How Does East Asia Achieve Its High Educational Performance?*

Ludger Wößmann

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

January 2003

An extended version, referred to in the paper as the "working-paper version of this study," will be available in the near future at http://www.uni-kiel.de/ifw/pub/kap/kap.htm as "Educational Production in East Asia: The Impact of Family Background and Schooling Policies on Student Performance."

Word count:
7,902 words in main text and references.
Approximately 8,900 words including tables.

* Financial support by the International Centre for the Study of East Asian Development (ICSEAD), Kitakyushu, Japan, for the research project "Education Production Functions for East Asian Countries" is gratefully acknowledged. I would like to thank Erich Gundlach and participants at the Eighth International Convention of the East Asian Economic Association in Kuala Lumpur for helpful discussions and comments, Andreas Ammermüller for research assistance in the construction of the database, and the National Bureau of Economic Research (NBER) as well as the Program on Education Policy and Governance (PEPG), Harvard University, for their hospitality when working on this research.

How Does East Asia Achieve Its High Educational Performance?

Abstract

East Asian students regularly take top positions in international league tables of educational performance. Using internationally comparable student-level data, I estimate how family background and schooling policies affect student performance in five high-performing East Asian economies. Family background is a strong predictor of student performance in South Korea and Singapore, while Hong Kong and Thailand achieve more equalized outcomes. There is no evidence that smaller classes improve student performance in East Asia. By contrast, school autonomy over salaries and regular homework assignments are related to higher student performance in several of the considered countries.

JEL Classification: O15, I20, H52

Keywords: education production function, East Asia, family background, class size, school autonomy

Most of the high-performing East Asian economies have achieved universal enrollment of children in primary and secondary education. However, many people in these countries fear that their schooling systems do not provide the skills necessary to excel in a modern economy, such as analytical skills, creativity, and independence of mind (cf., e.g., Dosanjh and Richardson, 2001; Ward and Richardson, 2002). It has been commented that "it is ironic that this debate … is taking place at a time when many in longer-established developed economies are urging a return to traditional educational systems" (Richardson 1996, p. 22) emphasizing basic skills and general rather than highly specialized education. Certainly, a strong foundation in basic skills is a prerequisite for success in more ambitious tasks. And the East Asian countries actually seem to do very well with regard to general education: Their students repeatedly take top places in international comparative studies of cognitive achievement. For example, the first four places of 39 participating countries in the middle-school math test of the 1995 Third International Mathematics and Science Study (TIMSS) are taken up by Singapore, Korea, Japan, and Hong Kong. This extraordinary performance of East Asian countries had already been evident in previous cross-country studies, and it has been repeated since.[1] These achievement studies do not only test the basic knowledge of students, e.g. by multiple-choice questions, but also require students to accomplish a transfer and application of their knowledge to less familiar real-world tasks when solving more advanced open-ended questions. The lead of East Asian students over students from other continents is generally especially large in the latter, more difficult questions (cf. Beaton et al. 1996, pp. 57–98). Psychologists studying Asian and American metropolitan areas also conclude that "contrary to popular stereotypes the high levels of achievement in Asian schools are not the result of rote learning and repeated drilling … instead the students are led to construct their own ways of representing … knowledge." (Stevenson 1992, p. 32)

The crucial question thus is how the high-performing East Asian economies have achieved their high educational performance, and how they can sustain the quality of their knowledge foundation and ensure a high-quality education for all children for their

[1] For example, middle-school Japanese children performed second in the first internationally comparative math study in 1964, and the two East Asian countries participating in the 2000 OECD study Programme for International Student Assessment (PISA), Japan and Korea, took the first two places in math and science.

future development into a skill-based economy. Outside the United States, in-depth evidence on the impact of family background and school policies in educational production is very limited (cf. Hanushek 2002, pp. 3–4, 43–5). To my knowledge, recent comparable empirical evidence does not exist for East Asian countries. This paper starts to provide such evidence by estimating the impact of family background, resources, and other educational policies on student performance in five East Asian countries.

The evidence presented in this paper is based on student-level micro data from TIMSS, combining performance information with abundant data on students' family background and schools' resource endowments and institutional constraints (Section 1). The TIMSS database allows an estimation of education production functions for five East Asian countries: Hong Kong, Japan, South Korea, Singapore, and Thailand. Furthermore, the data and thus the estimated effects are directly comparable across these countries, as well as to countries in America and Europe. As discussed below, the multi-grade structure of the TIMSS sampling design also allows a credible identification of causal effects of class size on student performance in some of the countries.

The first set of analyzed influence factors is the impact of family background on students' educational performance in the different countries (Section 2). The research question is to what extent the different schooling systems provide equal educational opportunities for children from different family backgrounds. For example, the strong priority placed on education in South Korea since its earliest days stems largely from the desire to put "smallholders on an equal educational footing with the owners of larger farms – which was an important aspect of avoiding polarization in the countryside and of enabling migrants from the countryside to adapt relatively easily to urban and industrial life" (Ward and Richardson 2002, p. 17). The evidence presented in this paper suggests that the rural-urban performance difference is indeed relatively small in Korea. However, social background has a much larger impact on student performance in Korea, as well as Singapore, than in Hong Kong and Thailand.

One response to the concerns about schooling quality in these countries has been to substantially increasing educational spending (cf., e.g., Wrigley and Richardson, 2001). All countries concerned in this paper have substantially lowered their pupil-teacher ratios over recent decades (Gundlach and Wößmann, 2001). To see whether such policies can help to ensure a high-quality education, Section 3 analyzes the impact of resource endowments on students' academic skills in the East Asian countries. Least-squares estimates of the coefficient on several resource measures such as endowment with materials, instruction time, and teacher characteristics reveal few statistically significant correlations with student performance. However, as these standard estimates may be substantially biased by non-random resource endowments, the paper combines instrumental-variables with school-fixed-effects estimation to disentangle the causal effect of class size on student performance from any effects of placements of students into differently sized classes. Accounting for such resource endogeneity and omitted variable biases, class size does not seem to have a noteworthy causal effect on student performance in Japan and Singapore, the two countries for which the data allow a meaningful assessment.

Given the dismal results for resource policies, the question arises whether other policy options affect educational achievement in the East Asian countries (Section 4). For example, one complaint often heart all over the region is "that the government's administration of schools and universities is cumbersome, centralized and resistant to change" (Economist, 1997). Rather than centralized administration, giving more autonomy to schools may induce more creativity and make better use of localized knowledge on effective teaching techniques, particularly in school systems where performance is regularly accounted for in central examinations as in East Asia (cf. Wößmann, 2003). Accordingly, large positive effects of salary autonomy are found in Japan and Singapore, but no such effects are evident in Hong Kong and Korea. As another policy option, regular homework assignments have a statistically significant positive effect on student performance in Japan and Singapore.

It should be noted that the evidence presented in this paper mainly allows answers to questions relating to within-country variations in student performance. Thus, it shows the importance of different sets of influence factors for the performance variation within each country, and it allows for a comparison of the size of these effects across countries. By contrast, for questions relating to the most important determinants of the cross-country variation in test scores, the most promising way is to use the entire international dataset in order to link cross-country performance differences to cross-country differences in potential determinants. Such cross-country analyses have been performed elsewhere, both at the country level (e.g., Lee and Barro, 2001) and at the student level (Wößmann, 2002). However, an analysis of the relative effects of the different influence factors on the within-country variation across countries can help to understand better how the East Asian countries achieve their high educational standards, and it can yield implications for educational and social policies both in these countries and in other countries that strive to learn from the East Asian education systems.

1. The TIMSS Database for East Asian Countries

The database used to estimate education production functions for the five East Asian countries draws from a large-scale cross-country comparative test of student achievement, the Third International Mathematics and Science Study (TIMSS). It combines individual student-level performance data with information from student, teacher, and school-principal background questionnaires for nationally representative samples of students in each of the countries. TIMSS was conducted in 1995 under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), an independent cooperation of national research institutes and governmental research agencies. The target population of middle school students to which each participating country administered the test was defined as those students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing. These are the first two grades of secondary school in all the East Asian countries, representing the seventh and eighth year of formal education.

Each participating country randomly sampled the schools to be tested in a stratified sampling design, and within each of these schools, generally one class was randomly chosen from each of the two grades and all of its students were tested, yielding a representative sample of students within each country. The number of sampled schools that participated in the TIMSS test in each country is about 150, with the exception of Hong Kong, where it is 86.[2] Sample sizes range from 5827 students in Korea to 11643 students in Thailand.

---

2 In Singapore, all eligible schools were included in TIMSS (Martin and Kelly 1998, p. B-23).

TIMSS gave rigorous attention to quality control, using standardized procedures to ensure comparability in school and student sampling, to prevent bias, and to assure quality in test design and development, data collection, scoring procedures, and analysis. The TIMSS achievement tests were developed through an international consensus-building process involving inputs from international experts in math, science, and measurement, and were endorsed by all participating countries. Students were tested in a wide array of content dimensions, expecting skills that range from routine to complex procedures. A quarter of the test items (meant to cover a third of the testing time) were in free-response format, sometimes requiring extensive responses, while the remainder of the items were multiple-choice questions. A test-curriculum matching analysis which restricted the analysis to items definitely covered in each country's curriculum showed that the overall achievement patterns in TIMSS were hardly affected by this restriction.

Student performance is measured on an international achievement scale with scores having an international mean of 500 and an international standard deviation of 100. The mean math performance in the East Asian countries ranges from 508.3 test-score points in Thailand to 622.3 in Singapore. The variation in performance as indicated by the standard deviation of test scores in each country is relatively low in Thailand at 83.4, and it is relatively high in Korea at 107.8.[3]

The performance data are merged with the specific background data from three different TIMSS background questionnaires for each individual student. From the student background questionnaires, I draw information on age and sex of the student, on whether the student was born in the country and lives with both parents, the level of the parents' education, and the number of books at home. The teacher background questionnaires contain data on the actual class sizes, as well as on teacher characteristics such as sex, years of experience, and education level. They also report the amount of homework assignments per week and whether teaching was thought to be limited by uninterested or interested parents. The school-principal background questionnaires provide information on the community location of the school, shortage

[3] For detailed descriptive statistics on all variables used in this paper, as well as for background information on the East Asian schooling systems, see [the working-paper version of this study]. [Note to referees: The tables with descriptive statistics are reported as a non-publishable appendix at the end of this version.]

of materials, instruction time, average class sizes in the two relevant grades, and on whether the school had responsibility for determining teacher salaries. Most of these background variables based on qualitative survey data were transformed into dummy variables for the estimations of this paper.

Complete performance data is available for all the students participating in TIMSS. In the background questionnaires, however, some students, teachers, and school principals failed to answer some questionnaire items. Since dropping all students with missing data on some explanatory variables from the analyses deletes the information available on the other explanatory variables, reduces the sample size, and might introduce bias if observations are not missing at random, I chose instead to impute missing values within each country for the analyses in this paper.[4] I use a set of "fundamental" explanatory variables F with original data available for virtually all students to impute missing data on each variable M for each student i within each country. Let S denote the set of students j with available data for M. Using the students in S, the variable M is regressed on F:

$$M_{j \in S} = F_{j \in S} \phi + \varepsilon_{j \in S} \tag{1}$$

The regression model is a weighted least-squares estimation (weighting each student by its sampling probability) if M is a discrete variable, a weighted probit model if M is a dichotomous (binary) variable, and a weighted ordered-probit model if M is a polychotomous qualitative variable with multiple categories. The coefficients ϕ from these regressions and the data on F_i are then used to impute the value of M_i for the students with missing data:

$$\widetilde{M}_{i \notin S} = F_{i \notin S} \phi \tag{2}$$

For the probit models, the estimated coefficients were used to forecast the probability of occurrence associated with each category for the students with missing data, and the category with the highest probability was imputed. For the purposes of this paper, this data imputation technique is applied within each country individually, resulting in a complete data set for all the students sampled in TIMSS.

---

[4] See [the working-paper version of this study] for details on missing data and the imputation method.

Given the international standardization of the test results, the cooperative nature of the test development, its endorsement by all participating countries, and the substantial efforts to ensure high-quality sampling and testing in all countries, the TIMSS student performance and background data should be comparable across countries. This should also make the empirical estimates presented in this paper directly comparable across the different countries. This makes the database uniquely capable of using student, class, and school level data to analyze the determination of student performance in the five East Asian countries.

2. Family Background and Student Performance in East Asia

2.1. The Empirical Model

To assess the influence of the students' family background on their educational performance in the different East Asian countries, I estimate education production functions for each country of the following form:

$$T_{ics} = B_{ics}\alpha_1 + D^B_{ics}\delta_1 + \left(D^B_{ics}B_{ics}\right)\delta_2 + \varepsilon_{ics} \quad , \tag{3}$$

where T is the test score of student i in class c in school s and B is the vector of family background variables. The coefficient vectors α_1, δ_1, and δ_2 are to be estimated. The inclusion of the imputation controls D^B and the structure of the error term ε are discussed below. The estimation does not control for other school characteristics, such as schools' resource endowments or teaching policies, because in this section I am interested in the total impact of family background on student performance, including any effect that might work through families' differential access to schools or their influence on school policies.

It helps to clarify in advance what the estimates of the coefficients α_1 on the family-background variables (and of the coefficients on the other explanatory variables in later sections), and especially differences in the estimates across countries, mean and do not mean. Because the TIMSS data were generated by the same data-generating process in the different countries and are therefore directly comparable across countries, the prior from a technical point of view should be that the coefficient estimates should be the same everywhere. Given the technical constraints on the pedagogical process, the size of the *effect* of any family-background characteristic on students' educational performance should be expected to be the same in any school system. If this is not the case, this implies that there must be differences in how the school systems work. This does *not* reflect different *distributions* of family-background characteristics in the different populations. Different distributions of family-background characteristics would not be an a priori reason for the gap in student performance between students with two different characteristics to be different. For example, the performance gap between children of parents with university degrees and children of parents without secondary education may be expected to be independent of the relative number of parents with different educational degrees in the population. If this gap is 25 TIMSS test-score points in one country but 50 points in another country, this would rather be a sign that the school systems work differently in the two countries, resulting in a different effect of parental education on student performance.

As discussed in the previous section, some of the data are imputed rather than original. Generally, data imputation introduces measurement error in the explanatory variables, which should make it more difficult to observe statistically significant effects. Still, to make sure that the results are not driven by imputed data, a vector of dummy variables D^B is included as controls in the estimation. The vector D^B contains one dummy for each variable in the family-background vector B which takes the value of 1 for observations with missing and thus imputed data and 0 for observations with original data. The inclusion of D^B as controls in the estimation allows the observations with missing data on each variable to have their own intercepts. The inclusion of the interaction term between imputation dummies and background data, $D^B B$, allows them to also have their own slopes for the respective variable. These imputation controls for every variable with missing values ensure that the results are robust against possible bias arising from data imputation.

Further problems in the econometric estimation equation (3) are that the explanatory variables in this study are measured at different levels, with some of them not varying within classes or schools; that the performance of students within the same school may not be independent from one another; and that the primary sampling unit (PSU) of the two-stage clustered sampling design in TIMSS was the school, not the individual student (see Section 1). As shown by Moulton (1986), a hierarchical structure of the data requires the addition of higher-level error components to avoid spurious results. Therefore, the error term ε of equation (3) has a school-level and a class-level element in addition to the individual-student element:

$$\varepsilon_{ics} = \eta_s + \nu_c + \upsilon_i \quad , \tag{4}$$

where η is a school-specific error component, ν is a class-specific error component, and υ is a student-specific error component. Clustering-robust linear regression (CRLR) is used to estimate standard errors that recognize this clustering of the survey design. The CRLR method relaxes the independence assumption and requires only that the observations be independent across the primary sampling units, which are schools in the case of TIMSS. By allowing any given amount of correlation within the primary sampling units, CRLR estimates appropriate standard errors when many observations share the same value on some but not all independent variables (cf. Deaton, 1997).

Finally, TIMSS used a stratified sampling design within each country, which produced varying sampling probabilities for different students (Martin and Kelly, 1998). To obtain nationally representative coefficient estimates from the stratified survey data, weighted least squares (WLS) estimation using the sampling probabilities as weights is employed. The WLS estimation ensures that the proportional contribution to the parameter estimates of each stratum in the sample is the same as would have been obtained in a complete census enumeration (DuMouchel and Duncan, 1983; Wooldridge, 2001).

2.2. Results

Table 1 presents the results of an estimation of equation (3) for each of the sample countries for TIMSS math performance.[5] To allow a comparison of the East Asian findings to countries from other regions in the world, all estimations are also executed for the United States and France, the latter being a European country with reasonably complete TIMSS data. With respect to students' characteristics, students in the upper grade (eighth grade) perform statistically significantly better than students in the lower grade (seventh grade) in all countries, with the gap being largest in Singapore and

5 All the results shown here for math are also reported for science in [the working-paper version of this study].

smallest in Japan. In Japan, much of the superior performance of older students seems to be captured by students' age rather than grade level, as older students perform statistically significantly better in both subjects in Japan. In Hong Kong, Korea, and Singapore, older students perform statistically significantly worse once the grade level is held constant.

In Hong Kong, Japan, and Korea, girls perform substantially worse than boys – a result similarly found in the two advanced economies (United States and France). Singapore and Thailand show no such performance difference between genders, with girls performing statistically insignificantly better than boys. The performance gap between native and immigrant children is quite different between the East Asian countries. In Korea and Thailand, children born in the respective country performed better – although the share of immigrant children is very low in these two countries. But in Hong Kong, children not born in the country actually performed better. No statistically significant performance difference between natives and immigrants is found in Singapore. Students living with both parents perform better in Hong Kong and Korea.[6]

Two sets of dummy variables reflect the educational background of the students' families: the highest level achieved by the parents and the number of books in the students' home. For both categorical variables, the lowest category – primary education and less than one shelf of books, respectively – was dropped as residual category in the estimation. In all the countries, children from more favorable backgrounds on both measures perform consistently better. The largest performance difference between children of parents with a university degree relative to children of parents without secondary education are found in Singapore. The same is true when comparing parents who finished university to parents who finished secondary school. The size of the coefficient says that, for example, the performance gap between students of parents with a university degree and students of parents without secondary education in Singapore in math was 52.7 test-score points – slightly more than half an international standard deviation in TIMSS test scores, and slightly less than the average difference in performance between seventh and eighth grade in Singapore.

---

6 In Japan, much of the family-background data are reported as being not administered or not internationally comparable.

Because parental education levels may be slightly differently defined in the different countries, possibly reflecting different years and courses of education, it is illuminating to look at the performance levels of students with different numbers of books at home, which can work as an internationally comparable additional proxy for the educational background of a student's family. Using this measure, the impact of family background on students' educational achievement is again substantially larger in Korea and Singapore than it is in Hong Kong and Thailand. This is true irrespective of whether one compares the highest category of books at home to the lowest one, the highest one to some intermediate one, or an intermediate one to the lowest one. On this measure, the impact of family background in Korea is even stronger than in the United States, a country with a schooling system generally known to produce relatively large performance differences between students from different backgrounds. In Hong Kong and Thailand, the measure points to a smaller impact of family background than the one found in the two advanced economies.

The statistically significant and quantitatively substantial coefficients on the family-background variables cannot necessarily be interpreted in the sense that, for example, increasing parental education for the whole population in the different countries would increase educational performance of the students by the amount estimated. Rather, the coefficient estimates may to some extent reflect heritable ability in that more able parents, who may have obtained more education because of their higher ability levels, have more able children, who then perform better on the performance tests. Heritable ability has been shown to be a likely source of the whole correlation between the quantitative educational attainment of mothers and their children in data on Minnesota twins (Behrman and Rosenzweig, 2002). This was not true for fathers, however, and other evidence shows that there was a causal impact of increased women's schooling on their children's schooling, working through home teaching, in the setting of rural India during the green revolution (Behrman et al., 1999). Whatever the sources and channels of transmission may be, the reduced-form results of Table 1 still represent the observed performance gap between children from different family backgrounds in the schooling systems of the different East Asian countries.

Student performance also differs by community location in most of the East Asian countries. In Hong Kong, Korea, Singapore, and Thailand, students in schools close to the center of a town perform statistically significantly better than students in schools located in villages or at the outskirts of a town (the residual category). This rural-urban performance gap is smaller in Korea than in the other three countries, and it is not statistically significant in Japan, France, and the United States. Student performance in geographically isolated areas is generally even worse than performance in village or outskirt areas, although except for Thailand, none of the TIMSS samples in the East Asian countries contains a noteworthy share of geographically isolated schools.[7]

The explanatory power of the family-background regressions, as measured by the proportion of the variation in test scores explained by the family-background variables (the R^2), ranges from 10.2 percent in Hong Kong to 16.9 percent in Singapore (without considering the variation "explained" by the imputation controls).[8] The standard finding of a large residual in microeconometric student-level estimations may be attributed to unobserved heterogeneity in the innate ability of students entering the error term in student-level education production functions. Across the East Asian countries, the explained performance variation is relatively small in Hong Kong and Thailand, both in comparison to Korea and Singapore and to the more advanced economies.

3. Resource Endowments and Student Performance in East Asia

3.1. Least-Squares Coefficients on Resources and Teacher Characteristics

The standard procedure to estimate the relationship between schools' resource endowments and their students' performance is to simply introduce resources into the previously estimated equation (3):

$$T_{ics} = B_{ics}\alpha_2 + R_{cs}\beta_1 + D_{ics}^B\delta_3 + \left(D_{ics}^B B_{ics}\right)\delta_4 + D_{cs}^R\delta_5 + \left(D_{cs}^R R_{cs}\right)\delta_6 + \varepsilon_{ics} \quad , \tag{5}$$

where R is a vector of resource measures such as class size, the availability of instructional materials, and teacher characteristics. The imputation controls D^R again ensure that the results are robust against possible bias arising from missing and thus imputed data in the resource variables.

---

7 The number schools classified as being located in geographically isolated areas is only 2 in the Hong Kong sample, 1 in Korea, 4 in Japan, and 0 in Singapore.

8 The low R^2 of the Japanese regression obviously reflects the fact that most of the family background data are missing in Japan.

Under the assumption that the resource endowment is exogenous to student performance – an assumption shown to be wrong in the next section at least in the case of class sizes in most countries – the coefficient vector β_1 estimated in a least-squares regression would reflect the impact of resources on student performance. The coefficient vector on resources obtained by this standard procedure may be substantially biased, however. One potential reason for bias is that the resource endowment may to some extent be endogenous to student performance, for example if weaker students are sorted into smaller classes (cf. West and Wößmann, 2003). Another potential reason for bias is the impact of further omitted variables which, like sorting, could be related to the resource endowment.

Table 2 presents the estimated least-squares coefficients on resources, controlling for all the family-background variables reported in Table 1 and for all the imputation controls. Class size is measured in natural logarithm units because the proportional impact of a one-student reduction in class size is greater the smaller the initial size of the class. Except for Thailand and Korea, the estimated coefficients on log class size are statistically significant and *positive* in the East Asian countries; that is, higher test scores are related to *larger* classes. If one were to interpret these coefficients causally, as much previous work for other countries has done (cf., e.g., Hanushek, 2002; Krueger, 2003), one would come to the counterintuitive conclusion that in most East Asian countries, students learn more in larger classes.

Students whose school principal reported no shortage of instructional materials perform statistically significantly better in some of the East Asian countries than students whose principal reported some shortages. However, students whose principal reported a lot of shortage do not perform statistically significantly worse, and in Japan, they even perform statistically significantly better. Only in Thailand is the length of instruction time statistically significantly related to student performance.

With respect to teacher characteristics, students of female teachers performed statistically significantly worse than students of male teachers in Japan. Teacher experience, measured in logs so as to allow for decreasing returns to experience, is statistically significantly positively related to student performance in Singapore and Thailand; in Korea, there is a statistically significant negative relation. Teachers' educational levels also do not seem to be strongly related to student performance. In

Thailand, students of teachers with the equivalent of a BA actually performed statistically significantly lower than students of teacher with less education than a BA.[9]

In conclusion, there is basically not much of a positive relationship between student performance and additional units of any of the measured resource variables. These findings mirror prior research in this field that found no strong or systematic relationship between larger resource endowments and student performance in the United States and in several developing countries (Hanushek, 2002). Note also that the increase in the explained proportion of the test-score variation (R^2) relative to the family-background regressions of Table 1 is minimal in most cases, and where it is not, this is nearly exclusively driven by the counterintuitive correlation between student performance and class size.

3.2. School-Fixed-Effects Instrumental-Variables Estimates of Class-Size Effects

While the family-background measures B in the estimated equations (3) and (5) can reasonably be expected to be exogenous to student performance because there appears to be no plausible inverse link from student performance to family background, there may potentially be endogeneity of schooling resources R. The quantitative estimates of the resource effects will be biased if the resources spent on students are determined by student performance T, that is if additional schooling resources are systematically allocated either to above-average performing students or to below-average performing students. The estimates of resource effects would also pick up the correlation between student performance and any omitted variable that is correlated with resource endowment. In both cases, unbiased econometric estimates can only result if the endogenous nature of schooling resources is properly accounted for (Hoxby, 2000).

In the case of the estimated coefficients on class size, I can exploit specific characteristics of the TIMSS data in a quasi-experimental estimation design in order to obtain unbiased estimates of the effects of class size on student performance. Akerhielm (1995) suggests to instrument the actual class size C_{cs} (one vector in the resource matrix R_{cs} of equation (5)) by the average class size in the school A_s in a two-stage least-squares estimation to control for the problem of endogenous resource allocation within

[9] The residual category that drives the statistically significant coefficients in Korea is made up of only one teacher.

schools.[10] The grade-average class size promises to be a valid instrument for actual class size: It is generally strongly linked to the size of the class actually tested in TIMSS; within each school, it is exogenous to the performance of the students (although this might not be the case between schools, a fact that I will return to shortly); and there is no reason to expect that it affects student performance in any other way than through the size of the class in which they are actually taught.[11] The first-stage estimation regresses (log) C_{cs} on (log) A_s and all other exogenous variables X_{ics}:

$$C_{cs} = \chi_1 A_s + X_{ics}\chi_2 + \mu_{ics} \quad , \tag{6}$$

where X_{ics} includes the family-background measures and the imputation controls. The second stage then employs $\hat{C}_{cs} = C_{cs} - \mu_{ics}$ instead of C_{cs} in lieu of R_{cs} in the estimation of equation (5). This specification eliminates any bias in the estimated class-size effects that would result from within-school sorting of low-performing students, at a given grade level, to smaller classes.

However, these IV estimates may still be biased by between-school sorting effects. If parents tend to send low-performing children to schools with smaller classes, the estimated resource effect would again be biased downward. But it could also go the other way if parents tend to send high-performing children to schools with smaller classes. Between-school sorting might also be relevant if students are tracked into different schools according to their ability, as is the case in Singapore.

In order to exclude any effects of either within- or between-school sorting from the estimates of class-size effects, Wößmann and West (2002) suggest an identification strategy specifically designed to exploit the multi-grade nature of the TIMSS database. They combine the aforementioned IV strategy with a school-fixed-effects estimation which disregards any between-school variation, as this may reflect between-school sorting effects. The combined school-fixed-effects instrumental-variables (SFE-IV) estimation then is:

---

10 Akerhielm (1995) also uses the overall grade-level enrollment of a school as a second instrument in addition to average class size. However, this may be a false instrument as there might be a direct relationship between overall enrollment and student performance that is unrelated to differences in class size (Angrist and Lavy, 1999). Moreover, none of the coefficients on enrollment in Akerhielm's first-stage regressions are statistically significant, suggesting that it is anyway not a good instrument.

11 See Wößmann and West (2002) for a more detailed discussion of the validity of the instrument.

$$T_{ics} = B_{ics}\alpha_3 + \hat{C}_{cs}\beta_2 + S_s\varphi + D^B_{ics}\delta_7 + \left(D^B_{ics}B_{ics}\right)\delta_8 + D^C_{cs}\delta_9 + \left(D^C_{cs}C_{cs}\right)\delta_{10} + \varepsilon_{ics} \quad , \tag{7}$$

where S_s is a complete set of school dummies and $\hat{C}_{cs}$ is again the result of a first-stage regression that instruments actual class size by grade-average class size and all other exogenous variables as in equation (6).[12] Because equation (7) includes school fixed effects, and because every class size at a given grade level is instrumented by the same average class size, this SFE-IV strategy requires comparable information on student performance from more than one grade level in each school. This is exactly the structure of the TIMSS data.

The grade-level dummy included in the background measures *B* controls for the average difference in performance between students from the two adjacent grades. Therefore, the remaining performance difference between students from the different grades is idiosyncratic to each school. Equation (7) relates this idiosyncratic variation in student performance to that part of the actual class-size difference between the two grades that is due to differences in average class size between the two grades. Thereby, the SFE-IV identification strategy effectively excludes both between-school and within-school sources of student sorting: Between-school sorting is eliminated by controlling for school fixed effects; within-school sorting is filtered out by instrumenting actual class sizes by grade-average class size. Arguably, the remaining variation in class size between classes at different grades of a school is caused by random fluctuations in cohort sizes between the two adjacent grades in each school, presumably reflecting natural fluctuations in student enrollment. The coefficient estimate β_2 can thus be interpreted as an unbiased estimate of the causal impact of class size on student performance.[13]

Table 3 reports the coefficient estimates on class size obtained by implementing the different identification strategies for the East Asian countries. The first row presents the

---

[12] The imputation dummies D^C for the class-size variable used in this section equal 1 if either the observation on actual class size or the observation on grade-average class size (the instrument) is imputed. In the IV and SFE-IV regressions, in addition to instrumenting class size, the interaction term $D^C C$ between the imputation dummy and actual class size is also instrumented, using an interaction term $D^C A$ between the imputation dummy and grade-average class size as an additional instrument.

[13] As there is no comparable quasi-experimental identification strategy for the other resource measures, these are not included in equations (6) and (7). Therefore, the resulting coefficient estimates on class size should be interpreted as the effect on student performance of class size and any other resource with which class size may be associated.

standard weighted least-squares (LS) estimates, where the slight differences to the coefficients reported in Table 2 stem from the exclusion of the other resource variables.[14] The second row reports results of the straight IV regression without controlling for school fixed effects, which should exclude biases due to within-school sorting but not due to between-school sorting. The third row reports results of a least-squares regression that does not instrument for class size but includes the whole set of school fixed effects (SFE), which excludes any effects of between-school sorting but might still be biased by within-school sorting effects. And finally, the fourth row reports results of the combined SFE-IV identification strategy that excludes both between- and within-school sorting effects.

The SFE-IV estimation is extremely demanding in terms of data requirements, because the variation on which it is based excludes both any between-school variation and any within-grade variation within schools. If the remaining within-school between-grade variation is low, this will be reflected in imprecise estimates of the class-size coefficient estimated by the SFE-IV strategy (cf. Wößmann and West, 2002). This is the case in Hong Kong and Thailand, where the standard errors of the SFE-IV estimates are too large to make any confident statement about the existence or magnitude of class-size effects in these countries. By contrast, in Japan and Singapore the SFE-IV estimates are very precise, with standard errors of about 20. These standard errors are so small that if a 10 percent reduction in class size were to change TIMSS test scores by just 4 test-score points or 4 percent of an international standard deviation, the change would be statistically significant at the 5 percent level. In other words, the random variations in class size identified by the SFE-IV strategy have considerable power to detect class-size effects in these two countries.

The SFE-IV estimates of the causal effect of class size on student performance are statistically indistinguishable from zero in Japan and Singapore. Given the precision of their estimation, they are equivalent to what Hoxby (2000, p. 1280) calls "rather precisely estimated zeros." These results suggest that there is no causal effect of class size on student performance in Japan and Singapore. By contrast, the SFE-IV estimate

[14] In order to be able to implement the school-fixed-effects strategy, I also had to exclude one school from the Hong Kong sample and one from the Thai sample which tested only classes at one of the two grade levels. In the United States and France, this exclusion rate was slightly larger.

for France in math is marginally statistically significant (at the 15 percent confidence level) and negative, suggesting a potential beneficial effect of reduced class sizes there.

The strong prevalence of statistically significant positive estimates of the coefficient on class size in least-squares estimations in East Asian countries is clearly linked to the endogeneity of class size with respect to student performance. The differences in the estimated coefficients between the four estimation strategies reported in Table 3 imply that there is substantial sorting of students into differently sized classrooms based on their achievement levels in the East Asian school systems. Particularly in Japan and Singapore, the differences between the LS and the SFE estimates suggest that low-performing students are sorted into schools with smaller classes.[15] Once the estimation is based on credibly exogenous variations in class size in the SFE-IV estimation, no statistically significant effect of class size on student performance is found in the East Asian countries. While the existence of any sizable causal effect of class size on student performance can be rejected in Japan and Singapore, no confident evaluation is possible in the other three countries given the imprecision of their SFE-IV estimates.

4. Institutional Features and Student Performance in East Asia

The lack of consistent evidence that resource endowments matter for student performance suggests that resources are inefficiently used in the school systems analyzed. In other countries, such inefficiencies have been related to the lack of suitable performance incentives in the school system (e.g., Hanushek et al., 1994). This opens the possibility for other schooling policies that focus on institutions rather than on resources to affect student performance. Theoretical work suggests that the institutional structure of the school system generates the incentives that drive actors' behavior in educational production and thus the performance achieved (cf., e.g., Bishop and Wößmann, 2003).

Because institutional features generally do not vary substantially within school systems, but rather across countries, empirically the institutional effects should be mainly an issue in cross-country rather than within-country research. Wößmann (2002)

[15] See West and Wößmann (2003) for a detailed analysis of the pattern of sorting between and within schools.

shows that many schooling institutions are strongly linked to the cross-country variation in student performance. The TIMSS background data reveal that some institutional features do also vary within some of the East Asian systems. Particularly, there is some limited variation in schools' autonomy in salary decisions, homework policies, and parental involvement in the education process. This section analyzes whether these within-country differences in institutional schooling policies add to an understanding of the within-country differences in student performance in East Asia.

As institutional features of the school systems may be viewed as largely exogenous to student performance, reasonable estimates of institutional effects may be obtained by adding the vector of institutional measures *I* as explanatory variables to the education production function of equation (5):

$$\begin{aligned} T_{ics} &= B_{ics}\alpha_4 + R_{cs}\beta_3 + I_{cs}\gamma \\ &\quad + D^B_{ics}\delta_{11} + \left(D^B_{ics}B_{ics}\right)\delta_{12} + D^R_{cs}\delta_{13} + \left(D^R_{cs}R_{cs}\right)\delta_{14} + D^I_{cs}\delta_{15} + \left(D^I_{cs}I_{cs}\right)\delta_{16} + \varepsilon_{ics} \end{aligned} \quad . \qquad (8)$$

D^I is again a set of imputation dummies to control for possible effects of the data imputation. The estimation keeps controlling for all family background and resource variables of Tables 1 and 2, as well as for their respective imputation controls.[16]

The coefficient estimates on the institutional variables are reported in Table 4. Students in schools that had autonomy in determining their teachers' salaries performed statistically significantly better than students in schools without salary autonomy in Japan and Singapore. In these countries, school autonomy in determining teacher salaries seems to positively affect students' educational performance.

The amount of homework assigned by the teacher is statistically significantly and positively related to performance in Japan and Singapore. Thus, to the extent that teachers' homework assignments can be viewed as exogenous to student achievement, they seem to favorably affect achievement in the East Asian countries, excepting Korea and Thailand. The estimates on homework assignments should be interpreted with care, however, as they may be particularly prone to endogeneity and omitted-variable biases.

16 Excluding the resource variables and their imputation controls, because their estimation may be biased by sorting effects, does not make any qualitative difference to the estimated coefficients on the institutional variables.

In Hong Kong, students whose teachers reported that their teaching was limited by parents uninterested in students' progress performed statistically significantly worse than students whose teachers did not report limitations by uninterested parents. Interestingly, students whose teachers reported that their teaching was limited by interested parents performed statistically significantly better than students whose teachers did not report such limitations. Apparently, even though teachers judged the interventions of interested parents as limiting their teaching, this "limitation" was positively related to the performance of their students – a result similarly found in the United States.[17]

5. Conclusions

Given the pivotal role of students' educational performance for the future economic prospects of societies, the empirical results of education production functions estimated for the five high-performing East Asian countries in this paper could have substantial implications for educational and social policies in the region and in other, lower-performing countries alike. For the East Asian countries, the evidence for the first time reveals the impact of family background and schooling policies in the different school systems. And by examining how the East Asian countries achieved their high educational performance, other countries can learn for their own educational production.

Although the fact that all East Asian countries performed extraordinarily well in international comparisons of student performance seems to suggest that they are very homogenous, the evidence presented in this paper reveals that their schooling systems actually feature a lot of heterogeneity. For example, family background is a much stronger predictor of children's educational performance in Korea and Singapore than in Hong Kong and Thailand, both in terms of estimated effect sizes and explanatory power. If providing more equal opportunities for successful learning independent of parental education and social status is an important goal of the education systems, the different size of family-background effects across countries reveals that the different

[17] The large negative coefficient on interested parents in Japan in math is due to only 2 teachers reporting limitations by interested parents.

schooling systems achieve this goal to a different extent. Furthermore, the evidence from the different countries suggests that those school systems that allow family background to exert its beneficial impact on student performance achieve the highest overall performance levels. In reverse, this may mean that although school systems that try to equalize educational performance for students from different backgrounds may be able to lower the variation in educational performance in the population, the overall educational performance of the system may suffer.

The high educational performance of East Asian countries also suggests that their schooling systems are highly efficient. While this is true in the sense of a cross-country comparison between East Asian countries and countries from other parts of the world, the internal efficiency of the East Asian school systems is less clear. The evidence presented in this paper reveals that resource endowments and especially class sizes do not seem to be strongly related to students' educational achievement. As in many other countries in the world, East Asian schools that are better equipped with educational resources do not seem to make efficient use of the additional resources. This cross-sectional finding mirrors the time-series evidence of Gundlach and Wößmann (2001) that increased spending and smaller class sizes did not lead to substantially better performance over time in the analyzed East Asian school systems.

With respect to other, more institutional schooling policies, giving schools autonomy in their salary decisions might strengthen educational performance, especially in Japan and Singapore. Given that performance standards are centrally set and examined in all the East Asian systems considered, additional autonomy might allow schools to find the best ways of how to achieve these standards. Additional focus on homework policies, which allow students to practice their knowledge at home, might be a worthwhile policy option, especially in Hong Kong, Japan, and Singapore. In Hong Kong, increased parental involvement in the teaching process also promises superior student performance.

Most of the results reported here for math performance also hold for science performance (cf. [the working-paper version of this study]). It remains to be seen whether the conclusions of this paper also apply for other subjects and skills than middle-school mastery of math and science. Some evidence suggests that East Asian students are not just capable of rote learning, but also do well in more creative tasks.

Learning the cognitive foundations is certainly a prerequisite for the mastery of more advanced applications, so that the two are complements rather than substitutes. To sustain the quality of this knowledge base and to tap the full potential of their student populations, East Asian school systems would be well advised to ensure an excellent educational performance for students from all family backgrounds and to care more for policies that ensure efficient educational production than for resource policies.

References

Akerhielm, K. (1995). 'Does Class Size Matter?', *Economics of Education Review*, vol. 14 (3), pp. 229–41.

Angrist, J.D. and Lavy, V. (1999). 'Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement', *Quarterly Journal of Economics*, vol. 114 (2), pp. 533–75.

Beaton, A.E., et al. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*, Chestnut Hill, MA: Boston College.

Behrman, J.R., Foster, A.D., Rosenzweig, M.R. and Vashishtha, P. (1999). 'Women's Schooling, Home Teaching, and Economic Growth', *Journal of Political Economy*, vol. 107 (4), pp. 682–714.

Behrman, J.R. and Rosenzweig, M.R. (2002). 'Does Increasing Women's Schooling Raise the Schooling of the Next Generation?', *American Economic Review* 92, vol. (1), pp. 323–34.

Bishop, J.H. and Wößmann, L. (2003). 'Institutional Effects in a Simple Model of Educational Production', *Education Economics*, forthcoming (Kiel Working Paper 1085).

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

Dosanjh, K. and Richardson, G. eds (2001). *Country Profile 2001: Singapore*, London: The Economist Intelligence Unit.

DuMouchel, W.H. and Duncan, G.J. (1983). 'Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples', *Journal of the American Statistical Association*, vol. 78 (383), pp. 535–43.

Economist (1997). 'South-East Asia's Learning Difficulties', *The Economist* (August 14).

Gundlach, E. and Wößmann, L. (2001). 'The Fading Productivity of Schooling in East Asia', *Journal of Asian Economics*, vol. 12 (3), pp. 401–17.

Hanushek, E.A., et al. (1994). *Making Schools Work: Improving Performance and Controlling Costs*, Washington, D.C.: The Brookings Institution.

Hanushek, E.A. (2002). 'Publicly Provided Education', in (A. Auerbach and M. Feldstein, eds), *Handbook of Public Economics*, vol. 4, pp. 2045–141, Amsterdam: Elsevier.

Hoxby, C.M. (2000). 'The Effects of Class Size on Student Achievement: New Evidence from Population Variation', *Quarterly Journal of Economics*, vol. 115 (4), pp. 1239–85.

Krueger, A.B. (2003). 'Economic Considerations and Class Size', ECONOMIC JOURNAL, vol. 113 (February), forthcoming (NBER Working Paper 8875).

Lee, J.-W. and Barro, R.J. (2001). ‘Schooling Quality in a Cross-Section of Countries’, *Economica*, vol. 68 (272), pp. 465–88.

Martin, M.O. and Kelly, D.L., eds (1998). *TIMSS Technical Report Volume II: Implementation and Analysis, Primary and Middle School Years*, Chestnut Hill, MA: Boston College.

Moulton, B.R. (1986). ‘Random Group Effects and the Precision of Regression Estimates’, *Journal of Econometrics*, vol. 32 (3), pp. 385–97.

Richardson, G., ed (1996). *Country Profile: Singapore, 1996-97*, London: The Economist Intelligence Unit.

Stevenson, H.W. (1992). ‘Learning from Asian Schools’, *Scientific American* (December), pp. 32–8.

Ward, R. and Richardson, G., eds (2002). *Country Profile 2002: South Korea*, London: The Economist Intelligence Unit.

West, M.R. and Wößmann, L. (2003). ‘Which School Systems Sort Weaker Students into Smaller Classes? International Evidence’, mimeo, Harvard University.

Wooldridge, J.M. (2001). ‘Asymptotic Properties of Weighted *M*-Estimators for Standard Stratified Samples’, *Econometric Theory*, vol. 17 (2), pp. 451–70.

Wößmann, L. (2002). ‘Schooling Resources, Educational Institutions, and Student Performance: The International Evidence’, Kiel Working Paper 983, Revised Version, December, Kiel: Institute for World Economics.

Wößmann, L. (2003). ‘How Central Exit Exams Affect Educational Achievement: International Evidence from TIMSS and TIMSS-Repeat’, forthcoming in (P.E. Peterson and M.R. West, eds), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington, D.C.: Brookings Institution Press.

Wößmann L. and West, M.R. (2002). ‘Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS’, PEPG Research Paper 02-02, Cambridge, MA: Harvard University.

Wrigley, D. and Richardson, G., eds (2001). *Country Profile 2001: Hong Kong*, London: The Economist Intelligence Unit.

Table 1: Family Background and Student Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	HON	JAP	KOR	SIN	THA	USA	FRA
Upper grade	37.795*	10.636+	38.731*	61.346*	26.377*	45.988*	67.387*
	(6.040)	*(4.502)*	*(4.406)*	*(5.021)*	*(4.296)*	*(4.439)*	*(4.359)*
Age	-11.571*	23.389*	-4.601	-14.685*	1.802	-22.088*	-24.737*
	(2.881)	*(3.850)*	*(3.975)*	*(1.922)*	*(2.180)*	*(2.610)*	*(2.358)*
Female	-13.519+	-10.385*	-15.926*	1.643	3.194	-9.006*	-10.691*
	(6.433)	*(3.254)*	*(3.182)*	*(4.320)*	*(3.121)*	*(2.335)*	*(2.044)*
Born in country	-17.544*	–	26.578	-5.065	28.679*	1.565	–
	(5.204)		*(17.753)*	*(3.914)*	*(9.100)*	*(4.566)*	
Living with both parents	9.258+	–	9.156+	5.222	0.400	15.476*	7.819*
	(4.285)		*(4.395)*	*(3.977)*	*(2.847)*	*(2.888)*	*(2.461)*
Parents' education							
Some secondary	0.019	–	0.738	–	1.204	11.061	8.377
	(3.185)		*(6.294)*		*(3.872)*	*(8.632)*	*(6.642)*
Finished secondary	13.341*	–	12.408°	13.754*	15.808*	17.203°	19.628*
	(3.716)		*(6.410)*	*(3.008)*	*(5.596)*	*(8.831)*	*(6.884)*
Some after secondary	29.154*	–	0.419	41.491*	39.924*	31.478*	21.237*
	(6.997)		*(7.844)*	*(4.520)*	*(6.283)*	*(8.288)*	*(7.155)*
Finished university	34.259*	–	41.639*	52.672*	40.557*	52.663*	38.249*
	(6.215)		*(7.218)*	*(5.738)*	*(9.150)*	*(9.160)*	*(6.936)*
Books at home							
One shelf (11-25)	17.943*	–	19.571*	8.573+	3.373	9.746+	-4.621
	(4.487)		*(6.704)*	*(3.597)*	*(2.208)*	*(3.779)*	*(4.977)*
One bookcase (26-100)	23.566*	–	57.779*	32.684*	9.736*	34.571*	8.747°
	(4.774)		*(4.997)*	*(3.611)*	*(3.076)*	*(3.560)*	*(4.672)*
Two bookcases (101-200)	18.297*	–	84.691*	43.718*	13.437*	53.481*	16.634*
	(5.353)		*(5.344)*	*(4.851)*	*(3.468)*	*(4.229)*	*(4.927)*
More than two bookcases	21.669*	–	97.397*	47.075*	10.980*	62.607*	11.165+
(>200)	*(5.908)*		*(5.235)*	*(5.387)*	*(3.930)*	*(4.747)*	*(5.234)*
Community location							
Close to town center	25.968°	-7.190	12.042*	16.791+	34.353+	-4.106	2.253
	(14.083)	*(6.467)*	*(3.589)*	*(8.124)*	*(13.548)*	*(6.639)*	*(5.232)*
Geographically isolated	-49.538°	-18.230	0.659	–	-10.975	-28.904*	–
	(25.728)	*(20.163)*	*(3.923)*		*(7.237)*	*(7.948)*	
Imputation controls	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6752	10271	5827	8285	11643	10973	6014
Schools [Unit of clustering]	86	151	150	137	147	183	134
R^2	0.144	0.038	0.179	0.154	0.119	0.185	0.230
R^2 (without imput. controls)	0.102	0.037	0.169	0.152	0.115	0.175	0.211

Significance levels (based on clustering-robust standard errors): * 1 percent. – + 5 percent. – ° 10 percent.

Table 2: Resources, Teacher Characteristics, and Student Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	HON	JAP	KOR	SIN	THA	USA	FRA
Class size (log)	106.206 *	123.908 *	-3.469	137.201 *	7.850	-3.716	63.962 *
	(35.471)	*(36.010)*	*(4.188)*	*(11.681)*	*(7.408)*	*(6.441)*	*(18.845)*
Shortage of materials							
None	16.000	7.754 °	0.921	13.521 °	24.237	-1.669	7.886
	(12.714)	*(4.360)*	*(3.680)*	*(7.151)*	*(18.338)*	*(6.055)*	*(5.076)*
A lot	-29.598	20.216 +	-0.087	-7.418	7.108	-28.585 +	4.839
	(32.201)	*(9.986)*	*(5.036)*	*(9.545)*	*(6.085)*	*(11.636)*	*(5.848)*
Instruction time	-3.288	–	-0.769	7.376	4.473 °	-1.939	1.030
	(5.108)		*(1.358)*	*(5.474)*	*(2.367)*	*(1.608)*	*(1.794)*
Teacher characteristics							
Female teacher	0.867	-9.718 +	3.898	2.989	-9.153	8.819 °	5.556
	(9.028)	*(4.051)*	*(3.133)*	*(4.766)*	*(6.275)*	*(5.278)*	*(3.943)*
Teacher's experience (log)	-2.638	-0.387	-3.771 °	8.191 *	9.181 *	2.873	2.370
	(4.339)	*(3.212)*	*(1.974)*	*(2.589)*	*(3.223)*	*(2.979)*	*(2.290)*
Teacher's education							
Secondary only	–	–	–	12.496	–	–	59.804 *
				(9.451)			*(13.942)*
BA or equivalent	-10.856	–	46.182 *	16.233	-18.566 °	–	52.564 *
	(9.264)		*(6.495)*	*(10.260)*	*(10.569)*		*(14.666)*
MA/PhD	13.777	–	47.056 *	11.998	-7.008	9.954 °	53.272 *
	(21.598)		*(8.427)*	*(14.651)*	*(21.552)*	*(5.880)*	*(15.372)*
Family background controls	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6722	10271	5827	8285	11643	10973	6014
Schools [Unit of clustering]	86	151	150	137	147	183	134
R^2	0.203	0.063	0.182	0.278	0.159	0.203	0.259
R^2 (without imput. controls)	0.150	0.062	0.172	0.270	0.141	0.187	0.229

Significance levels (based on clustering-robust standard errors): * 1 percent. – + 5 percent. – ° 10 percent.

Table 3: The Coefficient on Log Class Size

Regressions within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Controlling for family-background variables and imputation controls. Clustering-robust standard errors in parentheses.

	HON	JAP	KOR	SIN	THA	USA	FRA
LS	107.924 *	126.077 *	-5.566	138.002 *	10.742	-3.294	60.824 *
	(30.775)	*(39.352)*	*(4.538)*	*(11.984)*	*(8.680)*	*(6.656)*	*(21.424)*
IV	261.893	151.598 *	66.028 +	155.356 *	-1926.856	-25.978	-13.591
	(160.843)	*(53.952)*	*(27.462)*	*(16.581)*	*(4666.453)*	*(25.666)*	*(32.477)*
SFE	96.727 *	-10.286	-13.245 +	89.849 *	4.899	-0.808	43.019 °
	(20.298)	*(15.222)*	*(5.547)*	*(15.366)*	*(5.999)*	*(7.903)*	*(21.838)*
SFE-IV	249.479	1.509	-46.547	11.093	-585.839	52.385	-81.209
	(752.850)	*(21.177)*	*(40.134)*	*(20.792)*	*(2075.300)*	*(42.658)*	*(53.996)*
Students	6712	10271	5827	8285	11610	10831	5669
Schools	85	151	150	137	146	179	119

Methods of estimation: LS = Least squares. – IV = Instrumental variables. – SFE = School fixed effects. – SFE-IV = Combination of school fixed effects and instrumental variables. – See text for details on the four methods of estimation.

Significance levels (based on clustering-robust standard errors): * 1 percent. – + 5 percent. – ° 10 percent.

Table 4: Institutions and Student Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	HON	JAP	KOR	SIN	THA	USA	FRA
School responsibility for	0.537	64.160 *	0.113	60.400 *	18.209	3.000	–
determining teacher salaries	*(13.839)*	*(13.400)*	*(3.771)*	*(9.993)*	*(11.780)*	*(8.805)*	
Homework	6.121	8.547 +	1.936	4.177 +	-1.952	14.265 *	4.582
	(4.408)	*(4.218)*	*(1.323)*	*(1.628)*	*(1.399)*	*(2.381)*	*(2.813)*
Teaching limited by							
Uninterested parents	-62.004 *	–	4.348	-10.258	4.247	-17.825 +	-17.097 +
	(17.329)		*(7.662)*	*(7.962)*	*(8.411)*	*(7.057)*	*(7.067)*
Interested parents	58.296 *	-46.875 *	-13.949	14.285	19.899	34.660 +	–
	(21.940)	*(18.010)*	*(9.189)*	*(16.046)*	*(12.724)*	*(16.754)*	
Family background controls	yes	yes	yes	yes	yes	yes	yes
Resource controls	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6722	10271	5827	8285	11643	10973	6014
Schools [Unit of clustering]	86	151	150	137	147	183	134
R^2	0.239	0.094	0.183	0.299	0.164	0.231	0.267
R^2 (without imput. controls)	0.180	0.090	0.172	0.292	0.149	0.207	0.235

Significance levels (based on clustering-robust standard errors): * 1 percent. – + 5 percent. – ° 10 percent.

Appendix Tables: Descriptive Statistics

(Only for reference to the referees and editors, not for publication in the journal.)

Table A1: Descriptive Statistics: Sample Size and Student Performance

Sample size: Absolute numbers. – Student performance: International test scores.
Standard deviation in parentheses. Standard deviation in percent of country mean test score in brackets.

	HON	JAP	KOR	SIN	THA	USA	FRA
Sample size							
Students	6752	10271	5827	8285	11643	10973	6014
Classes	171	302	300	274	293	529	253
Schools	86	151	150	137	147	183	134
Student performance							
Math score	575.8	588.3	592.3	622.3	508.3	487.8	514.4
Standard deviation	*(100.8)*	*(100.5)*	*(107.8)*	*(93.2)*	*(83.4)*	*(90.9)*	*(78.3)*
Standard deviation/score (in percent)	*[17.5]*	*[17.1]*	*[18.2]*	*[15.0]*	*[16.4]*	*[18.6]*	*[15.2]*
Science score	508.7	551.5	550.1	576.2	508.9	521.4	473.9
Standard deviation	*(88.7)*	*(90.4)*	*(93.9)*	*(102.7)*	*(72.6)*	*(106.2)*	*(78.9)*
Standard deviation/score (in percent)	*[17.4]*	*[16.4]*	*[17.1]*	*[17.8]*	*[14.3]*	*[20.4]*	*[16.7]*
Position in international ranking							
Math, 7th grade (out of 37 countries)	4	3	2	1	17	22	19
Math, 8th grade (out of 39 countries)	4	3	2	1	19	27	13
Science, 7th grade (out of 37 countries)	15	4	2	1	17	11	28
Science, 8th grade (out of 39 countries)	23	3	4	1	20	16	27

Table A2: Descriptive Statistics: Student and Family Background

Country means. Standard deviations in parentheses. – Only non-imputed data. Weighted by sampling probabilities.

	HON	JAP	KOR	SIN	THA	USA	FRA
Upper grade	0.500	0.512	0.504	0.502	0.492	0.502	0.487
	(0.500)	*(0.500)*	*(0.500)*	*(0.500)*	*(0.500)*	*(0.500)*	*(0.500)*
Age	13.688	13.902	13.710	13.939	13.884	13.735	13.805
	(0.884)	*(0.576)*	*(0.611)*	*(0.835)*	*(0.716)*	*(0.719)*	*(0.910)*
Sex (female)	0.449	0.483	0.438	0.492	0.594	0.498	0.496
	(0.497)	*(0.500)*	*(0.496)*	*(0.500)*	*(0.491)*	*(0.500)*	*(0.500)*
Born in country	0.870	–	0.991	0.920	0.989	0.926	–
	(0.337)		*(0.096)*	*(0.272)*	*(0.105)*	*(0.261)*	
Living with both parents	0.901	–	0.876	0.907	0.852	0.791	0.862
	(0.299)		*(0.330)*	*(0.290)*	*(0.355)*	*(0.406)*	*(0.345)*
Parents' education							
Primary	0.189	–	0.079	0.229	0.636	0.015	0.092
	(0.392)		*(0.269)*	*(0.420)*	*(0.481)*	*(0.122)*	*(0.289)*
Some secondary	0.394	–	0.178	0.000	0.113	0.059	0.246
	(0.489)		*(0.383)*	*(0.000)*	*(0.317)*	*(0.235)*	*(0.431)*
Finished secondary	0.280	–	0.414	0.565	0.114	0.192	0.334
	(0.449)		*(0.493)*	*(0.496)*	*(0.318)*	*(0.394)*	*(0.472)*
Some after secondary	0.053	–	0.090	0.134	0.027	0.375	0.145
	(0.224)		*(0.286)*	*(0.341)*	*(0.161)*	*(0.484)*	*(0.352)*
Finished university	0.084	–	0.238	0.072	0.111	0.359	0.183
	(0.278)		*(0.426)*	*(0.259)*	*(0.314)*	*(0.480)*	*(0.387)*
Books at home							
Less than one shelf (<=10)	0.208	–	0.088	0.108	0.187	0.081	0.054
	(0.406)		*(0.283)*	*(0.310)*	*(0.390)*	*(0.273)*	*(0.226)*
One shelf (11-25)	0.281	–	0.109	0.219	0.301	0.124	0.186
	(0.450)		*(0.312)*	*(0.413)*	*(0.459)*	*(0.330)*	*(0.389)*
One bookcase (26-100)	0.301	–	0.335	0.408	0.334	0.279	0.361
	(0.459)		*(0.472)*	*(0.491)*	*(0.472)*	*(0.449)*	*(0.480)*
Two bookcases (101-200)	0.103	–	0.240	0.145	0.093	0.209	0.196
	(0.304)		*(0.427)*	*(0.352)*	*(0.290)*	*(0.407)*	*(0.397)*
More than two bookcases	0.107	–	0.228	0.120	0.086	0.306	0.204
(>200)	*(0.309)*		*(0.420)*	*(0.325)*	*(0.280)*	*(0.461)*	*(0.403)*
Community location							
Geographically isolated	0.026	0.012	0.007	0.000	0.165	0.034	0.000
	(0.160)	*(0.107)*	*(0.081)*	*(0.000)*	*(0.371)*	*(0.180)*	*(0.000)*
Close to town center	0.679	0.382	0.540	0.392	0.234	0.442	0.391
	(0.467)	*(0.486)*	*(0.498)*	*(0.488)*	*(0.423)*	*(0.497)*	*(0.488)*

Table A3: Descriptive Statistics: Resources

Country means. Standard deviations in parentheses. – Only non-imputed data. Weighted by sampling probabilities.

	HON	JAP	KOR	SIN	THA	USA	FRA
Math class size	38.838	36.556	55.934	33.196	53.591	27.400	25.376
	(5.583)	*(4.026)*	*(24.807)*	*(7.074)*	*(28.312)*	*(15.637)*	*(3.277)*
Grade-average class size	40.136	36.302	49.893	32.515	42.804	25.624	25.357
	(3.687)	*(4.584)*	*(5.282)*	*(6.251)*	*(5.395)*	*(4.541)*	*(2.570)*
Shortage of materials							
None	0.629	0.521	0.367	0.733	0.115	0.456	0.385
	(0.483)	*(0.500)*	*(0.482)*	*(0.442)*	*(0.319)*	*(0.498)*	*(0.487)*
A lot	0.058	0.071	0.180	0.024	0.452	0.064	0.178
	(0.234)	*(0.256)*	*(0.384)*	*(0.153)*	*(0.498)*	*(0.245)*	*(0.383)*
Instruction time	8.625	–	9.247	8.366	9.947	7.683	7.039
(in 100 hours of 60 minutes per year)	*(1.615)*		*(1.829)*	*(0.512)*	*(1.544)*	*(2.228)*	*(1.506)*
Math teacher's sex (female)	0.386	0.248	0.496	0.599	0.690	0.688	0.484
	(0.487)	*(0.432)*	*(0.500)*	*(0.490)*	*(0.463)*	*(0.463)*	*(0.500)*
Math teacher's experience	9.124	13.273	12.095	17.540	9.739	15.076	19.784
(in years)	*(8.985)*	*(9.166)*	*(9.185)*	*(12.378)*	*(7.651)*	*(9.751)*	*(10.297)*
Math teacher's education							
Less than secondary	0.000	–	–	0.089	0.000	–	0.007
	(0.000)			*(0.285)*	*(0.000)*		*(0.086)*
Secondary only	0.354	–	0.003	0.350	0.053	–	0.338
	(0.478)		*(0.058)*	*(0.477)*	*(0.224)*		*(0.473)*
BA or equivalent	0.617	–	0.907	0.512	0.908	0.568	0.396
	(0.486)		*(0.291)*	*(0.500)*	*(0.289)*	*(0.495)*	*(0.489)*
MA/PhD	0.028	–	0.090	0.048	0.039	0.432	0.259
	(0.166)		*(0.286)*	*(0.215)*	*(0.194)*	*(0.495)*	*(0.438)*

Table A4: Descriptive Statistics: Institutional Features

Country means. Standard deviations in parentheses. – Only non-imputed data. Weighted by sampling probabilities.

	HON	JAP	KOR	SIN	THA	USA	FRA
School responsibility for	0.103	0.076	0.374	0.067	0.961	0.892	0.000
determining teacher salaries	*(0.305)*	*(0.265)*	*(0.484)*	*(0.249)*	*(0.194)*	*(0.310)*	*(0.000)*
Math							
Homework assignment	1.362	0.716	1.268	2.636	3.417	1.647	1.542
(in hours per week)	*(0.936)*	*(0.747)*	*(1.027)*	*(1.444)*	*(2.290)*	*(1.075)*	*(0.693)*
Teaching limited by							
Uninterested parents	0.100	–	0.071	0.090	0.142	0.149	0.128
	(0.299)		*(0.257)*	*(0.286)*	*(0.349)*	*(0.356)*	*(0.335)*
Interested parents	0.055	0.008	0.021	0.030	0.062	0.043	–
	(0.228)	*(0.089)*	*(0.145)*	*(0.170)*	*(0.241)*	*(0.204)*	

Table A5: Missing Values

Unweighted percentage of students with missing data.

	HON	JAP	KOR	SIN	THA	USA	FRA
Age	0.008	0.006	0.000	0.001	0.014	0.001	0.065
Sex	0.001	0.000	0.000	0.000	0.009	0.000	0.035
Born in country	0.028	1.000	0.019	0.006	0.011	0.018	1.000
Living with both parents	0.018	1.000	0.001	0.008	0.006	0.020	0.037
Parents' education	0.122	1.000	0.058	0.003	0.069	0.101	0.450
Books at home	0.020	1.000	0.003	0.007	0.017	0.023	0.044
Community location	0.113	0.006	0.013	0.000	0.116	0.150	0.107
Math class size	0.190	0.006	0.069	0.006	0.592	0.320	0.135
Science class size	0.225	0.010	0.113	0.026	0.620	0.574	0.174
Grade-average class size	0.111	0.000	0.009	0.003	0.136	0.205	0.122
Shortage of materials	0.111	0.016	0.020	0.016	0.123	0.152	0.097
Instruction time	0.218	1.000	0.064	0.000	0.162	0.320	0.396
Math teacher characteristics							
Teacher's sex	0.064	0.006	0.036	0.006	0.355	0.138	0.074
Teacher's experience	0.048	0.017	0.043	0.010	0.413	0.140	0.103
Teacher's education	0.070	1.000	0.036	0.020	0.359	0.142	0.092
School determines teacher salaries	0.123	0.020	0.020	0.007	0.308	0.172	0.123
Math							
Homework assignment	0.132	0.026	0.054	0.011	0.376	0.280	0.124
Uninterested parents limit	0.131	1.000	0.066	0.025	0.358	0.287	0.103
Interested parents limit	0.137	0.010	0.070	0.023	0.358	0.290	1.000