

Shrinking Classroom Age Variance Raises Student Achievement

Evidence from Developing Countries

Liang Choon Wang

The World Bank
Development Research Group
Human Development and Public Services Team
January 2011



Abstract

Large classroom variance of student age is prevalent in developing countries, where achievement tends to be low. This paper investigates whether increased classroom age variance adversely affects mathematics and science achievement. Using exogenous variation in the variance of student age in ability-mixing schools, the author finds robust negative effects of classroom age variance on

fourth graders' achievement in developing countries. A simulation demonstrates that re-grouping students by age in the sample can improve math and science test scores by roughly 0.1 standard deviations. According to past estimates for the United States, this effect size is similar to that of raising expenditures per student by 26 percent.

This paper is a product of the Human Development and Public Services Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at lwang12@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Shrinking Classroom Age Variance Raises Student Achievement
– Evidence from Developing Countries*

Liang Choon Wang⁺

JEL Codes: I20, O15.

* I thank Luis Benveniste, Michael Ewens, Deon Filmer, Ha Nguyen, and Adam Wagstaff for comments and suggestions. The findings, interpretations, and conclusions expressed in this paper are those of the author and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent.

⁺ Development Research Group's Human Development and Public Services Team, The World Bank, Mailstop MC 3-311, 1818 H St NW, Washington, DC 20433, USA; email: lwang12@worldbank.org.

1. Introduction

Developing countries often face shortages of infrastructure and teachers when striving for universal primary school attendance. It is fairly common to see students of diverse ages attending the same grade level in relatively poor countries (figure 1). Since students in poorer countries tend to have lower achievement (figures 2), could the larger variance of student age be one of the factors responsible for their relatively low achievement? As cognitive skills are linked to economic growth (Hanushek and Woessmann 2008), earnings (Murnane, Willett, and Levy 1995), and productivity (Bishop 1989), understanding how variance of student age affects achievement can be important for countries pursuing the millennium development goals.

A large variance of student age within the classroom may pose challenges to teachers in providing instruction appropriate to students with different academic readiness. These classrooms may also be prone to discipline and behavior problems as students with different mental maturities interact with each other. As a result, large classroom age variance may impede learning. Nonetheless, an age heterogeneous classroom may provide a venue for younger students to learn closely from older students and for older students to gain from helping and studying with younger peers. Hence, the effect of classroom age heterogeneity on achievement may be ambiguous.

The lack of resources and the non-enforcement or absence of compulsory schooling laws in developing countries often lead to students of diverse ages beginning formal schooling at the same time. Similarly, a successful promotion of universal primary education may also generate a sudden influx of relatively old first graders into schools. Identifying whether increased classroom age variance impedes student achievement will permit policy makers to respond with appropriate strategies to ameliorate its adverse effect, if it exists. For example, principals may consider

grouping students into classrooms on the basis of age and assigning teachers most qualified to teach the respective groups. If grouping students into different classrooms by age is not feasible, educators may form students into different age groups within the classroom and tailor instruction accordingly to minimize the adverse effect of classroom age heterogeneity. To the extent that test scores influence future earnings and economic growth, appropriate policy responses altering classroom age heterogeneity can have long-term economic consequences.

Past studies on the effects of age differences between students on their outcomes mostly focus on how they relate to school entry age policies in developed countries (Bedard and Dhuey 2006; Black, Devereux, and Salvanes forthcoming; Cascio and Schanzenbach 2007; Datar 2006; Elder and Lubotsky 2009). These studies argue that older students may outperform younger students in the same grade level because: (a) older students have accumulated more human capital prior to formal schooling as a result of their greater absolute age; and/or (b) the superior physical and mental capabilities of older students due to their relative age advantage reinforce their confidence over time and attract more school inputs at the expense of younger students. Findings by Elder and Lubotsky (2009), for example, show that absolute age differences explain the achievement gap better than relative age differences, implying that classroom age variance may have little negative effect on achievement. Although these results are important for evaluating school entry age policy in developed countries, they may not be informative for policy in developing countries where the variance of student age within grade level is significantly larger and the schooling decision is often complicated by the lack of resources and accessible schools.

A recent experimental study on the effect of ability grouping on achievement in Kenyan primary schools shows that reducing ability heterogeneity in classrooms generates an effective

teaching and learning environment that benefits all types of students (Duflo, Dupas, and Kremer forthcoming).¹ Given the positive correlation between age and achievement, ability grouping basically narrows the relative age differences within the classroom and Duflo et al.'s (forthcoming) results imply that decreased classroom age variance should also lead to achievement gain. Several factors may explain the different implications based on studies of the effects of school starting age versus the study on the effect of ability grouping on achievement. First, studies examining the effects of school starting age tend to focus on why school starting age matters for outcomes and whether it is worthwhile to delay school entry age, rather than to identify whether increased classroom variance of student age influences achievement. Second, differences in the education systems studied and identification strategies employed by previous studies may be responsible for the disparate findings. Third, samples drawn from places where variance of student ages within grade level is small due to the enforcement of compulsory attendance laws and entry age policy may not be suitable candidates for examining the effect of classroom age heterogeneity, as the small variation in student age across a relatively small number of classrooms or schools might yield imprecise estimates.

This paper uses exogenous variation in the classroom variance of student age in 14 developing countries to examine its effects on student achievement. To utilize variation in classroom age variance that is arguably exogenous, I employ a school fixed effects estimator and focus on the variation of student age within ability-mixing elementary schools sampled from two waves of the Trends in International Mathematics and Science Study (TIMSS). Because ability-

¹ Studies on the effects of ability grouping and tracking on student achievement using observational data from developed countries show mixed findings. Examples include Betts and Shkolnik (1999), Figlio and Page (2002), Manning and Pischke (2006), and Hanushek and Woessman (2006).

mixing schools do not assign students into classrooms on the basis of student ability, differences in student age across classrooms are likely orthogonal to other determinants of student achievement.² Nonetheless, it is difficult to rule out implicit sorting on the basis of age. To address potential selection bias, I simulate the average and standard deviation of classroom age that each student would experience if the school assigned students into classrooms on the basis of age. The variation in classroom age variables not explained by age sorting permits the implementation of an instrumental variable strategy. More importantly, the large number of classrooms sampled provides significant variation in classroom age variance to precisely identify its effect on student outcomes. Although this paper focuses on developing countries, the findings may also be applicable to schools in developed nations that have large classroom age variances due to the practice of combination classrooms or the implementation of grade promotion and retention policy.³

I find that greater classroom age variance leads to lower fourth graders' achievement in mathematics and science. For every one month increase in the classroom standard deviation of student age, average achievement falls by 0.03 standard deviations for both math and science. However, classroom average age has an insignificant negative effect on achievement. The negative effect of classroom age variance appears to (weakly) persist as the cohort of fourth

² An ability-mixing school is one in which the school principal claimed that students were not assigned into different classrooms based on their ability in mathematics and science.

³ This does not necessarily imply that the current results inform the effects of combination or multi-grade classrooms on student achievement, since teachers instructing these classrooms usually receive special training and use pedagogies different to those in traditional single-grade classrooms (see Benveniste and McEwan (2000) for a case study on multi-grade schools in Colombia). Studies examine the effects of multi-grade classrooms (e.g., see Sims [2008]) often face the difficulty associated with identifying causal relationship because multi-grade schools and students attending such schools likely differ in many aspects that are not easy to control for.

graders entered into the eighth grade. Similarly, there is weak evidence suggesting that boys and students above the median age are less affected by classroom age variance. On the other hand, increased classroom age variance is not associated with negative behavioral problems that students encounter in schools. The findings imply that the adverse effect of classroom age heterogeneity is likely restricted to academic achievement. Finally, a policy simulation demonstrates that by switching from age mixing to age sorting students, achievement in both mathematics and science can improve by roughly 0.1 standard deviations. In other words, age grouping students may help an average school achieve the benefit associated with increasing expenditures per student by roughly 26 percent according to Sander's (1999) estimate or that of cutting class size by 2.5 students based on Angrist and Lavy's (1999) estimate. Given the low administrative cost, age grouping seems like a cost effective method to raise average achievement.

2. Identification Strategy and Econometric Specifications

2.1 The Effect of Classroom Age Heterogeneity on Achievement

Differences in the variance of student age across countries and schools are likely correlated with other unobserved influences of achievement, such as income, the extent of urbanization, and educational expenditures. In contrast, differences in the variance of student age across classrooms within a school are more likely exogenous, if the school does not assign students and/or teachers into classrooms based on the students' prior achievement. Because students are essentially randomly assigned into classrooms in ability-mixing schools, it is unlikely that other

determinants of student achievement are systematically correlated with classroom age variance.⁴ As long as principals do not selectively assign teachers according to the classroom variance of student age, a school fixed effects estimator will yield a consistent estimate of the effect of classroom age variance on achievement.

The school fixed effects specification is:

$$y_{icjk} = \alpha_{jk} + \delta\sigma_{cjk} + \lambda_1 a_{icjk} + \lambda_2 a_{icjk}^2 + \phi \bar{a}_{cjk} + x'_{icjk} \beta + u_{icjk} \quad (1)$$

The dependent variable y_{icjk} is the achievement of student i in classroom c of school j in country k . α_{jk} is a set of school fixed effects, which ensures that I exploit the variation in age variance across classrooms within schools. σ_{cjk} is the standard deviation of student age (in years) for classroom c , which measures the extent of classroom age heterogeneity. The coefficient of interest, δ , is expected to be negative if classroom age variance impedes achievement. a_{icjk} is student i 's age measured in years. Because within-grade-level age range is large in developing countries, the achievement is expressed as a quadratic function of age to account for potential non-linearity of achievement in age. The average age of students in classroom c is \bar{a}_{cjk} . The coefficient ϕ captures the “social” age effect of the classroom on student i .⁵ Given the age dispersion of a classroom, if being present in a relatively “old” classroom hurts the student’s achievement, then ϕ is expected to be negative. x is a set of background characteristics and

⁴ For examples, see Kang (2007) and Wang (2010) that exploit the variation in peer quality across classrooms in ability-mixing schools to study the effects of peers on student outcomes.

⁵ This follows Manski’s (1993) definition of exogenous social effect, where the group’s average predetermined characteristics includes student i ’s own predetermined characteristics. The measure differs slightly from a typical peer effect study where the social effect excludes student i ’s own characteristics. Because own age is separately included in the regression equation, the current approach only alters the interpretation of the social effect.

teacher characteristics specific to student i . If classroom age variance is exogenously determined, the exclusion of x should have little effect on the estimate of δ .

2.2 Instrumental Variables

Although principals may claim to mix students of all ability types in classrooms, they may still implicitly sort students into classrooms by age or assign teachers of different quality based on the ex-post age distributions of classrooms. For example, a principal may assign a more competent teacher to teach a classroom that has slightly younger students or have more diverse age groups. The school fixed effects specification (1) will not adequately address this type of selection bias.

One way to correct for this form of selection bias is to exploit the variation in classroom age variables that are unrelated to age sorting. I simulate the hypothetical age distribution of a student's classroom under the assumption that students were sorted into classrooms by age.⁶ If age sorting is present, then the actual age distribution of a student's classroom and the simulated age distribution of a student's classroom and other observables are expected to be positively correlated. The variation in actual classroom average of age or standard deviation of age not explained by the simulated one and other observables is likely free of the effect of age sorting. Specifically, I will generate instrumental variables (IV) for classroom average age and classroom standard deviation of age using the regression residuals of the following regressions:

$$\bar{a}_{cjk} = \hat{\alpha}_{jk}^M + \hat{\gamma}^M \bar{a}_{cjk}^{Sim} + \kappa^M \sigma_{cjk} + \hat{\tau}_1^M a_{icjk} + \hat{\tau}_2^M a_{icjk}^2 + x'_{icjk} \hat{\psi}^M + \hat{e}_{icjk}^M \quad (2)$$

$$\sigma_{cjk} = \hat{\alpha}_{jk}^{SD} + \hat{\gamma}^{SD} \sigma_{cjk}^{Sim} + \kappa^{SD} \bar{a}_{cjk} + \hat{\tau}_1^{SD} a_{icjk} + \hat{\tau}_2^{SD} a_{icjk}^2 + x'_{icjk} \hat{\psi}^{SD} + \hat{e}_{icjk}^{SD} \quad (3)$$

⁶ This is done by ranking students in the sample by age and then assigning them into classrooms in the sample.

σ_{cjk}^{Sim} is the simulated classroom standard deviation of age and \bar{a}_{cjk}^{Sim} is the simulated classroom average age when students are perfectly sorted into classrooms by age. The coefficients $\hat{\gamma}^M$ and $\hat{\gamma}^{SD}$ capture the relationship between age sorting and classroom average age and classroom standard deviation of age, respectively. They indicate whether estimates of δ and ϕ in equation (1) may suffer from selection bias. By construction, the residuals \hat{e}_{icjk}^M and \hat{e}_{icjk}^{SD} obtained from the regressions are orthogonal to age sorting and other observables, and can be used as instrumental variables for \bar{a}_{cjk} and σ_{cjk} in equation (1).

This instrumental variable strategy effectively exploits the variation in age distributions across classrooms for students who are not perfectly matched to their classmates and teachers on the basis of their age and other observables. The instrumental variables will be highly relevant if the simulated age distribution and observed characteristics of students and teachers do not have much explanatory power. This will be the case if the claim of ability-mixing corresponds to the random assignment of students of different age groups and teachers of different quality into classrooms. The validity of these instruments rests on the assumption that the extent of non-random selection and other threats to identification are fully captured by the simulated age sorting distributions and observables.

2.3 Behavior Outcomes and Attitudes

One of the concerns against early school entry or mixing students of different ages relates to the possibility that an age heterogeneous classroom increases the chances of younger students being bullied, teased, or left out by older students. These behavioral issues may harm students' self esteem, which in turn affects their learning outcomes. To assess whether increased classroom

variance of student age may lead to increased behavioral problems, I replace the dependent variable in regression equation (1) with a set of variables measuring whether students experienced behavioral problems inflicted by others and whether they find school enjoyable. I also estimate the effects of classroom age heterogeneity on these measures of behaviors and attitudes separately for students who are younger than the median age of other fourth graders in their countries.

2.4 Differential Effects of Classroom Age Heterogeneity on Achievement

Classroom age heterogeneity may have differential effects on students depending on their age and gender. For example, young students may require more teacher attention and if teachers tailor instruction to the median or average students, a diverse classroom may have a stronger negative effect on their achievement than on older students' achievement. Similarly, the effect of age heterogeneity on achievement may differ depending on gender. For instance, boys are perhaps more likely to be distracted than are girls in heterogeneous classroom. It is also possible that girls are more vulnerable to age heterogeneity. To examine whether there exist differential effects, I estimate equation (1) separately by student age group and gender.

3. Data

The data used are sourced from the Trends in International Mathematics and Science Study (TIMSS) in 2003 and 2007. TIMSS provides student-level data on mathematics and science

achievement of fourth graders and eighth graders in a large number of countries.⁷ In addition to internationally comparable standardized test scores, TIMSS also collected student surveys, teacher surveys, and school surveys.

Because TIMSS asked principals whether they grouped students into classrooms on the basis of ability in mathematics and science, and sampled at least two classrooms from numerous schools in several countries, I am able to exploit classroom level variance of student age within each ability-mixing school through a school fixed effects estimator.⁸ The focus on ability-mixing schools is important, as principals in these schools are less likely to selectively assign teachers according to the age distribution of students or the prior achievement of students across classrooms. Similarly, as age and achievement are positive correlated, the classroom age distributions of ability-mixing school are also less likely correlated with students' prior achievement and other determinants of achievement. Since eighth graders tend to attend various mathematics and science classes with different levels of difficulties and with different set of peers, even within schools claiming not to group students based on ability, classroom age heterogeneity measured in the eighth grade is more likely confounded with unobserved factors and measurement error. Consequently, I focus primarily on fourth graders and only examine eighth graders to assess whether the effect persists into the eighth grade. I include student data

⁷ However, countries were not consistently covered across different waves of TIMSS or across grade levels within each wave of TIMSS. Furthermore, each student was only tested once and individual schools and students were not followed over time in TIMSS, limiting the use of various estimation techniques.

⁸ An ability-mixing school is one in which students of different ability levels are mixed in a classroom. I only include schools that do not ability group students in math and science classes, based on principals' responses to the survey.

from 14 countries classified as low and middle income countries by the World Bank in 2007 and estimate the models using pooled data from TIMSS 2003 and 2007 in most of the analysis.⁹

Table 1 reports descriptive statistics of the variables used in this study. Since the focus of this study is on classroom age variance, it is crucial that there is a considerable amount of variation in the classroom standard deviation of student age and classroom age distributions are fairly symmetric on average. Indeed, the classroom standard deviation of age has a standard deviation of 0.21 years (or 2.5 months) and a range of 1.7 years. Furthermore, the average classroom age skewness is only 0.24, indicating that the extent of asymmetry in classroom age distributions is reasonably low. Hence, using classroom standard deviation of age as the measure of age heterogeneity appears sensible. Nevertheless, alternative measures of age dispersion are also considered to assess whether the estimates are sensitive.

Table 2 verifies the claim that the classroom variance of student age is orthogonal to other influences of achievement and that teachers are not systematically assigned to students depending on classroom age heterogeneity. It reports the regression estimates of a set of student background characteristics and teacher characteristics as the dependent variable against the classroom standard deviation of student age and classroom average of student age, after controlling for school fixed effects, own age, and own age squared. If classroom age variance is exogenously determined, it should not be correlated with student background characteristics and teacher characteristics. Except in one instance where parental nativity status is significant at the 10% level, all other predetermined student and teacher characteristics are not significantly

⁹ These countries are selected because of their development status and their samples of multiple classrooms per school. Four of these countries are in TIMSS 2003 and thirteen in TIMSS 2007. Three of them classified as low or middle income countries appear in both waves of TIMSS. I also estimate the models using TIMSS 2003 and TIMSS 2007 data separately. The results are presented in a robustness check section.

correlated with classroom standard deviation of student age. Thus, I am quite confident with the identification strategy used to estimate the effects of classroom age variance.

However, Table 2 shows that classroom average age is significantly correlated with a few observables at the 5% or 10% level. In particular, it appears that students in classrooms with higher average age also tend to have less qualified teachers. This means that there may be some extent of age sorting and non-random assignment of teachers, which may bias the school fixed effects estimates. Thus, I will need to rely on the instrumental variable estimator to isolate the potential of selection bias and to make causal inferences on the estimated effects of classroom average age.

Table 3 presents evidence that there exists some form of age sorting. Column (1) shows that simulated classroom average age (under age sorting) is positively correlated with the actual classroom average age. This means that the school fixed effects estimate of classroom average age effect will likely suffer from selection bias and highlights the need to implement IV estimation. In contrast, column (2) shows that the actual classroom standard deviation of age is fairly exogenous to selection bias, as it is not significantly correlated with the simulated classroom standard deviation of age. Hence, I must rely on IV estimates to make causal interpretation of the estimated effects of classroom average of age in the following section.

4. Empirical Results and Discussion

4.1 Classroom Age Heterogeneity and Achievement

The regression estimates for math achievement based on equation (1) and its variants is presented in Table 4. Table 5 reports the estimates for science achievement.

Table 4 shows that classroom standard deviation of age has an adverse effect on mathematics achievement. The estimated effect is significantly negative in all specifications. Comparing to the simple Ordinary Least Squares (OLS) specification, the country and school fixed effects specifications tend to show a smaller negative effect of classroom age heterogeneity, highlighting the bias inherent in a simple cross-country or cross-school analysis. The school fixed effects specifications without (column 3) and with (column 4) student and teacher characteristics yield similar estimates of the effect of classroom age heterogeneity, supporting the claim that classroom age variance within a school is exogenous.¹⁰ Since the variation in classroom age variance is fairly exogenous, the instrumental variable (IV) estimate is similar to the school fixed effects estimate. The preferred IV specification (5) shows that for every one month increase in the classroom standard deviation of age, average math achievement is expected to fall by 0.03 standard deviations.¹¹

Table 4 also shows that the estimated effect of classroom average age on achievement is mostly negative, which means that being placed in an older classroom hurts a student's achievement. However, the estimated effect is insignificant in the school fixed-effects specification. The IV estimate shows that the correction for potential selection bias increases the size of the negative effect of classroom average age, but the estimated effect remains statistically insignificant. Note that as the instrumental variables have very high partial F statistics, the estimates are unlikely to suffer from a weak instrumental variable problem.

¹⁰ Although the coefficients of student and teacher characteristics are not reported, student and teacher characteristics are jointly significant in explaining achievement.

¹¹ These numbers measured in month of age are obtained by dividing the coefficient estimates by 12.

Table 5 presents estimates for science achievement. Similar to the effect of classroom standard deviation of age on math achievement, the effect on science achievement is also significantly negative across various specifications. The school fixed effects specifications and the preferred IV specification (5) yield similar point estimates of the effect of classroom age heterogeneity. The preferred IV estimate indicates that for every one month increase in the classroom standard deviation of age, average science achievement is expected to fall by 0.03 standard deviations. Similarly, the effect of classroom average age is estimated to be negative, but statistically insignificant.

Because the standard deviation of classroom standard deviation of student age is 0.21 years (or 2.5 months), the estimated effect size of a one standard deviation increase in classroom age heterogeneity is roughly -0.075 standard deviations for mathematics and -0.081 standard deviations for science. To gauge how large these effect sizes are, it is helpful to use past estimates on the effects of class size reduction and increased school expenditures on achievement to make a simple comparison (even though some of these estimates were debated).¹² For example, Sander's (1999) instrumental variable estimate shows that for every one dollar increase in the spending per student, math achievement in Illinois is predicted to increase by 0.0034 points. Converting this effect size to standard deviation of change in test score with respect to percentage change in expenditures, the current estimates are roughly equivalent to an increase in expenditures per student of 23 percent. Similarly, comparing to Angrist and Lavy's (1999) largest instrumental variable estimate of the effect of class size reduction on math achievement in

¹² For examples, see Hanushek (1995, 1997), Krueger (2003), and Woessmann (2000) for the debates on the effectiveness of school resources on student achievement.

Israel, the effect of a one standard deviation decrease in classroom standard deviation of student age is almost as large as cutting class size by 2 students.

In sum, the results show fairly robust and large adverse effect of classroom age variance on student achievement. In contrast, the effect of classroom average age is negative but statistically insignificant, implying that grouping students by age will not significantly benefit younger students at the expense of older students. The results imply that grouping students by age can significantly improve test scores without redistributing (much) achievement gain from older students to younger students.

4.2 Effects on Behavior and Attitude toward Schooling

Having classmates of various ages may increase the incidence of students, especially young ones, being bullied and shunned by classmates, as well as make schooling experience less enjoyable.

Columns (1) to (3) of Table 6 report the estimated effect of classroom age heterogeneity on the likelihood of an average student reported being bullied, left out of activities, and not liking school, respectively. Columns (4) to (6) report the estimates for the sample of students at the median age or younger.¹³ The top panel presents school fixed effects estimates, and the bottom panel presents IV estimates. As the preferred IV estimates are statistically insignificant, there is little evidence suggesting that greater classroom age heterogeneity increases the chances that students reported being bullied, left out of activities in school, or not liking school. Together with the estimates reported in the previous section, the results imply that the negative effect of

¹³ Median age is defined in accordance with the grade-level age distribution of the fourth grader's school. The estimates are not sensitive to using the grade-level age distribution of the fourth grader's country.

classroom age variance is more likely academic specific. Nonetheless, these findings should be interpreted with caution because the surveys asked students about their experience in school, but not in class, and it is possible that school level measures are noisily related to classroom level measures.

4.3 Who Loses More? Effects by Age and Gender

One concern for mixing students of large age differences in the same classroom is in its potential adverse effect on younger students. Table 7 reports the differential effects of classroom age heterogeneity on achievement of students who are above the median age and students who are at the median age or below. The point estimates reveal that younger students tend to be more affected by greater classroom age variance. The differences between old and young students are larger in science than in math, and the school fixed effects estimates and IV estimates are similar. For a one month increase in the classroom age standard deviation, the differential effect on the change in math achievement between young and old students is at most 0.004 standard deviations. Even though the size of the differential effect is minute, the overall pattern is consistent with the view that younger students are disadvantaged more than older students in age-diverse classrooms. Similarly, the estimated effects of classroom average age show that young students are more hurt when placed in relatively older classrooms, even though both the school fixed effects and IV estimates are statistically insignificant.

Table 8 presents estimates for boys and girls separately. The coefficient estimates of classroom standard deviation of age are more negative for girls than for boys, especially in science. However, similar to the differential effects by age, the differential effects by gender are also small in magnitude. For a one month increase in the classroom age standard deviation, the

differential effect on the change in math achievement between boys and girls is at most 0.003 standard deviations. The estimated effects of classroom average age on achievement show that boys appear to be more disadvantaged by being placed with relatively old classmates, but the estimates are statistically insignificant.

To summarize, the estimates presented in this section show weak evidence that younger students and girls are more disadvantaged by increased classroom age variance.

4.4 Robustness Checks

4.4.1 Sensitivity to Functional Form of Age

In the analysis presented above, all regressions included age and age squared as explanatory variables. Table 9 presents estimates from regressions using different order of age polynomial as regressors. The estimated effects of classroom standard deviation of age and classroom average age are fairly insensitive to different functional form assumption of the relationship between achievement and student's age. The effect of age on achievement is negative in the linear specification, but increasing at a decreasing rate in the quadratic specification. Since the average age in the sample is roughly 10.7 years, the quadratic specification is more consistent with previous literature on the positive relationship between achievement and age. The cubic specification yields estimated effect of classroom age heterogeneity similar to the quadratic specification. Hence, the quadratic functional form chosen for the main analysis appears reasonable.

4.4.2 *Alternative Measures of Classroom Age Dispersion*

The measure of classroom age heterogeneity used so far has been the standard deviation of student age. Table 10 shows that the adverse effect of classroom age heterogeneity is robust to alternative measures of dispersion. Columns (1) and (3) present estimates based on the difference between the 75th and 25th percentile of the classroom age distribution and columns (2) and (4) report estimates based on age range (i.e., maximum – minimum). Since the standard deviation of the 75th-25th percentile age difference is 0.256 (Table 1), the effect size is roughly 0.07 standard deviations for both mathematics and science. Similarly, as the standard deviation of age range is 1.092, the effect size is also roughly 0.06 standard deviations. These estimated effect sizes are similar to the 0.08 using the standard deviation of age as the measure of classroom age dispersion.

4.4.3 *Estimating Using TIMSS 2003 and TIMSS 2007 Separately*

The estimates presented above are all based on pooled TIMSS data. For pooled data to be sensible, the point estimates should be similar across both waves of TIMSS. Table 11 shows that the estimated effects of classroom age heterogeneity remain significantly negative despite the reduction in sample size. Furthermore, even though only three countries overlap in TIMSS 2003 and TIMSS 2007, the estimated effects of classroom age heterogeneity are not too different between the two waves of TIMSS. Thus, the estimates are not sensitive to pooling both waves of TIMSS.

4.5 Does the Effect Persist in the Eighth Grade?

Estimates presented in section 4.3 shows weak evidence that older students are less adversely affected by classroom age variance than younger students in the fourth grade. However, does classroom age variance continue to impede achievement as students enter higher grade levels? Since the cohort of fourth graders in TIMSS 2003 attended grade eight in 2007, it is possible to examine whether eighth-grade classroom age variance continues to affect their achievement. However, because individuals were not followed over time in TIMSS, I can only compare the performance of the same grade cohort in countries sampled in both waves of TIMSS, leaving us a sample from three countries.

There are a number of limitations when the same grade cohort is compared in the two waves of TIMSS. First, the same set of students was not followed over time and sampling differences across the two waves of TIMSS make comparison less reliable. Second, in addition to sampling variation, individuals most negatively affected by classroom age heterogeneity might repeat grade or drop out of school, and hence not observed in the eighth grade sample, leading to potential selection bias against the finding of a negative effect. Third, students who attended ability mixing schools in the fourth grade may switch to ability grouping schools in the eighth grade, making the comparison less meaningful. Fourth, the structure of eighth grade courses may also introduce error in the measure of classroom age heterogeneity, as eighth graders are more likely to take different courses that vary in difficulties with different set of peers. Despite these shortcomings, comparing the effects of classroom age variance on achievement for the same cohort of students within ability-mixing schools is the only option to gauge whether classroom age variance persists to impede achievement as students advance to higher grade levels.

Table 12 compares the estimates for fourth graders and eighth graders using data from Armenia, Latvia, and Lithuania. First, note that although the smaller sample size greatly reduces the statistical significance of the estimates, the estimated effect of classroom standard deviation of age on fourth graders' math achievement remains statistically significant at the 10% level (column 1) and is similar in magnitude to that using the full sample. Column (3) and column (4) show that the negative effect is much smaller in the eighth grade than in the fourth grade. Specifically, the negative effect is not statistically significant for science and only significant at the 15% level for mathematics. Based on the IV estimates, the reduction in the effect size is roughly 18% for mathematics and 38% for science. Given the many caveats highlighted above and the smaller sample size, it appears that the negative effect of classroom age heterogeneity on achievement, especially for mathematics, does persist to some extent.

4.6 Policy Simulation: Age Grouping and Achievement Gain

Given the significant negative effects of classroom age variance on achievement, school principals may improve student achievement by reducing the extent of age variance within the classroom through grouping students by age. The question is how much achievement gains are feasible for all ability-mixing schools to switch to age grouping, given its age distribution and the point estimates presented above? Does age grouping lead to greater or less inequality between students of different ages?

4.6.1 Mean Effects of Re-assignment

The preferred point estimates for mathematics and science reported respectively in specification (5) of Table 4 and Table 5 can be used to simulate the achievement gains attainable by

reassigning students into classrooms on the basis of their age.¹⁴ First, I construct the classroom standard deviation of age and the classroom average of age for each classroom by age grouping students. Note that by re-grouping students, classroom standard deviation of age shrinks for all students, but students assigned into a younger classroom will have a lower classroom average age than those assigned into an older classroom. The reduction in classroom standard deviation of age leads to achievement gain, but regrouping lowers the achievement of older students, and increases that of younger students, because the coefficient of classroom average of student age is negative (even though it is statistically insignificant). Second, the differences in the classroom standard deviation of age and classroom average of age between the original age mixing scenario and the new age grouping scenario are then multiplied by the respective point estimates to derive the net achievement gains for mathematics and science. Finally, averages by country are reported in Table 13.

Column (1) of Table 13 shows the standard deviation of age at the grade level for each country. Column (2) reports the average differences of classroom standard deviation between age grouping classrooms and age mixing classrooms. Column (3) reports the average differences of classroom average age between age grouping classrooms and age mixing classrooms. Columns (4) and (5) present the average predicted achievement gain for mathematics and science, respectively. The bottom row reports the averages for the sample of countries.

Table 13 shows that countries with the largest within-grade standard deviation of age also tend to realize the greatest reduction in classroom standard deviation of age if schools were to

¹⁴ Estimates reported in Tables 4 and 5 are used because the differential effects of classroom age heterogeneity (between boys and girls or between young and old students) are small in magnitudes. Simulation is conducted using TIMSS 2007 sample.

switch from age mixing to age grouping. The reduction in classroom standard deviation of age ranges from 0.14 years in Russia to 0.63 years in Morocco. The corresponding achievement gain in mathematics ranges from 0.05 standard deviations in Russia to 0.21 standard deviations in Morocco. The achievement gain is slightly greater for science – 0.06 and 0.23 standard deviations in Russia and Morocco, respectively. Overall, the average reduction in classroom age standard deviation is 0.26 years, and the average achievement gain is 0.09 standard deviations for mathematics and 0.10 standard deviations for science. These gains from reassignment are roughly equivalent to raising expenditures per student by 26 percent in accordance with Sander's (1999) estimates. Similarly, using Angrist and Lavy's (1999) estimated effect of class size reduction on achievement as a comparison, these predictions suggest that regrouping students can bring about an effect equivalent to cutting average class size by approximately 2.5 students.

Figure 3 and figure 4 plot the predicted achievement gain in mathematics against national incomes and achievement, respectively. Figure 3 illustrates that poorer countries tend to gain the most by switching from age mixing to age grouping students. In particular, countries that have lower average achievement, such as El Salvador and Morocco, are also the ones that will benefit the most through age grouping students (figure 4). Given that grouping students by age involves little administrative cost, it is an attractive option to raise achievement, especially for countries with large age variance, low achievement, and low incomes.

4.6.2 Distributional Effects of Re-assignment

A policy change may be difficult to justify if some students will be significantly disadvantaged by the change. Since re-assignment will lower the classroom average age for relatively young students and raise it for relatively old students, the former will gain while the latter will lose

through the (insignificant) negative effect of classroom average age. Similarly, because classroom age heterogeneity has a slightly more negative effect on students below the median age, age grouping may benefit them more. Although the gain from reduced classroom age heterogeneity is likely greater than any loss from having older classmates for students above the median age, it can be useful to compare achievement gains of the two groups of students to evaluate the distributional effect of re-assignment. I simulate the achievement gain for students above the median age and for students at the median age or younger based on estimates reported in Table 8.

Table 14 summarizes the simulation results by student age group and country. Except Tunisia, where re-assignment lowers math achievement of students above the median age by 0.006 standard deviations, all other countries experience achievement gains in math and science for all students through re-assignment. Specifically, the gains are greater for students at the median age or below than for students above the median age. Therefore, the simulation shows that age grouping not only improves average achievement, but also reduces achievement differences between older and younger students.

5. Conclusions

This paper presents evidence that increased classroom age variance is detrimental to student achievement in mathematics and science. Using arguably exogenous variation in classroom variance of age within ability-mixing schools in 14 developing countries, I show that a one-month increase in the classroom standard deviation of student age will lead to approximately a 0.03 standard deviation reduction in fourth graders' math and science achievement. However, the effect of classroom average age is statistically insignificant. There is weak evidence suggesting

that younger students and girls are more negatively affected by increased classroom age heterogeneity. Classroom age variance also appears to impede student achievement as they progress into the eighth grade. Although classroom age variance hurts academic achievement, it does not significantly increase the incidence of behavioral problems or make schooling experience less enjoyable for students.

The robust negative effect of classroom age variance and the insignificant effect of classroom average age on student achievement suggest that grouping students by age can lead to test score improvements. A simulation shows that by switching from age mixing students into age grouping students, schools can reduce classroom standard deviation of age by 0.26 years on average. The corresponding average achievement gain is 0.09 standard deviations in math score and 0.10 standard deviations in science score. According to Sander's (1999) estimates using U.S data, such effect sizes are similar to that of raising expenditures per student by 26 percent. Furthermore, gains are experienced by students of all age groups, but more so for students at the median age or younger, leading to smaller achievement differences between older and younger students. Countries that have larger within-grade-level variance of student age and lower average achievement are the ones that tend to gain the most from age grouping.

Given the low administrative cost, age grouping shows promise as a method to improve learning outcomes. Nevertheless, since the estimates are based on observational data and it is not possible to completely rule out the presence of unobserved influences which are correlated with classroom age distribution, readers should be cautious in attaching a causal interpretation to the estimates. It will certainly improve the confidence in recommending age grouping to policy makers in countries with large classroom age variances if randomized controlled experiments can be conducted to ascertain whether findings reported here stand up to scrutiny.

References

- Angrist, Joshua D. and Victor Lavy. (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, vol.114 (2), pp. 533-575.
- Bedard, Kelly and Elizabeth Dhuey. (2006). "The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects." *Quarterly Journal of Economics*, vol.121, pp. 1437-1472.
- Benveniste, Luis A. and Patrick J. McEwan. (2000). "Constraints to Implementing Educational Innovations: The Case of Multigrade Schools," *International Review of Education*, vol.46 (1-2), pp. 31-48.
- Betts, Julian and Jaime Shkolnik. (1999). "Key Difficulties in Identifying the Effects of Ability Grouping on Student Achievement." *Economics of Education Review*, vol.19 (1), pp. 243-266.
- Bishop, John. (1989). "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review*, vol.79 (1), pp. 178-197.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. (forthcoming). "Too Young to Leave the Nest? The Effects of School Starting Age." *Review of Economics and Statistics*.
- Cascio, Elizabeth U. and Diane Schanzenbach. (2007). "First in the Class? Age and the Education Production Function." NBER Working Paper No. 13663.
- Datar, Ashlesha. (2006). "Does Delaying Kindergarten Entrance Give Children a Head Start?" *Economics of Education Review*, vol.25, pp. 43-62.
- Duflo, Esther, Pascaline Dupas and Michael Kremer. (forthcoming). "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*.
- Elder, Todd E. and Darren H. Lubotsky. (2009). "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources*, vol.44 (3), pp. 641-683.
- Figlio, David and Marianne Page. (2002). "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?" *Journal of Urban Economics*, vol.51, pp. 497-514.
- Hanushek, Eric A. (1995). "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer*, vol.10 (2), pp. 227-246.

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

Hanushek, Eric A. and Ludger Woessmann. (2008). "Does Educational Tracking Affect Performance and Inequality? Differences-in-differences Evidence Across Countries." *Economic Journal*, vol.116 (510), pp. C63-C76.

Hanushek, Eric A. and Ludger Woessmann. (2008). "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature*, vol.46 (3), pp. 607-668.

Kang, Changhui. (2007). "Classroom Peer Effects and Academic Achievement: Quasi-randomization Evidence from South Korea." *Journal of Urban Economics*, vol.61, pp. 458-495.

Krueger, Alan. (2003). "Economic Considerations and Class Size." *Economic Journal*, vol. 113, F34-F63.

Manski, Charles. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, vol.60 (3), pp. 531-542.

Manning, Allen and Jorn-Steffen Pischke. (2006). "Comprehensive Versus Selective Schooling in England & Wales: What Do We Know?" Centre for the Economics of Education (LSE) Working Paper No. CEEDP006.

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

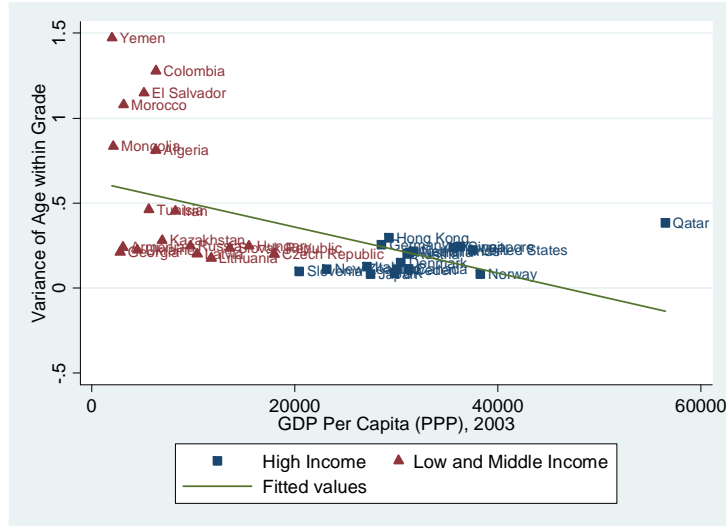
Sander, William. (1999). "Endogenous Expenditures and Student Achievement." *Economics Letters*, vol.64 (2), pp. 223-231.

Sims, David. (2008). "A Strategic Response to Class Size Reduction: Combination Classes and Student Achievement in California." *Journal of Policy Analysis and Management*, vol.27, pp. 457-478.

Wang, Liang Choon. (2010). *Three Essays in Labor Economics*. Ph.D. Dissertation, University of California, San Diego.

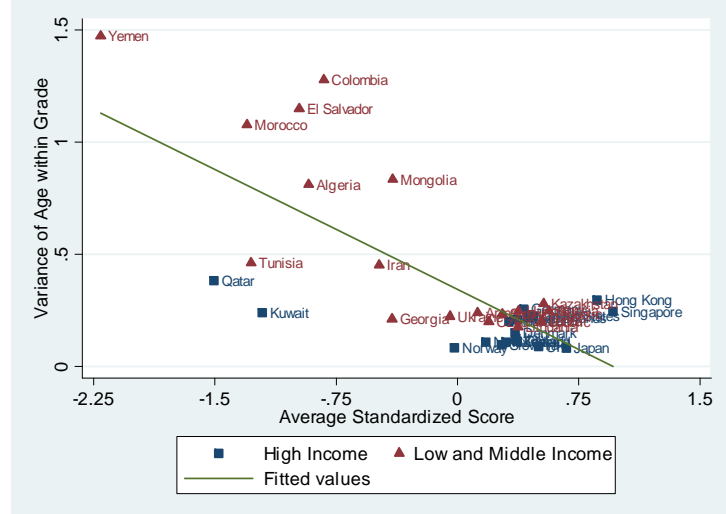
Woessmann, Ludger. (2001) "New Evidence on the Missing Resource-Performance Link in Education." Kiel Working Paper No. 1051, Kiel Institute of World Economics.

Figure 1: Variance of Fourth Graders' Age against GDP per Capita (PPP)



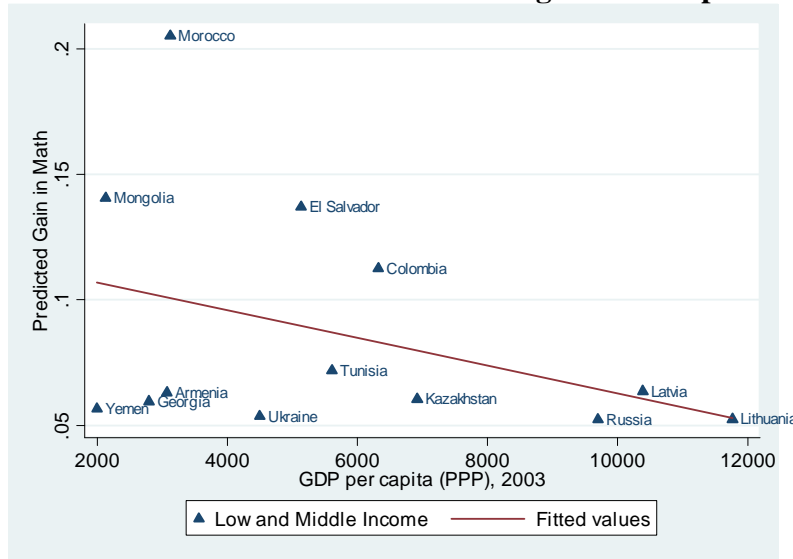
Notes: Data sourced from Third International Mathematics and Science Study (TIMSS) 2007 and the World Development Indicators. 37 economies are included in the sample. The variance of age for United Kingdom is calculated based on the weighted average of the figures of England and Scotland. The variance of age for Canada is calculated by the weighted average of Alberta, British Columbia, Quebec, and Toronto provinces. GDP per capita (PPP) in 2003 is selected so that income matches the time that the fourth grade cohort in TIMSS 2007 commenced primary education.

Figure 2: Variance of Age against Average Test Score of Fourth Graders



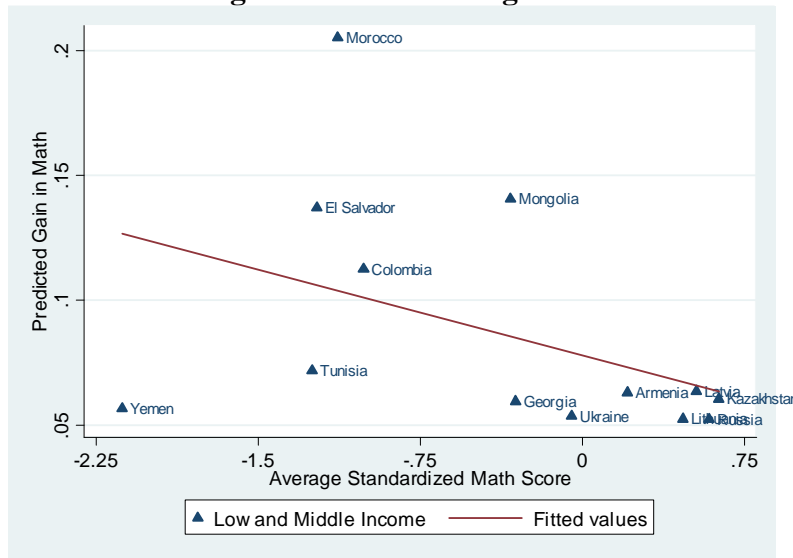
Notes: Data sourced from TIMSS 2007 and the World Development Indicators. 37 economies are included in the sample. The test score for United Kingdom is calculated based on the weighted average of the figures of England and Scotland. The test score for Canada is calculated by the weighted average of Alberta, British Columbia, Quebec, and Toronto provinces. GDP per capita (PPP) in 2003 is selected so that income matches the time that the fourth grade cohort in TIMSS 2007 commenced primary education. Average standardized score is the average of the standardized international scale mathematics and science scores.

Figure 3: Predicted Gain in Math Achievement against GDP per Capita (PPP)



Notes: Author's own calculation using TIMSS 2007 and the World Development Indicators. 13 economies are included in the sample. GDP per capita (PPP) in 2003 is selected so that income matches the time that the fourth grade cohort in TIMSS 2007 commenced primary education.

Figure 4: Predicted Gain against Actual Average Achievement in Mathematics



Notes: Author's own calculation using TIMSS 2007. 13 economies are included in the sample. Average standardized math score is the actual scale math score standardized internationally.

Table 1: Descriptive Statistics

<i>Variables</i>	<i>Obs.</i>	<i>Weighted Mean</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
<u>Student Characteristics</u>						
Math	22841	0.101	0.052	0.873	-3.88	2.56
Science	22841	0.013	-0.015	0.827	-3.82	2.57
Classroom Age SD (years)	22841	0.460	0.470	0.210	0.06	1.74
Classroom Age 75 th -25 th Percentile (years)	22841	0.582	0.585	0.256	0.08	3.25
Classroom Age Range (years)	22841	1.805	1.906	1.092	0.08	7.08
Classroom Age Skewness (years)	22841	0.236	0.241	0.793	-3.06	3.11
Classroom Ave. Age (years)	22841	10.68	10.64	0.397	9.63	12.36
Age (years)	22841	10.68	10.64	0.641	6.17	15.00
Bullied	22841	0.405	0.404	0.491	0	1
Left Out	22841	0.164	0.164	0.370	0	1
Like School	22841	0.841	0.852	0.355	0	1
Native Born	22841	0.833	0.834	0.372	0	1
Parents Native Born	22841	0.884	0.876	0.329	0	1
Speak National Language	22841	0.853	0.857	0.350	0	1
Boy	22841	0.504	0.506	0.500	0	1
Books	22841	0.797	0.802	0.399	0	1
Calculator	22841	0.793	0.795	0.404	0	1
Computer	22841	0.511	0.510	0.500	0	1
Study Desk	22841	0.784	0.784	0.412	0	1
Dictionary	22841	0.778	0.792	0.406	0	1
<u>Teacher Characteristics</u>						
Math Teaching Experience (years)	22841	20.00	20.26	11.02	0	50
Math Teacher Certificate	22841	0.686	0.667	0.471	0	1
Major in Math	22841	0.347	0.340	0.474	0	1
Male Math	22841	0.196	0.066	0.248	0	1
Science Teaching Experience (years)	22841	18.56	18.77	11.54	0	50
Science Teacher Certificate	22841	0.652	0.637	0.481	0	1
Major in Science	22841	0.241	0.248	0.432	0	1
Male Science	22841	0.194	0.048	0.214	0	1

Notes: Author's own calculated based on data sourced from TIMSS 2003 and TIMSS 2007. The weighted means are computed based on TIMSS sampling weights. Only observations with achievement and age available are included. Mathematics and Science scores reported are the international scale scores standardized to a standard normal distribution. The mean test scores reported above do not have zero means because only the subset of ability mixing schools is included. 538 schools are in the sample. See data appendix for variable construction.

Table 2: Verification of Exogenous Variation in Classroom Age Heterogeneity

	<i>Classroom Age</i>			<i>Classroom Age</i>	
	S.D.	Average		S.D.	Average
<u>Student characteristics</u>			<u>Ave. Math Teacher</u>		
Native born	0.001 (0.038)	0.002 (0.039)	Teaching experience	0.008 (2.995)	-4.432** (2.202)
Parents native born	-0.058* (0.031)	-0.014 (0.026)	Teacher certificate	0.136 (0.107)	-0.190** (0.093)
Speak national language	-0.003 (0.037)	-0.000 (0.029)	Major in math	0.097 (0.124)	0.042 (0.078)
Boy	0.076 (0.062)	0.053 (0.035)	Male	-0.081 (0.123)	0.040 (0.102)
<u>Home characteristics</u>			<u>Ave. Science Teacher</u>		
Some books	-0.022 (0.041)	-0.056** (0.026)	Teaching experience	-1.162 (2.731)	-1.288 (1.964)
Calculator	0.068 (0.044)	-0.059* (0.031)	Teacher certificate	0.130 (0.105)	-0.157* (0.090)
Computer	-0.028 (0.038)	-0.052 (0.032)	Major in science	0.068 (0.115)	-0.018 (0.089)
Study desk	0.006 (0.041)	-0.033 (0.029)	Male	-0.113 (0.118)	0.036 (0.098)
Dictionary	-0.058 (0.041)	-0.019 (0.030)			

Notes: Classroom standard deviation of age is the key independent variable. The constant term and coefficients of age and age squared not reported. Depending on whether the dependent variable is a student characteristic or a teacher characteristic, either a student non-response indicator or a teacher non-response indicator is included to control for missing values. Teacher characteristics are averages because multiple teachers are involved in some cases. Regressions are weighted by the sampling weights. Robust standard errors clustered by school reported in parentheses. See data appendix for variable construction.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Simulated Age Distribution and Instrumental Variables

	(1)	(2)
	----- Classroom Age -----	
	Ave.	S.D.
Simulated Classroom Ave. Age	0.021** (0.010)	
Simulated Classroom S.D. Age		0.021 (0.014)
Observations	22841	22841
R-squared	0.922	0.812

Notes: Only students with both mathematics and science test scores available are included. Indicators for non-responses to student survey and teacher survey are included to control for missing values. Simulated classroom average age and standard deviation of age are constructed based on the assumption that students are perfectly sorted by age and assigned into classrooms of equal size. All regressions include school fixed effects, student and teacher characteristics, and other age variables. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Classroom Age Heterogeneity on Mathematics Achievement

	(1)	(2)	(3)	(4)	(5)
Classroom S.D. Age	-1.063*** (0.160)	-0.346*** (0.106)	-0.369*** (0.102)	-0.359*** (0.093)	-0.366*** (0.093)
Classroom Ave. Age	0.705*** (0.061)	-0.243*** (0.076)	-0.101 (0.074)	-0.064 (0.068)	-0.077 (0.067)
Age	1.432*** (0.206)	0.791*** (0.169)	0.856*** (0.139)	0.634*** (0.127)	0.631*** (0.125)
Age squared	-0.071*** (0.010)	-0.041*** (0.008)	-0.044*** (0.006)	-0.033*** (0.006)	-0.033*** (0.006)
Fixed Effects	No	Country	School	School	School
Student and Teacher Characteristics	No	No	No	Yes	Yes
Instrumental Variables (IV)	No	No	No	No	Yes
First-stage Summary:					
Partial F for S.D. Age IV					66466
- Shea Partial R-squared					0.992
Partial F for Ave. Age IV					21112
- Shea Partial R-squared					0.984
Observations	22841	22841	22841	22841	22841
R-squared	0.182	0.392	0.541	0.586	0.110

Notes: Only students with both mathematics and science test scores available are included. Indicators for non-responses to student survey and teacher survey are included in specification (4) to control for missing values. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Classroom Age Heterogeneity on Science Achievement

	(1)	(2)	(3)	(4)	(5)
Classroom S.D. Age	-0.774*** (0.147)	-0.394*** (0.117)	-0.391*** (0.108)	-0.388*** (0.102)	-0.388*** (0.100)
Classroom Ave. Age	0.678*** (0.059)	-0.160** (0.076)	-0.081 (0.067)	-0.032 (0.062)	-0.045 (0.062)
Age	0.858*** (0.223)	0.493** (0.203)	0.622*** (0.167)	0.415*** (0.154)	0.415*** (0.152)
Age squared	-0.044*** (0.010)	-0.027*** (0.009)	-0.033*** (0.008)	-0.022*** (0.007)	-0.022*** (0.007)
Fixed Effects	No	Country	School	School	School
Student and Teacher Characteristics	No	No	No	Yes	Yes
Instrumental Variables	No	No	No	No	Yes
First-stage Summary:					
Partial F for S.D. Age IV					130000
- Shea Partial R-squared					0.996
Partial F for Ave. Age IV					12940
- Shea Partial R-squared					0.975
Observations	22841	22841	22841	22841	22841
R-squared	0.141	0.358	0.517	0.559	0.100

Notes: Only students with both mathematics and science test scores available are included. Indicators for non-responses to student survey and teacher survey are included in specification (4) to control for missing values. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Classroom Age Heterogeneity on Behaviors and Attitudes

	(1)	(2)	(3)	(4)	(5)	(6)
	----- All Students -----			----- Young Students -----		
	Bullied	Left out	Like school	Bullied	Left out	Like school
<u>School FE Results</u>						
Classroom S.D. Age	0.021 (0.056)	0.014 (0.039)	-0.042 (0.035)	-0.025 (0.065)	-0.011 (0.052)	-0.017 (0.050)
Classroom Ave. Age	0.034 (0.035)	0.035 (0.029)	-0.008 (0.025)	0.079* (0.048)	0.053 (0.041)	-0.047 (0.044)
Age	-0.187 (0.119)	-0.183 (0.116)	0.032 (0.072)	-0.019 (0.200)	0.002 (0.158)	0.356** (0.162)
Age squared	0.009 (0.006)	0.009* (0.005)	-0.002 (0.003)	-0.000 (0.010)	-0.000 (0.008)	-0.018** (0.008)
Observations	22841	22841	22841	12539	12539	12539
R-squared	0.128	0.093	0.144	0.145	0.108	0.156
<u>IV Results</u>						
Classroom S.D. Age	0.028 (0.056)	0.011 (0.039)	-0.040 (0.035)	-0.020 (0.066)	-0.012 (0.051)	-0.011 (0.050)
Classroom Ave. Age	0.018 (0.034)	0.023 (0.029)	-0.008 (0.025)	0.066 (0.048)	0.042 (0.040)	-0.043 (0.043)
Age	-0.185 (0.118)	-0.184 (0.113)	0.032 (0.071)	-0.020 (0.195)	-0.000 (0.155)	0.358** (0.158)
Age squared	0.009 (0.006)	0.009* (0.005)	-0.002 (0.003)	-0.000 (0.010)	-0.000 (0.008)	-0.019** (0.008)
First-stage Summary:						
Partial F for S.D. Age IV	17589	17589	17589	16927	16927	16927
- Shea Partial R-squared	0.983	0.983	0.983	0.982	0.982	0.982
Partial F for Ave. Age IV	3096	3096	3096	3163	3163	3163
- Shea Partial R-squared	0.949	0.949	0.949	0.944	0.944	0.944
Observations	22841	22841	22841	12539	12539	12539
R-squared	0.004	0.009	0.043	0.004	0.008	0.037

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student characteristics listed in Table 2, as well as indicators for non-responses to the student survey. Young students are those at the median age or below. Median age is defined according to the grade-level age distribution of the school. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Classroom Age Heterogeneity on Achievement by Student Age

	(1)	(2)	(3)	(4)
	----- Mathematics -----		----- Science -----	
	Old	Young	Old	Young
<u>School FE Results</u>				
Classroom S.D. Age	-0.325*** (0.121)	-0.341*** (0.094)	-0.331*** (0.120)	-0.383*** (0.101)
Classroom Ave. Age	-0.055 (0.094)	-0.076 (0.070)	0.022 (0.086)	-0.098 (0.072)
Age	-1.080*** (0.320)	0.432* (0.221)	-1.859*** (0.375)	0.329 (0.235)
Age squared	0.038*** (0.014)	-0.021* (0.011)	0.072*** (0.016)	-0.016 (0.012)
Observations	10301	12539	10301	12539
R-squared	0.624	0.585	0.606	0.555
<u>IV Results</u>				
Classroom S.D. Age	-0.338*** (0.118)	-0.353*** (0.093)	-0.335*** (0.115)	-0.385*** (0.099)
Classroom Ave. Age	-0.046 (0.090)	-0.089 (0.069)	0.020 (0.082)	-0.110 (0.070)
Age	-1.082*** (0.311)	0.429** (0.216)	-1.859*** (0.364)	0.327 (0.229)
Age squared	0.038***	-0.021*	0.072***	-0.016
First-stage Summary:				
Partial F for S.D. Age IV	63814	63299	63814	63299
- Shea Partial R-squared	0.993	0.993	0.993	0.993
Partial F for Ave. Age IV	15145	21491	15145	21491
- Shea Partial R-squared	0.985	0.984	0.985	0.984
Observations	10301	12539	10301	12539
R-squared	0.125	0.101	0.122	0.090

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. “Old” students are those above the median age of students in the same school and grade level. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Classroom Age Heterogeneity on Achievement by Gender

	(1)	(2)	(3)	(4)
	----- Mathematics -----		----- Science -----	
	Boy	Girl	Boy	Girl
<u>School FE Results</u>				
Classroom S.D. Age	-0.321*** (0.111)	-0.346*** (0.103)	-0.339*** (0.121)	-0.372*** (0.108)
Classroom Ave. Age	-0.088 (0.087)	-0.045 (0.068)	-0.075 (0.081)	0.006 (0.065)
Age	0.697*** (0.168)	0.557*** (0.177)	0.540** (0.219)	0.341** (0.168)
Age squared	-0.036*** (0.008)	-0.029*** (0.008)	-0.028*** (0.010)	-0.019** (0.008)
Observations	11547	11293	11547	11293
R-squared	0.604	0.593	0.585	0.558
<u>IV Results</u>				
Classroom S.D. Age	-0.322*** (0.109)	-0.358*** (0.102)	-0.338*** (0.117)	-0.372*** (0.105)
Classroom Ave. Age	-0.106 (0.085)	-0.055 (0.067)	-0.089 (0.079)	-0.009 (0.063)
Age	0.697*** (0.164)	0.554*** (0.172)	0.540** (0.213)	0.341** (0.164)
Age squared	-0.036*** (0.007)	-0.028*** (0.008)	-0.028*** (0.010)	-0.019** (0.008)
First-stage Summary:				
Partial F for S.D. Age IV	57659	64796	57659	64796
- Shea Partial R-squared	0.992	0.993	0.992	0.993
Partial F for Ave. Age IV	16326	24241	16326	24241
- Shea Partial R-squared	0.984	0.985	0.984	0.985
Observations	11547	11293	11547	11293
R-squared	0.124	0.097	0.109	0.088

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Sensitivity of Estimates to Functional Forms of Age

	(1)	(2)	(3)	(4)	(5)	(6)
<u>School FE Results</u>						
Classroom S.D. Age	-0.394*** (0.092)	-0.359*** (0.093)	-0.359*** (0.093)	-0.413*** (0.100)	-0.388*** (0.102)	-0.389*** (0.102)
Classroom Ave. Age	-0.061 (0.067)	-0.064 (0.068)	-0.063 (0.068)	-0.030 (0.062)	-0.032 (0.062)	-0.029 (0.063)
Age	-0.071*** (0.011)	0.634*** (0.127)	1.102* (0.663)	-0.068*** (0.012)	0.415*** (0.154)	1.656** (0.662)
Age squared		-0.033*** (0.006)	-0.076 (0.064)		-0.022*** (0.007)	-0.139** (0.064)
Age cubed			0.001 (0.002)			0.004* (0.002)
Observations	22841	22841	22841	22841	22841	22841
R-squared	0.585	0.586	0.586	0.559	0.559	0.559
<u>IV Results</u>						
Classroom S.D. Age	-0.401*** (0.092)	-0.366*** (0.093)	-0.366*** (0.093)	-0.412*** (0.098)	-0.388*** (0.100)	-0.389*** (0.100)
Classroom Ave. Age	-0.076 (0.066)	-0.077 (0.067)	-0.073 (0.067)	-0.045 (0.061)	-0.045 (0.062)	-0.041 (0.062)
Age	-0.070*** (0.011)	0.631*** (0.125)	1.092* (0.655)	-0.067*** (0.011)	0.415*** (0.152)	1.647** (0.653)
Age squared		-0.033*** (0.006)	-0.076 (0.063)		-0.022*** (0.007)	-0.138** (0.063)
Age cubed			0.001 (0.002)			0.004* (0.002)
First-stage Summary:						
Partial F for S.D. Age IV	65702	66813	66542	130000	130000	130000
- Shea Partial R-squared	0.992	0.992	0.992	0.996	0.996	0.996
Partial F for Ave. Age IV	21231	22275	21622	13015	12940	13575
- Shea Partial R-squared	0.984	0.985	0.985	0.975	0.975	0.976
Observations	22841	22841	22841	22841	22841	22841
R-squared	0.109	0.110	0.110	0.099	0.100	0.100

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 10: Sensitivity to Alternative Measures of Classroom Age Heterogeneity

	(1)	(2)	(3)	(4)
	----- Mathematics -----		----- Science -----	
75 th -25 th Percentile Age Difference	-0.253*** (0.070)		-0.252*** (0.069)	
Max-Min Age Difference		-0.052*** (0.017)		-0.059*** (0.017)
Classroom Ave. Age	-0.068 (0.075)	-0.105 (0.069)	-0.039 (0.072)	-0.075 (0.064)
Age	0.679*** (0.128)	0.672*** (0.129)	0.470*** (0.151)	0.453*** (0.155)
Age squared	-0.035*** (0.006)	-0.034*** (0.006)	-0.025*** (0.007)	-0.024*** (0.007)
Observations	22841	22841	22841	22841
R-squared	0.586	0.585	0.559	0.558
IV Results				
75 th -25 th Percentile Age Difference	-0.256*** (0.069)		-0.249*** (0.068)	
Max-Min Age Difference		-0.053*** (0.017)		-0.060*** (0.016)
Classroom Ave. Age	-0.052 (0.072)	-0.095 (0.067)	-0.021 (0.070)	-0.066 (0.061)
Age	0.678*** (0.126)	0.671*** (0.127)	0.471*** (0.149)	0.452*** (0.153)
Age squared	-0.035*** (0.006)	-0.034*** (0.006)	-0.025*** (0.007)	-0.024*** (0.007)
First-stage Summary:				
Partial F for S.D. Age IV	42980	57179	40143	42383
- Shea Partial R-squared	0.993	0.993	0.992	0.991
Partial F for Ave. Age IV	2711	6036	2758	5266
- Shea Partial R-squared	0.945	0.971	0.939	0.960
Observations	22841	22841	22841	22841
R-squared	0.111	0.109	0.100	0.099

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. The mean and standard deviation of 75th-25th percentile age difference are 0.585 and 0.256 respectively. The mean and standard deviation of max-min age difference are 1.906 and 1.092 respectively. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Estimates by TIMSS 2003 and TIMSS 2007

	(1)	(2)	(3)	(4)
	----- TIMSS 2003 -----		----- TIMSS 2007 -----	
	Mathematics	Science	Mathematics	Science
<u>School FE Results</u>				
Classroom S.D. Age	-0.538** (0.223)	-0.393* (0.223)	-0.347*** (0.096)	-0.381*** (0.113)
Classroom Ave. Age	-0.261 (0.230)	-0.230 (0.185)	0.001 (0.062)	0.011 (0.061)
Age	1.092** (0.476)	0.829* (0.443)	0.577*** (0.124)	0.388** (0.165)
Age squared	-0.053** (0.022)	-0.040** (0.020)	-0.030*** (0.006)	-0.021*** (0.008)
Observations	6487	6487	16354	16354
R-squared	0.511	0.508	0.614	0.578
<u>IV Results</u>				
Classroom S.D. Age	-0.548** (0.224)	-0.387* (0.220)	-0.343*** (0.095)	-0.383*** (0.111)
Classroom Ave. Age	-0.278 (0.227)	-0.230 (0.186)	-0.023 (0.062)	-0.006 (0.060)
Age	1.087** (0.467)	0.834* (0.436)	0.577*** (0.122)	0.387** (0.163)
Age squared	-0.053** (0.021)	-0.040** (0.020)	-0.030*** (0.006)	-0.021*** (0.008)
First-stage Summary:				
Partial F for S.D. Age IV	11473	32893	77873	15000
- Shea Partial R-squared	0.989	0.992	0.995	0.998
Partial F for Ave. Age IV	4461	1590	29052	16611
- Shea Partial R-squared	0.973	0.934	0.990	0.986
Observations	6487	6487	16354	16354
R-squared	0.137	0.103	0.105	0.102

Notes: Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 12: Comparison of the Effects in the Fourth Grade and Eighth Grade

	(1)	(2)	(3)	(4)
	TIMSS 2003		TIMSS 2007	
	---- Fourth Grade ----		---- Eighth Grade ----	
	Mathematics	Science	Mathematics	Science
<u>School FE Results</u>				
Classroom S.D. Age	-0.424*	-0.254	-0.364	-0.171
	(0.221)	(0.220)	(0.243)	(0.347)
Classroom Ave. Age	-0.258	-0.236	-0.062	-0.234
	(0.235)	(0.191)	(0.230)	(0.268)
Age	1.480***	1.164**	0.192	0.887
	(0.465)	(0.470)	(0.701)	(1.023)
Age squared	-0.071***	-0.056***	-0.010	-0.033
	(0.021)	(0.021)	(0.023)	(0.034)
Observations	6036	6036	4934	4934
R-squared	0.517	0.513	0.266	0.339
<u>IV Results</u>				
Classroom S.D. Age	-0.435*	-0.244	-0.358	-0.092
	(0.223)	(0.216)	(0.243)	(0.327)
Classroom Ave. Age	-0.282	-0.239	-0.066	-0.262
	(0.232)	(0.193)	(0.226)	(0.259)
Age	1.474***	1.172**	0.197	0.949
	(0.454)	(0.462)	(0.688)	(0.982)
Age squared	-0.071***	-0.056***	-0.011	-0.035
	(0.020)	(0.021)	(0.023)	(0.033)
First-stage Summary:				
Partial F for S.D. Age IV	9156	26686	50154	1819
- Shea Partial R-squared	0.989	0.991	0.994	0.959
Partial F for Ave. Age IV	4290	1501	780000	2123
- Shea Partial R-squared	0.973	0.932	1.00	0.963
Observations	6036	6036	4934	4934
R-squared	0.137	0.103	0.053	0.034

Notes: The sample includes Armenia, Latvia, and Lithuania. Only students with both mathematics and science test scores available are included. All regressions include school fixed effects, a set of student and teacher characteristics listed in Table 2, as well as indicators for non-responses to student and teacher surveys. Teacher characteristics vary across subjects. Regressions are weighted by sampling weights. Robust standard errors clustered by schools are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 13: Simulation – Age Grouping and Achievement Gains in Grade Four

Country	Grade-level Age S.D.	Average Grouping – Mixing <u>Differences in Classroom Age</u>		--- Predicted Gain ---	
		S.D.	Ave.	Math	Science
Armenia	0.490	-0.172	0.000	0.063	0.067
Colombia	1.131	-0.310	0.012	0.113	0.120
El Salvador	1.072	-0.434	0.285	0.137	0.156
Georgia	0.460	-0.167	0.020	0.059	0.064
Kazakhstan	0.530	-0.164	-0.007	0.060	0.064
Latvia	0.448	-0.175	0.009	0.064	0.068
Lithuania	0.420	-0.145	0.009	0.052	0.056
Mongolia	0.913	-0.472	0.419	0.140	0.164
Morocco	1.038	-0.626	0.310	0.205	0.229
Russia	0.498	-0.144	0.007	0.052	0.056
Tunisia	0.680	-0.199	0.011	0.072	0.077
Ukraine	0.473	-0.151	0.024	0.054	0.058
Yemen	1.214	-0.170	0.074	0.057	0.063
Average	0.721	-0.256	0.090	0.087	0.095

Notes: The simulation is based on ability-mixing schools with at least two classrooms sampled in TIMSS 2007 (ability-grouping schools are excluded). Grade level standard deviation of age is the sample standard deviation of age for the whole country. The point estimates used to construct the predicted gains are sourced from specification (4) in Table 3 and Table 4.

Table 14: Simulation – Age Grouping and Distributional Effects in Grade Four

<i>Country</i>	Age Grouping – Mixing Differences in				----- Predicted Gain -----			
	<u>Classroom S.D. Age</u>		<u>Classroom Ave. Age</u>		---- Math ----		---- Science ----	
	Young	Old	Young	Old	Young	Old	Young	Old
Armenia	-0.175	-0.159	-0.282	0.326	0.087	0.039	0.098	0.060
Colombia	-0.386	-0.210	-0.585	0.662	0.188	0.040	0.213	0.083
El Salvador	-0.462	-0.406	-0.602	0.645	0.217	0.108	0.244	0.149
Georgia	-0.185	-0.137	-0.276	0.314	0.090	0.032	0.101	0.052
Kazakhstan	-0.155	-0.163	-0.261	0.310	0.078	0.041	0.088	0.061
Latvia	-0.226	-0.115	-0.245	0.296	0.101	0.025	0.114	0.045
Lithuania	-0.138	-0.157	-0.236	0.285	0.070	0.040	0.079	0.058
Mongolia	-0.262	-0.350	-0.447	0.503	0.132	0.095	0.150	0.127
Morocco	-0.251	-0.555	-0.381	0.428	0.123	0.168	0.139	0.195
Russia	-0.130	-0.156	-0.245	0.298	0.068	0.039	0.077	0.058
Tunisia	-0.328	-0.033	-0.310	0.385	0.143	-0.006	0.160	0.019
Ukraine	-0.176	-0.124	-0.253	0.298	0.085	0.028	0.096	0.047
Yemen	-0.288	-0.111	-0.475	0.554	0.144	0.012	0.163	0.048
Average	-0.243	-0.206	-0.354	0.408	0.117	0.051	0.133	0.077

Notes: The simulation is based on ability-mixing schools with at least two classrooms sampled in TIMSS 2007 (ability-grouping schools are excluded). Grade level standard deviation of age is the sample standard deviation of age for the whole country. “Young” students are at the median age or younger; “Old” students are those above the median age of their school. The point estimates used to construct the predicted gains are sourced from Table 6.

Data Appendix

1. Sample selection

The four countries sampled from TIMSS 2003 (T03) are Armenia, Latvia, Lithuania, and Moldova. The thirteen countries sampled from TIMSS 2007 (T07) are Armenia, Colombia, El Salvador, Georgia, Kazakhstan, Latvia, Lithuania, Mongolia, Morocco, Russia, Tunisia, Ukraine, and Yemen. These countries are selected because they were classified as low and middle income countries by the World Bank in 2007 and they sampled multiple classrooms in several schools in TIMSS. The principals of the sampled schools stated that their students were not grouped into different classrooms on the basis of their ability in mathematics and science. Note that schools that grouped students according to ability in either only math or science are also excluded. Students with missing test scores and age are dropped from the final sample.

2. Variable Construction

a. Standardized test scores

Scaled scores reported by TIMSS are standardized with respect to the standard normal distribution (within each wave of TIMSS) using the full TIMSS sample.

b. Age variables

The precision of age is only up to the month of birth. All age variables used are measured in years and based on the variable “asdage” in TIMSS data files. Median age is defined by the median age of each student’s school. “Old” means above the median age, and “young” means at the median age or below.

c. Other dependent variables

“Bullied” is a dummy variable taking the value of 1 if a student was reported to have been hurt (T03’s “as4ghurt” or T07’s “asbghurt”), made to do things (“as4gmade” or “asbgmade”), or teased (“as4gmfun” or “asbgmfun”) by other students in school. “Left out” is a dummy variable taking the value of 1, if a student was ever left out of activities by other students in school (“as4gleft” or asbgleft”). “Like school” is a dummy variable taking the value of 1 if a student agreed with the statement that he/she liked going to school (“as4galbs” or “asbgalbs”).

d. Nativity and language variables

“Native born” is a dummy variable taking the value of 1, if a student was born the in country (T03’s “as4gborn” or T07’s “asbgborn”). “Parent native born” is a dummy variable taking the value of 1, if a student’s father or mother was born in the country (T03’s “asbgmbrn” and

“asbgfbrn” or T07’s “asdgborn”). “Speak national language” is a dummy variable taking the value of 1, if a student always or almost always speaks the language of test at home (T03’s “as4golan” or T07’s “asbgolan”).

e. Things available at home

“Some books” is a dummy variable taking the value of 1, if a student was reported to have at least 11 books at home. “Calculator” is a dummy variable taking the value of 1, if a student was reported to have a calculator at home. “Study desk” is a dummy variable taking the value of 1, if a student was reported to have a study desk at home. “Dictionary” is a dummy variable taking the value of 1, if a student was reported to have a dictionary at home.

f. Teaching characteristics

Teacher’s experience is the average years of teaching experience of a student’s teachers, because some students have multiple teachers for each subject.

g. Teaching certificate

Teaching certificate is the average of the binary variable indicating whether a student’s teacher in a subject has the relevant teaching certificate. Average is used because some students have multiple teachers for a subject.

h. Teacher’s major

Teacher’s major is the average of the binary variable indicating whether a student’s teacher in a subject majored in the subject during college. Average is used because some students have multiple teachers for a subject.