistically significantly negative coefficients. For the merged dataset EAST that contains all seven transition countries and country dummies, there are statistically significantly positive coefficients for the overall estimates and statistically significantly positive ones for math in the middle and upper segment. The variation of the coefficients is greatest in the middle segment and lowest in the lower segment. For science more negative coefficients are estimated as compared to math, where only the LVA coefficient in the upper segment is statistically significantly negative. The inconsistency of the coefficients across countries and across segments can be interpreted as an indication of a possible bias of the class size coefficient in the production function but is clearly no proof of it.

4.3.2. Eliminating effects of between- and within-school selection

In order to control for selection that takes place between schools when measuring the class size effect, a dummy for each school in the country is added to the regression, leading to the SFE model (3) derived in Section 3.3.2. The few schools that tested only one class are excluded from the estimation of this model. The class size coefficients of the SFE model shown in Table 8 are smaller than for the survey regressions in 11 out of 14 cases. The standard errors increase only slightly. The results indicate that when excluding the effect of between-school selection of students, a positive effect of class size on test scores appears less likely. The change is most drastic in CSK, where the statistically significantly positive class size coefficient turns negative for math and is around zero for science. In LVA for math, a statistically significantly negative coefficient results from the control for SFE. Only two coefficients in LTU and ROM for math remain significantly positive. In these two countries, the different model has hardly any influence on the coefficients.

In a second step, the selection within schools is additionally controlled for by using the average class size at the grade level as an instrument for actual class

Table 8
The coefficient on class size in different models

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Math							
Least squares	2.18** (.92)	.50 (.38)	.44 (.59)	.23 (.51)	1.62* (.53)	−.36 (.31)	1.37* (.47)
School fixed effects	−.70 (1.80)	.18 (.51)	−.20 (.82)	.01 (.65)	1.64*** (.85)	−1.06* (.39)	1.51* (.56)
SFE + IV	.23 (2.89)		.49 (1.60)				−.89 (4.96)
Science							
Least squares	1.22*** (.69)	.24 (.34)	−.50 (.54)	.82 (.67)	−.20 (.44)	−.43*** (.26)	.67 (.69)
School fixed effects	.14 (1.15)	.11 (.56)	−.43 (.74)	1.03 (.82)	−.01 (.57)	.25 (.48)	.52 (.62)
SFE + IV	−.88 (1.78)		.04 (1.27)				−.61 (2.44)
Students LS/SFE	6672/6659	5978/5962	5606/5576	7101	5056	4976/4917	7471/7462
Schools LS/SFE	150/149	150/149	122/121	145	145	143/141	163/162
Average difference (std. dev.) in class size between the two grades of a school							
Math class	−.52 (4.10)	.68 (5.28)	−1.55 (3.64)	−.75 (4.89)	−.06 (4.37)	−.08 (7.18)	−2.87 (5.77)
Science class	−.40 (4.08)	−.22 (4.89)	−1.34 (3.68)	−.75 (4.05)	.49 (5.72)	1.76 (9.27)	−2.21 (5.69)
Instrument	.34 (4.38)		−1.64 (3.64)				−.52 (5.27)

Significance levels (based on robust standard errors): *1%, **5%, and ***10%.
Separate regressions within each country and subject, weighted by students' sampling probabilities. Dependent variable: TIMSS math/science test score. Controlling for all explanatory variables in least-squares regression and for student background variables, teacher characteristics and education in SFE regressions. Clustering-robust standard errors in parentheses.

size (cf. Section 3.3.2). The model and corresponding technique (SFE + IV) can only be applied to three countries, for which sufficient data are available on the instrument. To check whether there is enough variation in our instrument—the average class size in a grade—between the two grades within individual schools, the bottom of Table 8 reports the country mean and its standard deviation in average class size between the 7th and 8th grade of a school. In the three countries for which this measure is available, the between-grade variation within schools, as measured by this standard deviation, is actually comparable in size to the original standard deviation of class size in our samples (cf. Table A3). Therefore, the variation we use in this identification seems large enough for the estimation. Furthermore, in order for this identification strategy to be identified, we have to assume that the class size effect is the same across grades. While this is not necessarily generally true (cf. Todd & Wolpin, 2003) and class size effects may well differ between, say, first grade and tenth grade, it may be reasonable to assume that the effect does not change significantly between the 7th and the 8th grade. We test this in the SFE model and indeed find that there is no statistically significant difference between the two grades in the class size estimate in any country or subject, except for weakly statistically significant differences for math in SLV and in LTU.

For math, the consistent SFE + IV model leads to slightly positive coefficients in CSK and SVN, which are statistically highly insignificant, though. In ROM, the former significantly positive coefficient turns negative. This is as well the case for science, where two of the three coefficients for class size are negative, however still insignificantly. The estimates correspond to the intuitive reasoning that a smaller classroom is a better learning environment for students and should benefit their scores. The relatively large standard errors of the coefficients do not allow for pinpointing the exact effect.[14] It can only be said that it is approximately close to zero or even slightly negative. The results do not support a positive resource–performance link, but they do show that the counter-intuitive least-squares coefficients are likely to be biased.

In CSK, selection between schools seems to be the major cause of bias in the class size coefficient. Low-performing students seem to be allocated to schools that have lower class sizes than those schools of the better students. In ROM instead, between-school selection seems to introduce no bias, but within-school selection does seem to. This might as well be the case in LTU and could explain the statistically significantly positive coefficient of class size after controlling for SFE. The difference in the origin of the selection bias seems to be related to the structure of the school system. In the more diverse systems in CSK and HUN, students of the considered age of about 13 can choose between two types of schools, and the more able students are

[14]Hausman tests reject the hypothesis that the coefficients are different between the SFE and the SFE + IV model in only one of the three countries. For the other two countries, the statistical power of the SFE + IV model seems too limited to detect statistically significant differences.

Table 9
Educational production in Eastern and Western Europe

	East	First	Second	West
Student and family characteristics				
Upper grade	58.36* (1.71)	66.05* (2.31)	49.78* (2.49)	55.72* (2.14)
Age	−22.52* (1.03)	−28.25* (1.39)	−16.22* (1.49)	−20.75* (1.31)
Female	−5.22* (1.00)	−8.07* (1.34)	−1.83 (1.49)	−9.89* (1.47)
Immigrant	5.71*** (3.16)	−2.46 (3.00)	10.15** (4.66)	−25.63* (3.20)
Living with both parents	.75 (1.45)	6.29* (1.91)	−1.08 (1.91)	5.44* (1.57)
Parents' education				
Finished secondary	11.07* (1.76)	17.79* (1.79)	−.25 (3.30)	9.10* (2.23)
Finished university	37.80* (2.10)	45.04* (2.44)	24.15* (3.73)	28.91* (2.90)
Books at home	16.68* (.67)	20.72* (.78)	13.62* (.95)	14.93* (.68)
Close to the center	3.59 (2.20)	2.64 (2.86)	6.81** (3.24)	3.84 (3.33)
Resources and teacher characteristics				
Math class size	.80* (.20)	.68** (.28)	.81* (.27)	1.62* (.37)
Great shortage of materials	−2.19 (2.55)	.58 (3.17)	−3.11 (3.60)	3.85 (3.45)
Math teacher char. and education				
Teacher is female	.69 (2.73)	−2.07 (3.24)	5.36 (4.50)	−4.83*** (2.69)
Teacher's exper. in ln	5.04* (1.47)	4.21** (1.95)	5.43** (2.26)	2.76 (2.04)
BA or equivalent	1.59 (3.97)	−6.86 (5.39)	5.12 (4.94)	−3.47 (4.81)
MA/Ph.D.	4.37 (6.00)	14.46 (9.11)	4.97 (7.68)	−4.91 (5.86)
Institutional setting				
School responsibilities				
Autonomy (budget, suppl., t.)	1.71 (1.85)	5.25** (2.26)	−3.06 (2.75)	3.17 (3.65)
Determining teacher salary	1.35 (2.55)	3.79 (3.21)	−5.00 (4.03)	−5.21 (4.53)
Strong influence on curriculum				
Teachers individually	4.48*** (2.51)	5.17 (3.82)	3.75 (3.27)	3.77 (4.20)
Teachers collectively	−1.61 (3.01)	−2.26 (4.41)	−.07 (3.65)	−14.05* (3.57)
Math class teacher has strong influence				
Money for suppl. or subject matter	2.55 (2.83)	3.72 (3.37)	−.59 (5.28)	−.33 (2.82)
Kind of supplies or textbooks	3.51 (2.27)	2.60 (2.70)	6.93*** (3.96)	.98 (2.50)
Homework	2.25** (.99)	3.71*** (1.95)	1.05 (1.09)	8.18* (1.85)
Uninterested parents limit teaching	−4.09*** (2.43)	−5.02 (3.55)	−2.72 (3.22)	−21.00* (5.51)
Cons	598.19* (18.26)	706.45* (24.20)	546.48 (26.57)	664.74* (23.70)
Students	42,815	25,357	17,458	21,933
Schools	1017	567	450	553
Mean math score	500.97 (93.61)	527.09 (91.93)	466.14 (84.00)	505.10 (86.34)

Significance levels (based on robust standard errors): *1%, **5%, and ***10%.
Separate least-squares regressions within each country group, weighted by students' sampling probabilities. Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

allocated into the gymnasium. In ROM instead, only general schools exist and students cannot be allocated to other school types. Therefore, if low performing students are to receive more resources, selection is more likely to take place within schools.

5. Comparison to Western European countries

In order to see how far the different Eastern European school systems have converged towards the Western European ones, we compare the educational production functions of Eastern and Western European countries. The merged functions of all Eastern European countries (EAST), of the first group of countries (FIRST) and of the second group of countries (SECOND) are compared to a sample of Western European countries (WEST) that includes Austria, Denmark, France and Germany. They all are central European countries and long time members of the EU and should hence be well suited for a comparison. The respective country dummies are added to the regressions.

Table 10
F-statistics on equal coefficients between groups

	First/second	First/west	Second/west
Math			
All variables	6.67*	5.40*	5.21*
Background	13.12*	8.61*	8.02*
Science			
All variables	7.22*	3.51*	5.63*
Background	14.18*	6.62*	10.81*

Prob. $F>0$: *1%, 5%, and 10%.

Table 11
R^2 and percentage decrease in R^2 when categories of variables are excluded

Excl. category		First	Second	West
All categories	Math	.1854	.1541	.2062
	Science	.1526	.1549	.2506
Background	Math (%)	85.81	73.46	67.26
	Science (%)	87.68	77.66	54.23
Resources	Math (%)	1.40	3.96	2.96
	Science (%)	.52	1.68	.92
Institutions	Math (%)	1.94	1.75	5.92
	Science (%)	.85	3.62	1.28

If the aforementioned assumption is true that the schooling systems of the first group of countries have already moved significantly towards those of democratic market economies, then the production function of the Western European sample should resemble more the function of the first group than that of the second group. Table 9 presents the coefficients of estimating model (1) with the math scores as the dependent variable. The mean math scores in the bottom right indicate that the average performance of the sample of Western European countries is only slightly above the average for Eastern Europe and right between the first and the second group of countries. The coefficients for WEST are mostly closer to the first group and are often opposed to the estimates of the second group. In the following, we compare the estimates of the Western European countries mainly to those of the first group and refer to the mean values of the explanatory variables.

In the category of student characteristics and background, the coefficients of WEST and FIRST are close to each other, except for 'immigrant', where the effect in WEST is clearly negative. Moreover, the Western countries have higher shares of immigrated students, which may complicate integration. The average education of students' parents is lower in WEST than in both Eastern groups. The number of books at home in WEST is instead slightly higher than in the second group but well below that of the first group. It is obvious that returns to the individual characteristics and the effects of student background are even higher in FIRST than in WEST and are relatively low in the second Eastern group. Schools in the latter group still seem to diminish the impact of student characteristics and produce a homogenous output of students. The lower standard deviation of the math test score of the second group reflects this focus of educational policy. In several aspects like the returns to individual student characteristics and the deviation of test scores, the first group seems to have surpassed the Western countries already.

This is underlined by *F*-tests on equal coefficients among the groups. Table 10 shows the results for the category of student background and for all variables. All *F*-statistics are statistically significant at the 1% significance level. The coefficients of both Eastern groups are closer to those of the Western sample than to each other, implying that the Western coefficients lie between those of FIRST and SECOND. For math, the *F*-statistics imply an equal difference of WEST to the Eastern groups while for science, the coefficients of WEST are much closer to FIRST than to SECOND.

Further support is presented in Table 11, which shows the percentage decrease in R^2 for the three groups when categories of variables are excluded. The background variables are most important for explaining the test score variation in all groups, while the resource and institutional variables have little impact on the explained variation. Student background effects contribute more to the R^2 in FIRST than in SECOND and WEST. Given the total R^2, background effects account for the biggest variation in FIRST at .159 ($= .185 \times 85.8\%$), followed by WEST at .139 and SECOND at .113. This supports our story that the differentiation of students according to their background in FIRST has surpassed even that in WEST. The equalization of educational outcomes is characteristic for the second group instead and is illustrated by the low effects of student characteristics on their performance and the lower deviation of test scores.

6. Conclusion

The analysis of the schooling systems of seven Eastern European countries by means of estimating their educational production functions reveals several distinctive features. First, the countries can be divided into two groups, which share similar characteristics in their economic development, the properties of their schooling systems and the effects that the various factors have on student test scores. The first group of countries, which includes the Czech Republic, Hungary, Slovakia and

Slovenia, commenced the political and economic reforms earlier than the second group and features a higher level of both political and economic development. In the second group, the Baltic States instead remained under tight Russian control until 1991 and suffered great economic damage from this strong link. Romania's political struggles delayed any reforms and turned it into the political and economic laggard of this sample of Eastern European countries. Concerning their schooling systems, the average student of the first group is endowed with more favorable characteristics and a higher level of directly measurable resources at schools, especially in the two countries of former Czechoslovakia. Further, the first group features the two most decentralized schooling systems with Hungary and Slovenia.

When regarding the relationship between the individual impact factors and student performance, as estimated by the educational production function, distinct patterns emerge for both groups. The effects of the very significant variables on student background, especially of family circumstances, are consistently higher in the first group, which introduced reforms to the education systems earlier. In these countries, the schooling systems seem already to be a mode of differentiation for the labor market. A comparison with the sample of Western European countries underlines the proximity of the first group to the Western schooling systems. It even surpasses the Western sample in several specific features, such as the high returns to individual characteristics of students and a higher variation in test scores. In the second group, returns to individual characteristics and especially to family background are instead lower and student performance varies less. This gives students more equal opportunities irrespective of their background but seems to be associated with a lower average performance. Thus, the second group still features the patterns of a schooling system whose primary role is to produce a homogenous output of students.

The coefficients for resources and the institutional setting are less significant and of a much lower magnitude in all countries and not consistent within groups. The merit of increased resources is illustrated for the experience and educational level of teachers in most countries. The effect of class size remains ambiguous, but positive coefficients were shown to be biased by student selection. Thus, it seems more likely that there is no effect or even a slightly negative one, especially for lower levels of class sizes. As the reduction of class size is very costly, it is doubtful whether a minor negative effect of larger classes justifies the enormous effort of decreasing class sizes. Although the institutional setting has comparatively little impact compared to student background, differences in the autonomy of schools and teachers are shown to be positively related to student performance in some countries and would be relatively effortless to modify.

The findings of this paper carry implications for the potential future development of the transition countries' economies. The tested students have by now reached an age of about 22 and are about to enter the labor market. The measured schooling quality is thus to take effect on the countries' economies in the coming years. Besides the relatively advantageous economic and political situation, the first group of transition countries also presides over well functioning schooling systems. The high development of institutions is visible in the favorable incentive system in schools, where decentralization has widely progressed and positively affects student scores. The quick transformation of the schooling system in the early years of transition despite the financial and political hurdles is a good foundation for economic growth in the coming years.

The second group of transition countries instead has not fully reformed its schooling systems yet. The countries still feature many traits from communist times. Institutional reforms may need to be continued in the coming years.

In the analysis of educational production in Eastern Europe, there is great scope for further research. It would be worthwhile that datasets became available which contain value-added information on the changes in inputs and performance of individual students, in order to check the robustness of the results found in this paper. It would also be desirable to find further methods for bias control that leave more variation in test scores and are applicable to other inputs as well. Further, it would be interesting to study the development of other outcomes of education like social values conveyed to students, which are essential to form a steady democracy. Another path for further studies would be to compare the findings from TIMSS to other studies that were performed later, like TIMSS-Repeat in 1999 or PISA in 2000. This could reveal whether the trends of the schooling systems towards decentralization and choice have continued and whether the distinction between the two groups of countries still holds.

Acknowledgements

Ammermüller gratefully acknowledges financial support from the European Commission IHP project "Education and Wage Inequality in Europe" contract number HPSE-CT-2002-00108. We thank two anonymous referees for their constructive comments.

Appendix A

Refer Tables A1–A3 for further details.

Table A1
Participation of students, classes and schools in TIMSS

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Students	6672	5978	5606	7101	5056	4976	7471
Classes	299	299	243	290	292	284	325
Schools	150	150	122	145	145	143	163
Sampled students (in %)[a]	.6	.4	3.0	.9	1.0	1.7	.3

[a]Number of sampled students over the number of children of compulsory school age in percent.

Table A2
Definition of variables and range of values

Variable name	Definition	Type	Min.	Max.
Math score	Intern. math test score	Numeric	141.6	887.44
Science score	Intern. science test score	Numeric	85.02	872.46
Student and family characteristics				
Upper grade	Grade level of students	Dummy	0 for 7th grade	1 for 8th grade
Age	Age of students in years	Numeric	10.1	20.4
Female	Sex of students	Dummy	0 for male	1 for female
Immigrant	Origin of students	Dummy	0 for other students	1 for immigrated student or parent
Living with both par.	Student's family situation	Dummy	0 for one/no parent	1 for both parents
Parents' education				
No or some secondary		Dummy	Ref.	Ref.
Finished secondary	Highest educational level reached by a parent	Dummy	0	1 for finished section and some after
Finished university		Dummy	0	1 for finished university
Books at home	Number of books at student's home in ln	Categorical	1.61	5.52
in ln				
Community location				
Non-urban area	Location of the student's community	Dummy	Ref.	Ref.
Close to center of town		Dummy	0 for ref.	1 for close to center
Resources and teacher characteristics				
Class size	Size of student's class	Numeric	3	58
No/some shortage	Degree of school's shortage of materials	Dummy	Ref.	Ref.
Great shortage of mat.		Dummy	0 for ref.	1 for great shortage
Teacher characteristics				
Teacher is female	Class teacher's sex	Dummy	0 for male	1 for female
Teacher's exp. in ln	Teacher's exp. in years in ln	Numeric	0	3.91
Teacher education				
Secondary and less	Highest educational level of class teacher	Dummy	Ref.	Ref.
BA or equivalent		Dummy	0	1 for bachelor/equ.
MA/Ph.D.		Dummy	0	1 for master/equ.
Institutional setting				
School responsibilities				
Autonomy (budget, supplies, teachers)	Degree of school's autonomy	Categorical	0 for autonomy in no field	3 for autonomy in all three fields
Determining teacher salary	School's responsibility over setting salaries	Dummy	0 for no resp.	1 for resp. over salary
Strong influence on curriculum				
Teachers individually	Influence of individual or subject teachers on curr.	Dummy	0 for no infl.	1 for infl. of at least one group

Table A2 (*continued*)

Variable name	Definition	Type	Min.	Max.
Teachers collectively	Infl. of teachers collect. or teacher unions on curr.	Dummy	0 for no infl.	1 for infl. of at least one group
Class teacher has strong influence on				
Money for supplies or subject matter	Strong infl. of class teacher	Dummy	0 for no strong infl.	1 for strong infl. in at least one field
Kind of supplies or textbooks	Strong infl. of class teacher	Dummy	0 for no strong infl.	1 for strong infl. in at least one field
Additional				
Homework	Homework for subject in hours per week	Numeric	1	9.6
Uninterested parents limit teaching	Teacher notices uncooperative parents negatively	Dummy	0 for no negative infl. on students	1 for negative infl. on students

Min. and max. values include all seven Eastern European countries in both math and science.

Table A3
Means and standard deviations for selected variables

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Math score	543.57 (93.70)	519.05 (93.52)	518.68 (87.43)	527.08 (90.73)	453.72 (81.60)	476.68 (80.82)	468.01 (87.81)
Science score	553.44 (86.90)	535.30 (92.48)	544.42 (88.11)	526.69 (90.16)	441.27 (88.26)	458.59 (83.35)	468.85 (102.33)
Student and family characteristics							
Upper grade	.50 (.50)	.49 (.50)	.48 (.50)	.49 (.50)	.52 (.50)	.47 (.50)	.50 (.50)
Age	13.89 (.65)	13.82 (.71)	14.27 (.63)	13.77 (.62)	13.84 (.66)	13.77 (.72)	14.12 (.71)
Female	.50 (.50)	.50 (.50)	.51 (.50)	.51 (.50)	.52 (.50)	.52 (.50)	.51 (.50)
Immigrant	.02 (.14)	.02 (.16)	.11 (.32)	.02 (.14)	.02 (.15)	.02 (.14)	.13 (.34)
Living with both parents	.89 (.32)	.92 (.27)	.90 (.30)	.83 (.37)	.85 (.36)	.80 (.40)	.74 (.44)
Parents' education							
Finished secondary	.55 (.50)	.72 (.45)	.64 (.48)	.58 (.49)	.56 (.50)	.61 (.49)	.57 (.50)
Finished university	.21 (.41)	.19 (.39)	.19 (.39)	.22 (.41)	.37 (.48)	.29 (.45)	.09 (.29)
Books at home							
Before log transf.	148.49 (80.94)	155.43 (90.35)	111.23 (83.10)	112.03 (79.06)	112.12 (86.61)	184.79 (81.58)	89.85 (95.00)
Community location							
Close to the center of a town	.40 (.49)	.45 (.50)	.38 (.48)	.29 (.45)	.61 (.49)	.30 (.46)	.51 (.50)
Resources and teacher characteristics							
Math class size	25.35 (3.36)	22.41 (5.19)	24.66 (3.70)	26.09 (4.32)	20.86 (3.93)	21.65 (6.75)	26.67 (6.46)
Science class size	25.52 (3.41)	22.14 (4.98)	24.42 (3.47)	26.54 (3.48)	21.63 (5.39)	23.75 (7.80)	26.23 (6.49)
Great shortage of materials	.03 (.18)	.17 (.38)	.20 (.40)	.02 (.14)	.22 (.42)	.63 (.48)	.17 (.37)
Math teacher characteristics							
Teacher is female	.84 (.36)	.85 (.36)	.88 (.32)	.79 (.41)	.85 (.35)	.94 (.25)	.67 (.47)
Teacher's exp. before log transformation	21.36 (11.43)	18.00 (8.92)	15.64 (7.03)	20.47 (9.33)	19.67 (10.36)	17.42 (10.31)	20.78 (9.79)
Math teacher education							
BA or equivalent	0 (0)	.09 (.29)	.05 (.22)	0 (0)	.81 (.39)	.93 (.25)	.45 (.50)
MA/Ph.D.	.99 (.11)	.02 (.15)	0 (0)	.99 (.08)	.16 (.36)	.02 (.14)	.00 (.07)
Science teacher characteristics							
Teacher is female	.77 (.42)	.78 (.42)	.82 (.38)	.66 (.47)	.85 (.36)	.83 (.38)	.77 (.42)

Table A3 (*continued*)

	CSK	HUN	SVN	SLV	LTU	LVA	ROM
Teacher's exp. bef. trans	21.12 (11.75)	18.56 (9.70)	16.57 (8.03)	18.88 (10.64)	19.54 (11.06)	18.00 (11.34)	20.01 (10.22)
Science teacher education							
BA or equivalent	0 (0)	.21 (.41)	.14 (.34)	0 (0)	.72 (.45)	.88 (.32)	.54 (.50)
MA/Ph.D.	.96 (.19)	.01 (.08)	0 (0)	1 (0)	.22 (.42)	.03 (.16)	.01 (.08)
Institutional setting							
School responsibilities							
Autonomy (budget, supplies, teachers)	2.59 (.66)	2.92 (.30)	2.98 (.13)	2.40 (.80)	2.79 (.44)	2.85 (.39)	1.51 (.86)
Determining teacher salary	.59 (.49)	.88 (.33)	.29 (.45)	.85 (.36)	.15 (.36)	.75 (.44)	.07 (.25)
Strong influence on curriculum							
Teachers individually	.83 (.37)	.79 (.41)	.24 (.42)	.16 (.37)	.50 (.50)	.44 (.50)	.08 (.27)
Teachers collectively	.66 (.47)	.54 (.50)	.15 (.36)	.10 (.30)	.24 (.43)	.56 (.50)	.04 (.20)
Math class teacher has strong influence on							
Money for supplies or subject matter	.13 (.33)	.77 (.42)	.12 (.33)	.19 (.39)	.04 (.20)	.12 (.33)	.17 (.37)
Kind of supplies or textbooks	.42 (.49)	.22 (.41)	.34 (.47)	.16 (.37)	.07 (.26)	.23 (.42)	.18 (.38)
Science teacher has strong influence on							
Money for supplies or subject matter	.16 (.37)	.71 (.46)	.17 (.38)	.27 (.44)	.14 (.35)	.26 (.44)	.21 (.40)
Kind of supplies or textbooks	.50 (.50)	.29 (.45)	.50 (.50)	.22 (.41)	.13 (.34)	.26 (.44)	.17 (.37)
Additional math							
Homework	.58 (.43)	1.36 (.51)	1.89 (.93)	1.18 (.71)	2.06 (.90)	1.87 (.74)	4.52 (1.83)
Uninterested parents limit teaching	.07 (.26)	.08 (.27)	.18 (.38)	.11 (.31)	.09 (.28)	.17 (.38)	.50 (.50)
Additional science							
Homework	.16 (.18)	.77 (.65)	.28 (.25)	.27 (.20)	.63 (.58)	.59 (.53)	.76 (.95)
Uninterested parents limit teaching	.06 (.23)	.10 (.30)	.17 (.38)	.08 (.28)	.06 (.24)	.09 (.28)	.37 (.48)

Standard deviations are reported in parentheses. Values are weighted by the sampling probability of the students.

References

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, *14*(3), 229–241.

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

Berryman, S. (2000). *Hidden challenges to the education systems in transition economies*. Washington, DC: The World Bank.

Boeri, T., & Terrell, K. (2002). Institutional determinants of labor reallocation in transition. *Journal of Economic Perspectives*, *16*(1), 51–76.

Card, D., & Krueger, A. B. (1996). Labor market effects of school quality: theory and evidence. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43–73). Washington, DC: The Brookings Institution.

Deaton, A. (1997). *The analysis of household surveys*. Baltimore and London: The John Hopkins University Press.

European Bank for Reconstruction and Development (EBRD) (2000). *Transition report* 2000—*employment, skills and transition,* London.

Filer, R. K., & Münich, D. (2000). Responses of private and public schools to voucher funding: the Czech and Hungarian experience. CERGE-IE Working Paper No. 160.

Gonzalez, E., & Smith, T. (Eds.) (1997). *User guide for the TIMSS international database—primary and middle school years.* Chestnut Hill: International Association for the Evaluation of Educational Achievement, TIMSS International Study Center, Boston College.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, *113*(485), F64–F98.

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

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

Heyneman, S. (1997). Educational choice in Eastern Europe and the former Soviet Union: a review essay. *Education Economics*, *5*(3), 333–339.

Hoxby, C. M. (2000). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics*, *115*(4), 1239–1285.

Kane, T. J., & Staiger, D. O. (2001). Improving school accountability measures. NBER Working Paper 8156, Cambridge, MA, National Bureau of Economic Research.

Klazar, S., Sedmihradský, M., & Vančurová, A. (2001). Returns of education in the Czech Republic. *International Tax and Finance*, *8*(4), 609–620.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*(2), 497–532.

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, *113*(485), F34–F63.

Lazear, E. (2001). Educational production. *Quarterly Journal of Economics*, *116*(3), 777–803.

Mukherjee, C., White, H., & Wuyts, M. (1998). *Econometrics and data analysis for developing countries*. London and New York: Routledge.

Mullis, I., Martin, M., Fierros, E., Goldberg, A., & Stemler, S. (2000). *Gender differences in achievement*. Chestnut Hill: International Association for the Evaluation of Educational Achievement, TIMSS International Study Center, Boston College.

Newell, A., & Reilly, B. (1999). Rates of return to educational qualifications in the transitional economies. *Education Economics*, *7*(1), 67–84.

Schultz, T. W. (1975). The value of the ability to deal with disequilibria. *Journal of Economic Literature*, *13*(3), 827–846.

Summers, A., & Wolfe, B. (1975). Equality of educational opportunity quantified: a production function approach. Philadelphia Fed Research Papers, Department of Research, Federal Reserve Bank of Philadelphia.

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Wooldridge, J. M. (2001). Asymptotic properties of weighted *M*-estimators for standard stratified samples. *Econometric Theory*, *17*(2), 451–470.

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

Wößmann, L., & West, M. (2002). Class-size effects in school systems around the world: evidence from between-grade variation in TIMSS. Kiel Working Papers No. 1099.