RISE Working Paper 21/078
September 2021

What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the Young Lives Data from Ethiopia, Peru, India, and Vietnam

Paul Glewwe, Zoe James, Jongwook Lee, Caine Rolleston, Khoa Vu

Abstract

Vietnam's strong performance on the 2012 and 2015 PISA assessments has led to interest in what explains the strong academic performance of Vietnamese students. Analysis of the PISA data has not shed much light on this issue. This paper analyses a much richer data set, the Young Lives data for Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam, to investigate the reasons for the strong academic performance of 15-year-olds in Vietnam. Differences in observed child and household characteristics explain 37-39 percent of the gap between Vietnam, and Ethiopia, while observed school variables explain only about 3-4 additional percentage points (although an important variable, math teachers' pedagogical skills, is not available for Ethiopia). Differences in observed child and household characteristics explain very little of the gaps between Vietnam and India and between Vietnam and Peru, yet one observed school variable has a large explanatory effect: primary school math teachers' pedagogical skills. It explains about 10-12 percent of the gap between Vietnam and India, raising the overall explained portion to 14-21 percent of the gap. For Peru, it explains most (65-84 percent) of the gap.

What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the Young Lives Data from Ethiopia, Peru, India, and Vietnam

Paul Glewwe
University of Minnesota

Zoe James
Cambridge University

Jongwook Lee
University of Minnesota

Caine Rolleston
University College London

Khoa Vu
University of Minnesota

Acknowledgements:

We thank Richard Akresh and Sylvie Lambert, as well as seminar participants at the 2019 Comparative and International Education Society conference, the 2019 Midwest International Economic Development Conference, Tsukuba University, Sophia University, University of Tokyo, Hitotsubashi University, Kobe University, 2020 ASSA, Stanford University (REAP), Paris School of Economics, Université Catholique Louvain, University of Bologna, and Lancaster University for many helpful comments. We also thank Rayyan Mobarak for excellent research assistance.

This is one of a series of working papers from "RISE"—the large-scale education systems research programme supported by funding from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), the Australian Government's Department of Foreign Affairs and Trade (DFAT), and the Bill and Melinda Gates Foundation. The Programme is managed and implemented through a partnership between Oxford Policy Management and the Blavatnik School of Government at the University of Oxford.

Please cite this paper as:
Glewwe, P., James, Z., Lee, J., Rolleston, C. and Vu, K. 2021. What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the Young Lives Data from Ethiopia, Peru, India, and Vietnam. RISE Working Paper Series. 21/078. https://doi.org/10.35489/BSG-RISE-WP_2021/078

I. Introduction

Vietnam's rapid economic growth in the last 30 years has transformed it from one of the world's poorest countries to a middle-income country (World Bank and MPI, 2016). While its economic achievements have attracted much attention, in recent years its accomplishments in education have generated even more international interest. Most striking is its performance on the 2012 and 2015 PISA assessments: In 2012 Vietnam ranked 16th in math and 18th in reading out of 63 countries,[1] ahead of both the US and the UK and much higher than any other developing country (OECD, 2014). Its 2012 PISA mathematics and readings scores (at 511 and 508), for example, were more than one standard deviation higher than those of Indonesia (375 and 396), another Southeast Asian country, which is the closest to Vietnam of all the 2012 PISA participating countries in terms of GDP per capita.[2] While its 2015 performance was slightly lower, ranking 21st in math and 31st in reading out of 68 countries, it still outperformed all other developing countries and outperformed the US in both subjects and the UK in math (but not in reading).

Vietnam's achievements in education are particularly notable given its relatively low GDP per capita. This is shown in Figures 1 and 2, which plot 2012 PISA scores in math and reading by the log of per capita GDP. Vietnam is in the upper left in both figures, higher than any other country above the line that shows the expected test score given per capita GDP. Vietnam is also the largest positive outlier (relative to the fitted line) when PPP (purchasing power parity) per capita GDP is used and when the 2015 PISA data are used (Dang et al., 2020).

Dang et al. (2020) used the 2012 and 2015 PISA data to try to understand Vietnam's very high performance on those assessments of student learning. However, the PISA data have several

[1] We consider only country-level PISA data. Thus, we exclude Shanghai, which is not representative of China as a whole, and the Perm territory, which is not representative of Russia. For convenience, we treat Hong Kong, Macao and Taiwan as countries, though Hong Kong and Macao are Chinese territories and Taiwan's status is under dispute.
[2] Indonesia's GDP per capita was $US 3,332 in 2015, while Vietnam's was $US 2,085 (World Bank, 2017).

limitations.[3] First, they exclude children who are not in school, which excludes about one third of Vietnam's 15-year-olds (Dang et al., 2020). Second, the PISA data are collected only when students are 15 years old, and not at any earlier age. Third, the school-level data include only the schools the students currently attend, not the schools attended previously. Fourth, the school-level data have some limitations. For example, the teacher absence question simply asks the school principal whether such absences hinder student learning, the possible responses being: a) Not at all; b) Very little; c) To some extent; or d) A lot.[4] Fifth, Vietnam appears to have done more than other countries to prepare its students for the PISA assessment, which could explain at least part of its strong performance on that assessment (see Dang et al., 2020, for details).

This paper uses a different data source to examine the nature and underlying determinants of Vietnam's apparent exceptional performance: the Young Lives data collected from Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam. While the number of countries in the Young Lives data (4) is much smaller than that in the PISA data (63), the former data have several advantages over the latter. First, the Young Lives test score data at age 15 include *all* 15-year-old children, regardless of whether they were in school. Second, the Young Lives data were collected from the children over 14 years, when they were 1, 5, 8, 12 and 15 years old, and include much more detailed information than the PISA student-level data. Third, the Young Lives data include very detailed data from the primary schools attended by a subsample of the Young Lives children when they were in grades 4 or 5, as well as very detailed data from the secondary schools that they attended, if any, when they were about 14 years old. Fourth, relative to the PISA data, the Young Lives school data are much richer, including school facility,

[3] Despite these limitations, Vietnam is still an outlier after correcting for them (see Dang et al., 2020).
[4] This information on teacher absence could vary across countries for a given level of teacher absence. For example, a rate of teacher absence of 10% of school days would be considered a serious problem in Vietnam, while in India such a rate would be much lower than average and thus likely would not be considered to be a problem.

principal and teacher questionnaires, and school observation data. Fifth, the Young Lives data attract little or no media attention and thus there is little reason to think that Vietnam "prepped" its 15-year-olds who participated in the Young Lives' academic assessments.

At best, the analysis of the PISA data by Dang et al. (2020) explains only one third of the gap between Vietnam's strong performance on that assessment and the performance one would expect given its income level. The Young Lives data, which are much more detailed than the PISA data, may be able to explain a larger proportion of the gap between Vietnam and the three other countries in the Young Lives data. This paper investigates what more can be learned about Vietnam's exceptional performance in education by analyzing the Young Lives data. It focuses on performance on the mathematics tests given at age 15, since comparisons of language ability can be confounded by linguistic differences across languages.[5]

Two different econometric methodologies are used. The first begins by regressing test scores on country-level dummy variables, which replicates the gaps in mean test scores between Vietnam and the other countries. It then adds explanatory variables to investigate the extent to which sets of variables explain the gap (reduce the coefficients on the country-level dummy variables). This is similar to the approach of Fryer and Levitt (2004), who studies test score gaps between white and black children in the United States. The second applies the Oaxaca-Blinder decomposition, which uses regression analysis to decompose test scores gaps into the portion due to differences between Vietnam and the other countries in observed variables and the portion due to differences between Vietnam and the other countries in the coefficients on those variables.

[5] All four countries administered a reading comprehension test and the Peabody Picture Vocabulary Test (PPVT) in Round 5. The PPVT is more likely to be comparable across countries, yet Cueto and León (2012, p.35), who advised the Young Lives study on reading and math tests, advise "not using the [PPVT] scores across language groups", which means not using them to compare across countries (and across languages within countries). They also say that "[w]hile local teams have worked to adapt the test to its local language, … some bias is likely [to] remain."

The results can be summarized as follows. The Young Lives data only partially explain the strong performance of Vietnam's 15-year-olds relative to their counterparts in Ethiopia, India and Peru. Differences in observed child and household characteristics explain 37-39% of the gap between Vietnam and Ethiopia, while observed school variables explain only about 3-4 additional percentage points (although an important variable, primary school math teachers' pedagogical skills, is not available for Ethiopia). In contrast, differences in observed child and household characteristics explain very little of the gaps between Vietnam and India and between Vietnam and Peru. Yet one school variable has a large explanatory effect: primary school math teachers' pedagogical skills. It explains 10-12% of the gap between Vietnam and India, raising the overall explained portion to 14-21% of the gap. For Peru, it explains *most* (65-84%) of the gap.

II. The Young Lives Study

This paper uses data from the Young Lives Study, which follows two cohorts of children over 15 years in four developing countries: Ethiopia, India (Andhra Pradesh and Telengana), Peru and Vietnam. This analysis uses only the younger of the two cohorts, which is a sample of about 2,000 children in each of the four countries. Data were collected from the younger cohort children in 2002 (when they were about 1 year old), 2006 (5 years old), 2009 (8 years old), 2013 (12 years old) and 2016 (15 years old).

The data collected in each round is very detailed, with questionnaires for children (started at age 8), parents, community leaders, school principals, and teachers. Most of the data collected in each round are from the household questionnaire. That questionnaire varied somewhat over the different rounds, yet it collected the following in all rounds: education levels of all household members, with more detail on members age 5-17; household members' income-generating activities, including ownership of productive assets; household consumption of food (both self-produced and purchased), and expenditures on non-food items; social capital networks; recent economic changes; dwelling characteristics; ownership of durable goods; child health, including height and weight measurements; and caregiver perceptions and attitudes.

The schools surveys collected information on the principal's personal characteristics, basic school building characteristics, fees charged, basic teacher characteristics, and educational materials (e.g. libraries, computers). The teacher surveys collected data on the teachers' personal characteristics, including education and teacher training, classroom conditions and pedagogical materials, and attitudes regarding teaching.

Several tests were administered to the younger cohort children at different ages. In 2006, when they were about 5 years old, they took the Peabody Picture Vocabulary Test (PPVT) and a very simple mathematics test. In 2009, when they were about 8 years old, they took the PPVT, the USAID's Early Grade Reading Assessment (EGRA), and a mathematics test. In 2013, when they were about 12 years old, they took the PPVT, a reading comprehension test, and a mathematics test. Finally, in 2016, when they were about 15 years old, they were again given the PPVT, a reading comprehension test, and a mathematics test. For more details on the Young Lives data, see https://www.younglives.org.uk/.

The Young Lives Younger cohort children are close to being national representative (representative of Andhra Pradesh and Telangana for India) for all four countries, but are not exactly nationally representative. The Ethiopia sample was selected in two steps. First, five (Addis Ababa, Amhara, Oromiya, SNNP, and Tigray) of Ethiopia's nine regions were selected to ensure national coverage; these regions cover 96% of Ethiopia's population. Within each region, three to five districts were selected to have a balanced representation of rural-poor, urban-poor, and less poor rural and urban households. Within each district, one peasant association (rural areas) or *kebele* (urban areas) was selected. In the second step, a village is randomly drawn from each peasant association or *kebele*, and all households are interviewed until 100 eligible households are located. Outes-Leon and Sanchez (2008) show that, although the Young Lives Ethiopian sample is not nationally representative, it represents "a wide range of living standards akin to the variability found in the Ethiopian population" and "covers the diversity of the children in the country and a wide variety of attributes and experiences."

The India sample is representative of the state of Andhra Pradesh (which later was split into Andhra Pradesh and Telengana). It consists of three regions: Coastal Andhra, Rayalseema and Telangana. In the first stage, one poor and one non-poor district were selected from each region, and 20 *mandals* were selected as sentinel sites from these six districts and Hyderabad (the state capital). In the second stage, each *mandal* was divided into four geographical areas, and one village was randomly selected from each area. Kumra (2008) found that the sampled households are similar to, but slightly better-off than, the average Andhra Pradesh household.

The Peru sample is nationally representative of 95% of Peru's districts; the 5% with the lowest poverty index were excluded from the sample frame. Twenty districts were randomly selected from this sample frame, and within each of these districts one census tract was randomly selected. Within each of these 20 census tracts, one cluster or block of households was randomly chosen, and all households were visited to determine whether they had children of the appropriate age. One hundred households with children of the appropriate age were randomly selected from these households. If the cluster or block did not have 100 such households, households in the nearest adjacent cluster or block were selected until 100 households were obtained. The 100 households from each of 20 different districts yielded a sample of 2,000 children. For further details, see Escobal and Flores (2008), who conclude that "Young Lives households are very similar to the average household in Peru".

Strictly speaking, the Young Lives sample for Vietnam is not nationally representative. Vietnam can be divided into eight socio-economic regions.[6] To ensure that the Young Lives sample included a major urban center, a ninth "region" ("Cities") was created consisting of the major urban provinces (Hanoi, Ho Chi Minh City, Da Nang, Hai Phong and Ba Ria-Vung Tau). Of these nine "regions", five (North-East, Red River Delta, Cities, South Central Coast, and Mekong River Delta) were chosen as "representative" of Vietnam in that they: (1) include urban, rural and mountainous areas; (2) include regions in the northern, central, and southern areas of Vietnam; (3) are relatively poor; and (4) reflect some unique factors of Vietnam, such as areas prone to natural disasters or heavily affected by past wars. From each of these five regions, a "typical" province was chosen after consulting government and international experts. Within each province, at least four communes were chosen, giving greater weight to poor communes. This selection of communes was not random; the criteria included: (1) whether the commune represents common provincial/regional features; (2) whether the local government showed commitment for the research; (3) feasibility in terms of research logistics; and (4) population size. Following this procedure, 31 communes were selected. Random sampling was then applied in each commune to select 100 children of the appropriate age. While the Young Lives sample is not strictly nationally representative, Glewwe, Chen and Katare (2015) conclude that "comparisons with the nationally representative Vietnam Household Living Standards Survey suggest that the Young Lives sample is broadly representative of Vietnam as a whole."

[6] These regions are: North-West, North-East, Red River Delta, North Central Coast, South Central Coast, South-East, Central Highlands, and Mekong River Delta.

For convenience, henceforth this paper refers to the Andhra Pradesh and Telangana data as the India data, but note that these data include only two of India's 28 states.

III. Vietnamese Children's Performance on the Young Lives Tests, and Possible Explanations

Vietnam outperformed all other developing countries (and many developed countries) on the 2012 and 2015 PISA assessments (see Dang et al., 2020). This is also true for the four Young Lives countries: Vietnam outperforms the other three on the mathematics test, as seen in Table 1.

More specifically, in the fifth round of data collection, in 2016, mathematics tests were administered to the younger cohort Young Lives children, who were about 15 years old. These tests varied somewhat across the four countries, but 23 math questions were used in all four countries. For these 23 questions, the mean number of questions correct was 5.5 for Ethiopia, 6.9 for India, 9.1 for Peru and 12.3 for Vietnam. In terms of standard deviations of the test score (for the combined distribution of all four countries), henceforth denoted by σ, the average mathematics scores in Vietnam were 1.4σ higher than those in Ethiopia, 1.1σ higher than those in India, and 0.7σ higher than those in Peru. It is not surprising that Vietnamese 15-year-olds outperform those in Ethiopia, given that Vietnam's GDP per capita is over three times higher ($2,085 vs. $641), but this cannot explain the gaps with India and Peru: India's GDP per capita ($1,606) is almost as high as Vietnam's and Peru's ($6,229) is three times higher than Vietnam's.

This also holds if IRT (item response theory) is used to compare the latent mathematics ability of these 15-year-olds. This was done using a 2-parameter IRT model. IRT analysis has the advantage of using *all* questions in each country's mathematics test, rather than being limited to the 23 questions that were in the math tests of all four countries. IRT also accounts for the fact that some questions are more difficult than others, and that some have more discriminating power than others. The mean latent mathematics skill for the Young Lives 15-year-olds is shown in the third column of Table 1. (Note that the mean of this variable for the entire sample is set to zero, and the standard deviation is close to one.) The ranking is similar for all four countries, and the gaps between pairs of countries are quite similar. In particular, in terms of the standardized score for the 23 mathematics questions, Vietnam is 0.67σ ahead of Peru, 1.12σ ahead of India, and 1.41σ ahead of Ethiopia, and for the latent IRT score, Vietnam is 0.73σ ahead of Peru, 0.97σ ahead of India, and 1.35σ ahead of Ethiopia.

Another way to see the disparity across countries in these mathematics scores is to compare their density functions. This is done in Figure 3 for the scores on the 23 questions that were used in all four countries. The distributions for Ethiopia and India are concentrated on the left of the diagram, peaking at about 5 answers correct (out of 23). Note that 17 of the 23 questions were multiple choice, with four possible responses (one had 5 possible responses), so that by randomly guessing a test taker could get a score of 4.2; this implies that about a third of Ethiopian 15-year-olds and a third of Indian 15-year-olds performed no better than someone who randomly guessed on all the multiple choice questions. In contrast, relatively few Vietnamese 15-year-olds had scores of 5 or less. Peru's fifteen-year-olds do somewhat better; their distribution peaks at around 8 questions correct. Perhaps the starkest contrast is for 15-year-olds with scores of 15 or higher. About 36% of Vietnamese 15-year-olds scored in this range, compared to only 0.8% of Ethiopian 15-year-olds and 4.0% of Indian 15-year-olds. Even for Peru, the wealthiest of these four countries, only 9.8% of 15-year-olds score in this range.

What explains the strong performance of 15-year-olds in Vietnam? Table 2 provides some possibilities based on the rich Young Lives child and household data. First, there is strong evidence that malnutrition in the first years of life can lead to low school performance (see

Alderman et al., 2017, for a recent summary). The first two rows of Table 2 show that Vietnamese children were better nourished at age 5, as measured by average height-for-age Z-scores and percent of children who are stunted, than children at that age in the three other countries. Yet these differences are small; for example, about 25% of Vietnam's 5-year-olds are stunted, compared to 31-36% in the other three countries.

Large family size is often negatively associated with children's educational outcomes, while family wealth is typically positively correlated. The next two rows of Table 2 indicate that Vietnamese children have fewer siblings, and are wealthier (in terms of an index of household durable goods and housing characteristics), than the children in the other three countries.[7] Yet these differences are small, except that Ethiopian children have many more siblings and much less wealth than do children in the other three countries.

Another possible reason for Vietnamese 15-year-olds' strong education performance is that their parents are highly educated, and so they may be more able to help their children with schoolwork. The next four rows of Table 2 show that Vietnamese parents have completed many more years of schooling than parents in Ethiopia and India; on average, Vietnamese mothers have 6.2 years of education, compared to 2.4 and 3.1 years in Ethiopia and India, respectively, and Vietnamese fathers have 7.0 years of education, compared to 3.5 and 4.7 years in Ethiopia and India. Yet mothers and fathers in Peru have 1.5-1.9 more years of schooling than their Vietnamese counterparts. Regarding parental assistance with schoolwork, at age 12 about 22% of Vietnamese 15-year-olds were assisted by their parents (as reported by the child's caregiver), which is a higher rate than parents in Ethiopia (14%) and India (16%), but lower than parents in

[7] That Vietnam's wealth index is similar to Peru's seems to contradict the GDP per capita figures in Table 1. Yet when GDP is adjusted for purchasing power parity, Peru's GDP per capita is only 92% higher than Vietnam's, not three times higher. Recall also that the Peru sample excluded the wealthiest 5% of Peru's districts (see Section II).

Peru (35%). Yet by age 15 Vietnamese parents were less likely to help their children (4%) than those in the other three countries (10-15%). Overall, relative to Vietnamese parents, Peruvian parents seem better prepared, and more willing, to help their children with their schoolwork, while Ethiopian and Indian parents seem less prepared and help less when their children are age 12 (but more when their children are 15).

There is a general perception for East Asia as a whole, which includes Vietnam, that children spend longer hours in school, and spend more time studying outside of school. The next nine rows in Table 2 show that this is only partially correct. School-age children in Vietnam spend *less* time in school than children of the same age in the other three Young Lives countries, although the differences with Ethiopia are very small. This is true at ages 8, 12 and 15; for example, at age 12 the average Vietnamese child spent 5.4 hours per day in school, which is slightly lower than in Ethiopia (5.6 hours), lower than in Peru (6.1 hours) and much lower than in India (8.0 hours). On the other hand, Vietnamese children spend more time studying at home, between 2.6 and 2.9 hours per day from age 8 to age 15, than children in the other three countries (between 1.0 and 2.1 hours per day). Perhaps more importantly, Vietnamese parents spend much more on private tutors and tutoring classes than do parents in other countries: about 10 times more at age 8, 20 times more at age 12, and 30-40 times more at age 15.

A final parental variable is aspirations for their children's education. Another common perception is that East Asian parents value education more, and have higher aspirations for their children's education, than do parents elsewhere. The next two lines in Table 2 show, somewhat surprisingly, that in all four Young Lives countries most parents expected, when their children were five years old, that they would complete a university-level or other post-secondary education, For all four countries, this greatly exceeds actual post-secondary completion. While

79% of Vietnamese parents hope that their child will complete a university level or other post-secondary degree, Peruvian parents were even more likely (87%) to express this desire, and Ethiopian parents were not far behind (72%); only Indian parents had much lower aspirations (58%). Parents were also asked (when their child was five years old) whether they thought that their child would achieve this ambitious expectation. The parents in Ethiopia, India and Peru were very optimistic, with 89-91% opining that their child would obtain this goal. Vietnamese parents were also optimistic, but somewhat less so; 79% stated that they thought that their child could attain university or post-secondary schooling.

Perhaps the strong academic performance of Vietnamese 15-year-olds is due to their going to better schools. Table 3 examines this. In most respects, Vietnamese primary schools appear to be of higher quality. In particular: 1. Almost all (95%) Vietnamese primary school teachers have a general (non-education) university degree, compared to 5% for Ethiopia, 79% for India and 84% for Peru; 2. Vietnamese primary school teachers have more years of experience (17) than their counterparts in Ethiopia (11) and India (7), although not more than Peruvian teachers (also 17); 3. Vietnamese primary school principals have 10 years of experience, on average, which is less than in Peru (13) but much higher than in Ethiopia (4) and India (6); 4. Reported teacher absence in primary schools is lowest in Vietnam (but data on teacher absence was collected differently in the four countries);[8] 5. Almost all Vietnamese primary schools have electricity (96%), which is higher than in the other three countries (53% for Ethiopia, 86% for India and 94% for Peru); 6. Vietnamese primary schools are more likely to have a library, which is the case for 74% of primary schools in Vietnam, but only 62% in Ethiopia, 21% in India, and

[8] Teacher absence is measured differently in the four countries. In Ethiopia and Vietnam, it was obtained from school records. In India, the school principal was asked how many days one or more teachers were absent in the last 30 days; the teacher level absence rate is calculated based on the median school having six teachers and the assumption that teacher absences were uncorrelated across teachers. In Peru, absence is based on teacher self-reports.

44% in Peru; and 7. Vietnam's primary schools are more likely to have computers for students' use (34%) than schools in Ethiopia (20%) and India (30%), but less likely than in Peru (58%).

Finally, in India, Peru and Vietnam (but not Ethiopia), primary school mathematics teachers took a test that measured their pedagogical skills in mathematics. The tests varied over the three countries, but for each of the three pairs of countries there are between 4 and 9 common test items. Figure 4 provides an example test item. The last primary school variable in Table 3 is constructed by applying IRT analysis to *all* the teacher mathematics pedagogical skill questions for the three countries (India, Peru and Vietnam) that administered this exam to their math teachers. The latent mathematical pedagogical ability of Vietnamese primary school math teachers is about 0.71σ higher than that of their counterparts in India, and an astonishing 1.52σ higher than that of their counterparts in Peru.

The remaining rows in Table 3 compare (upper) secondary schools in Vietnam with secondary schools in the other countries. In some (but not all) respects, Vietnam's secondary schools appear to be better: 1. The school principals have more years of experience at the current school (6.4 years, vs. 3.0 to 5.2 years for the other countries); 2. Almost 100% have electricity, which is also the case for Peru but higher than for Ethiopia (73%) and India (70%); 3. Almost all (92%) have a library, unlike in Ethiopia (75%), India (80%) and Peru (65%); 4. They have more computers per student (0.08) than in Ethiopia (0.01) and India (0.03);[9] 5. They are more likely to have an internet connection (97%) than secondary schools in Ethiopia (19%), India (53%) and Peru (86%); 6. Teachers have more experience (14.0) than in Ethiopia (9.6) or India (11.8), though less than in Peru (16.9); and 7. Almost all teachers have a bachelor's teacher training degree (97%), unlike in Ethiopia (40%), India (82%) and Peru (25%). Yet in other respects the

[9] This variable, and several others in Table 3, were not collected in the Peru secondary school survey.

secondary schools in the other three countries seem better than those in Vietnam: 1. About 35% of Vietnam's secondary school principals have master's degrees, which is higher than in Ethiopia (0%) but lower than in India (71%) and Peru (48%); 2. Average class size (39) is much higher than in India (30) or Peru (25), but not Ethiopia (42); 3. Minutes per week spent on math classes (182) is lower than in Ethiopia (210) and much lower than in India (325); 4. Fewer teachers have a general (non-education) master's degree (23%) than in India (61%) and Peru (26%), but more likely than in Ethiopia (0%); and 5. Students are less likely to report that their math teacher checks their homework (1.17 on a scale from 0 to 2) than in India (1.39) and Peru (1.21).

Overall, the data in Tables 2 and 3 show that Vietnamese 15-year-olds have several advantages over 15-year-olds in the other three Young Lives countries that could lead to higher learning outcomes, such as better nutritional status, fewer siblings, greater wealth, and (except for Peru) better educated parents. They also spend more hours studying at home, and their parents spend much more on private tutoring. In most respects, the primary and secondary schools that they attend appear to be better, including primary school math teachers with better pedagogical skills. On the other hand, they are at a disadvantage in terms of parental assistance with homework at age 15, length of the school day, and parental confidence that they will complete university or other post-secondary education, and in several respects their schools appear worse than those in the other countries. The obvious question is: which of these child, parent and school characteristics have explanatory power for the learning of 15-year-olds in these four countries? This is examined in the next section.

IV. What Observed Variables Explain the Differences across the Young Lives Countries?

In theory, the differences in math scores across the four Young Lives countries are due to differences across those countries in the causal factors that determine student learning. To the extent that those causal factors are found in the Young Lives data, they can be used to explain the math score differences across these four countries. This section explains how to do this, and presents estimates showing the extent to which the Young Lives data explain these differences.

A. Regression Methodology. Multiple regression analysis is very useful for conducting the analysis proposed above. To start, the following simple regression replicates the differences in the mean of the (normalized) mathematics test scores across the four countries:[10]

$$\text{Test Score}_i = \beta_1\text{Ethiopia}_i + \beta_2\text{India}_i + \beta_3\text{Peru}_i + \beta_4\text{Vietnam}_i + u_i \qquad (1)$$

where Ethiopia_i is a dummy variable equal to 1 for Ethiopian 15-year-olds (and is 0 otherwise), and India_i, Peru_i and Vietnam_i are analogously defined. Note that there is no constant term in equation (1); this ensures that the OLS estimate of the β terms corresponding to the four dummy variables are exactly equal to the standardized scores shown in Table 1. Thus, this regression is not an estimate of any causal relationship; it simply provides the mean values of the dependent variable (the math test score) for each of the four countries.

In principle, there exists a causal process that determines test scores in all four countries. This can be expressed as:

$$\text{Test Score}_i = \boldsymbol{\beta}'\mathbf{x}_i + u_i \qquad (2)$$

where $\mathbf{x}_i$ includes *all* variables that determine test scores, which implies that u_i is simply random measurement error in the test scores, so there is little reason to expect u_i to be correlated with the

[10] This regression can also replicate differences in non-normalized math scores; we use normalized scores for ease of interpretation.

x variables. This linear approximation is not particularly restrictive as long as the **x** variables include interactions between combinations of **x** variables (including squares and other powers of the **x** variables). If one had accurate data for all variables that have a causal impact on test scores, one could estimate equation (2) by ordinary least squares (OLS) and obtain the causal impacts of all of those variables. Yet it is virtually impossible to obtain data on all causal factors, so OLS estimates of equation (2) that use only observed variables could produce biased estimates of the causal impacts of the (observed) **x** variables (biased estimates of $\boldsymbol{\beta}$ in equation (2)) because all unobserved variables are relegated to the error term, and they are likely to be correlated with at least some observed variables. On average, however, the more **x** variables in the regression (the more causal factors that are observed) the less bias in the OLS estimates.

The $\boldsymbol{\beta}$ parameters in equation (2) are assumed to be the same for all four countries. One could challenge this assumption; for example, the impact of a longer school day could be smaller in countries with less effective schools. This assumption will be relaxed in Section V. Yet, in theory, such differential impacts can be accounted for in equation (2) by, for example, including interaction terms between the length of the school day and indicators of school or teacher quality. Thus, the assumption that the $\boldsymbol{\beta}$'s are the same for all four countries may be reasonable.

The main problem when estimating equation (2) is that some variables that determine students' test scores are not in the data. This will now be discussed in detail. If $\boldsymbol{\beta}$ is the same for all four countries, equation (2) can be written as:

$$\text{Test Score}_i = \boldsymbol{\beta}'\mathbf{x}_{iE} + \boldsymbol{\beta}'\mathbf{x}_{iI} + \boldsymbol{\beta}'\mathbf{x}_{iP} + \boldsymbol{\beta}'\mathbf{x}_{iV} + u_i \qquad (3)$$

where, using the same country dummy variables used in equation (1), $\mathbf{x}_{iE} = \mathbf{x}_i \times \text{Ethiopia}_i$, $\mathbf{x}_{iI} = \mathbf{x}_i \times \text{India}_i$, $\mathbf{x}_{iP} = \mathbf{x}_i \times \text{Peru}_i$, and $\mathbf{x}_{iV} = \mathbf{x}_i \times \text{Vietnam}_i$.

To see the usefulness of equation (3), consider the term for Ethiopia. It can be written as:

$$\begin{aligned} \boldsymbol{\beta}'\mathbf{x}_{iE} &= \boldsymbol{\beta}_o'\mathbf{x}_{iE,o} + \boldsymbol{\beta}_u'\mathbf{x}_{iE,u} \qquad (4) \\ &= \boldsymbol{\beta}_o'\mathbf{x}_{iE,o} + \boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u} + \boldsymbol{\beta}_u'(\mathbf{x}_{iE,u} - \bar{\mathbf{x}}_{E,u}) \end{aligned}$$

The first line of equation (4) divides the **x** variables for Ethiopia into two sets, those that are observed, denoted by $\mathbf{x}_{iE,o}$, and those that are unobserved, denoted by $\mathbf{x}_{iE,u}$. The $\boldsymbol{\beta}$ vector is similarly divided into $\boldsymbol{\beta}_o$ for the observed variables and $\boldsymbol{\beta}_u$ for the unobserved variables.

The second line in equation (4) divides the unobserved variables into two parts, their means for Ethiopia, multiplied by their associated $\boldsymbol{\beta}$'s, and the deviation of those variables from these means. The first part, $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$, does not vary over observations from Ethiopia and thus it is a dummy variable for Ethiopia. More precisely, it can be replaced by the dummy variable for Ethiopia, Ethiopia_i, the coefficient of which equals $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$.

The $\boldsymbol{\beta}'\mathbf{x}_{iE}$ term in equation (4) has two "extreme" cases. First, assume that all variables are observed; then $\boldsymbol{\beta}'\mathbf{x}_{iE}$ equals $\boldsymbol{\beta}_o'\mathbf{x}_{iE,o}$, and equation (3), or equivalently equation (2), includes all causal variables, so OLS regressions produce unbiased estimates of $\boldsymbol{\beta}$, and a dummy variable for Ethiopia or any other country would be statistically insignificant. Second, if no variables are observed, then $\boldsymbol{\beta}'\mathbf{x}_{iE}$ becomes the sum of: 1. A dummy variable for Ethiopia multiplied by the coefficient on that dummy variable, which is equal to $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$; and 2. The (unobserved) variation around the means for each variable, $\boldsymbol{\beta}_u'(\mathbf{x}_{iE,u} - \bar{\mathbf{x}}_{E,u})$, and those deviations are uncorrelated with the dummy variable because the dummy variable does not vary.

The typical case is between these two extremes: some variables are observed and others are not. In general, adding more observed variables to the regression has two effects. First, the means of those variables are implicitly moved from $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$ to $\boldsymbol{\beta}_o'\mathbf{x}_{iE,o}$, the latter of which accounts for both the means of those variables and their variation within Ethiopia. In general, this will reduce the size of $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$, the coefficient associated with the Ethiopia dummy variable, so adding observed variables to equation (1), which yields equation (4), reduces the estimated coefficients for the country dummy variables. Intuitively, as more observed variables are added they will explain more of the differences in test scores across the four countries, and the dummy variables will explain less, reducing their coefficients.[11] Second, this removal of unobserved variables from the regression removes from the residual factors that may be correlated with the observed variables, reducing the correlation of the error term with the observed variables and thus reducing bias in OLS estimates of equation (4).

It is possible that moving a given variable from unobserved to observed will increase, rather than reduce, the coefficient on the Ethiopia dummy variable. For example, teacher absence is likely to lead to lower student learning, so the β associated with that variable will be negative, and thus removing the associated $\beta_u'\bar{x}_{E,u}$ for teacher absence will increase $\boldsymbol{\beta}_u'\bar{\mathbf{x}}_{E,u}$. Yet such an increase is not a problem because the comparison of interest is the *relative* size, not the absolute size, of the dummy variables for each country. In this example, removing the associated $\beta_u'\bar{x}_{E,u}$ for teacher absence would increase all four dummy variables if β_u is negative, but it would raise it the least for Vietnam since its rate of teacher absence is the lowest, reducing the gap between the Vietnam dummy variable and the dummy variables for the other three countries, so that more of the gap is explained by observed variables and less is explained by unobserved variables.

[11] Fryer and Levitt (2004) use this approach to study the test score gap between black and white students in the U.S.

The above discussion leads to the following regression equation:

$$\text{Test Score}_i = \boldsymbol{\beta}_o'\mathbf{x}_{i,o} + \beta_1\text{Ethiopia}_i + \beta_2\text{India}_i + \beta_3\text{Peru}_i + \beta_4\text{Vietnam}_i + u_i \qquad (5)$$

where u_i represents both measurement error in the test score variable and variation within each of the four countries in the unobserved determinants of test scores. The goal of estimating equation (5) is to investigate the extent to which the (relative sizes of the) dummy variables for the four countries decrease as more observed variables are added to the regression equation. More intuitively, the goal is to see how much of Vietnam's exceptional performance in education can be explained by the observed variables in the Young Lives data.

B. Results. Tables 4 and 5 show estimates of equation (5), starting with only the country dummy variables and then adding child and household level variables. Table 4 does this when the normalized score on the 23 mathematics questions is used as the dependent variable, while Table 5 uses the latent mathematics skills of each student obtained from the IRT analysis. For both tables, the first column simply shows that OLS estimation of equation (1) reproduces the (normalized) mean test scores shown in Table 1. That is, the first column in Table 4 shows that Vietnamese 15-year-olds score 1.41σ higher than 15-year-olds in Ethiopia, 1.12σ higher than those in India, and 0.67σ higher than those in Peru. Similarly, the first column in Table 5 shows that Vietnamese 15-year-olds score 1.35σ higher than those in Ethiopia, 0.97σ higher than those in India, and 0.73σ higher than those in Peru. These gaps, shown in brackets immediately below the parameter estimates in both tables, are very large; explaining them is the goal of this paper.

Recall that Vietnamese 15-year-olds are less likely to be stunted, come from wealthier homes (as measured by a wealth index), have fewer siblings, and (except for Peru) have more educated parents than 15-year-olds in the three other countries. The second column of estimates in Tables 4 and 5 adds those child and household characteristics as explanatory variables in equation (5). As expected, 15-year-olds from wealthier households, with more educated mothers, and with better nutrition (as indicated by higher height-for-age Z-scores) have higher test scores, and 15-year-olds with more siblings have lower test scores. More interesting for the purposes of this paper is that the gaps between the Vietnam dummy variable and the Ethiopia and India dummy variables have decreased: in Table 4 (Table 5) the gap between Ethiopia and Vietnam dropped from 1.41σ to 0.92σ (from 1.35σ to 0.84σ), and the gap between India and Vietnam dropped from 1.12σ to 0.85σ (from 0.97σ to 0.70σ). In contrast, the gap between Peru and Vietnam increased slightly, from 0.67σ to 0.68σ (from 0.73σ to 0.74σ). Thus, these variables alone "explain" about a third of the gap between Ethiopia and Vietnam and about a fourth of the gap between India and Vietnam, but they explain none of the gap between Peru and Vietnam.

The last column of estimates in Tables 4 and 5 examines whether hours spent in school, hours studying at home, spending on private tutoring, and parental aspirations can further explain the gap in math test scores between Vietnam and the other three countries. As expected, at all ages (8, 12 and 15) all of these variables have positive predictive power for test scores, and most are statistically significant. Regarding the gaps, adding these variables has little effect on the gap between Vietnam and Ethiopia, increases the gap between Vietnam and India, and slightly reduces the gap between Vietnam and Peru. More specifically, in Table 4 the gap relative to Ethiopia falls slightly, from 0.92σ to 0.88σ, while the gap relative to India increases from 0.85σ to 1.02σ (but still below the unconditional gap in column 1 of 1.12σ), and the gap relative to Peru drops by small amount, from 0.68σ to 0.59σ. The results in Table 5 are similar. The small reduction for Ethiopia reflects that, as seen in Table 2, hours in school are similar for Ethiopia and Vietnam while Vietnamese students spend more time studying at home and more time in private tutoring, and parental aspirations are somewhat higher in Vietnam, so these variables explain a little more of the gap. The increase in the gap for India likely reflects that time in school is much higher in India than in Vietnam, as seen in Table 2, which makes the gap even harder to explain; those impacts outweigh the advantages that Vietnamese students have in time studying at home, spending on tutoring and parental aspirations. Also, adding these educational input variables to the regression reduces the estimated impacts of wealth, mother's education, number of siblings and nutritional status, all of which are areas where Vietnam has an advantage over India, which reduces their power to explain the gap between these two countries. Finally, the small reduction in the gap for Peru likely reflects Vietnamese students' advantages with respect to hours studying at home and spending on tutoring; unlike India, the reduced effects of wealth and parental education are not areas where Vietnam has an advantage over Peru.

Tables 7-10 examine whether the primary and secondary school variables can further explain the gaps between Vietnam and the other three countries. However, before turning to those results some discussion is needed concerning several complications with the school data. First, matching the Young Lives children to their primary schools was difficult, especially for Ethiopia, India and Peru, as explained in Appendix A. For Ethiopia and India, only about two thirds of the younger cohort children with primary school codes could be matched to a primary school from which data were collected, so only 1,156 were matched in Ethiopia and 1,158 were matched for India. This problem is even worse in Peru because, due to budget and logistical constraints, primary school data were collected from only 14 of the 20 sites, and not every school in these 14 sites was sampled. In the end, only 642 younger cohort children in Peru could be matched to a primary school from which data were collected. Only in Vietnam can more than

80% of the younger cohort children be matched to a primary school from which data were collected; 1,662 children were matched for that country.

A second problem with the primary school data is measurement error, which can take two forms. First, the information reported may not be precisely equal to the relevant conditions at the school. Students were in those primary schools for five to seven years, depending on the country, and school characteristics in one year and could be different in other years. In addition, simple errors may have occurred in respondents' answers and in the recording of their answers by the survey teams. Second, in three countries (all but Peru) there were substantial discrepancies between the household survey data and the school data regarding the school in which the child was actually enrolled, as explained in Appendix A.

A third problem is that the secondary school data are not matched to individual children in the Young Lives data. Indeed, many of them did not enroll in secondary school. Also, the secondary school data could also suffer from measurement error.

One way to match almost all younger cohort children to both the primary school data and secondary school data is to calculate site-level means of primary and secondary school variables, for all sites in each country.[12] Yet this may lead to attenuation bias because these means measure with error the characteristics of the school the child attended; if these errors were random this attenuation bias would lead to underestimation of school characteristics' impacts. In fact, this measurement error is not random. As Appendix B shows, under plausible assumptions of the within-site variation of school characteristics, using site-level means does not lead to attenuation bias if the school data are measured without error; if they are measured with error then regression estimates using site-level means have *less* attenuation bias than estimates using school-level data.

[12] This can be done only for 14 sites in Peru for the primary school data and 14 sites for the secondary school data, and only 12 sites have both primary and secondary school data.

The result that regressions using site-level means suffer less from measurement error yields a test of the existence of measurement error. If the primary school variables do not have measurement error, estimates using site-level data should be similar to those using school-level data, since neither would suffer from attenuation bias. Yet if measurement error is present, regressions using site-level means should have larger (in absolute value) coefficients than those from regressions using on school-level data. This is investigated in Table 6 by adding almost all primary and secondary school variables to the regressions in the last column of Tables 4 and 5.[13]

The first two columns in Table 6 use the normalized score on the 23 math questions as the dependent variable, while the third and fourth columns use the latent mathematics skills obtained from the IRT analysis. In all four columns, include only those children who could be matched to a primary school, so that the comparisons using site-level means and school-level data are on the same sample. Turning to the first two columns, most primary school variables are statistically insignificant, yet the four with some statistical significance show the same pattern: the coefficients when the primary school variables are site-level means are three to five times larger than those when the variables are measured at the school level. This same pattern is found when the latent value from the IRT analysis is used. These results indicate that there is substantial measurement error in the school variables, so one should use site-level means for the primary school variables to minimize attenuation bias (and to maximize the sample, since almost all students can be matched to site level means). Thus, all remaining regressions use site-level means for the primary school variables (site-level means must also be used for the secondary school variables, since those schools cannot be matched to the younger cohort children).

[13] A few of these school variables were excluded because of their low explanatory power, as explained below.

Can the school and teacher characteristics in both primary and secondary schools further explain the test score gaps between 15-year-olds in Vietnam and 15-year-olds in the three other Young Lives countries? As seen in Table 3, 11 primary school and 14 secondary school variables are available for all four countries. To select variables with explanatory power, preliminary regressions were run using these 25 of school variables (and all the child and household variables used in the last columns of Tables 4 and 5). Two regressions were estimated, one using the normalized math score based on the 23 common questions, and the other using the latent math skill variable generated by the IRT analysis, as the dependent variable. School variables with t-statistic below one in both regressions are not used in the subsequent analysis. This removed from the analysis three primary school variables (teacher has a general university degree, school has electricity, and using tracking/streaming to assign students to classrooms) and four secondary school variables (school principal has a master's degree in education, school has an internet connection, teacher has a bachelor's degree in education, and teacher is a contract teacher). Thus the regression analysis uses eight primary school variables and ten secondary school variables.

The results when these primary and secondary school site-level means are added as explanatory variables are shown in Tables 7 (normalized math score) and 8 (latent math skill). The first four columns reproduce the first two columns and last two columns of Table 4 (for Table 7) and Table 5 (for Table 8), although the sample sizes are smaller because observations without school variables (more specifically, 8 of the 20 sites in Peru) are excluded. Yet the overall findings from Tables 4 and 5 using these slightly small samples are similar; for Table 7 household variables can explain a little more than one third of the gap between Vietnam and Ethiopia, about one tenth of the gap between Vietnam and India, and none of the gap between

Vietnam and Peru.[14] Similar results hold for Table 8, except that – as in Table 5 – the household variables explain only about 2% of the gap between Vietnam and India.

The last two columns in Tables 7 and 8 add the eight primary school and ten secondary school site-level mean variables that offered at least some explanatory power in preliminary regressions. Seven of these eighteen school variables are significant at the 5% level in both tables, and all but one have the expected sign; the sole exception is that female primary school teachers reduce students' mathematics skills. The other three robustly significant primary school variables have the expected signs: principal years of experience and libraries lead to higher math scores, and higher rates of teacher absence reduce scores. Turning to the secondary school variables, the three robustly significant variables indicate that principals who help teachers, teachers who help students, and teachers with a master's degree all increase student test scores.

Do these school variables help explain the gap in test scores between Vietnam and these other three countries? For Ethiopia, adding the school variables decreases the gap by a modest amount. The child and household variables alone decreased the gap in Table 7 (Table 8) from 1.413 to 0.888 (1.358 to 0.827), explaining 37% (39%) of the gap, and adding the school variables decreased it to 0.843 (0.767), increasing the explained gap to 40% (44%).

Turning to India, the child and household variables alone decreased the gap only modestly in Table 7, from 1.116 to 1.016, explaining just 9% of the gap, and adding the school variables increased it to 1.096, so that only 2% of the gap is explained. The results in Table 8 fail to explain any of the gap between Vietnam and India; adding child and household variables slightly decreases the gap from 0.968 to 0.950, and adding the school variables increases it

[14] There are some relatively small differences, in Tables 7 and 8 relative to Tables 4 and 5, in how much of the gap between Vietnam and Peru can be explained by the child and household variables. This primarily reflects a different sample for Peru; the results in Tables 4 and 5 use data from all 20 sites in Peru, while the results in Tables 7 and 8 use only the 12 sites in Peru that were included in the data collection for primary schools and secondary schools.

substantially, to 1.068. To see why the school variables do not explain the gap between India and Vietnam, consider their means in Table 3. India compares favorably to Vietnam among four of the eight highly significant school variables in Tables 7 and 8; it has fewer female primary school teachers, more secondary teachers with master's degrees, and higher reports of principals helping teachers and teachers helping students.

Finally, turn to the comparison of Vietnam and Peru. Adding the child and household variables alone increases the gap in Table 7 (Table 8) from 0.555 to 0.562 (0.613 to 0.653). In Table 7, adding the school variables explains a small part of the gap, decreasing it to 0.507, but in Table 8 adding those variables has almost no effect, decreasing the gap very slightly to 0.645.

The results thus far are puzzling; even though Vietnam's primary and secondary schools appear better in many respects than those in Ethiopia, India and Peru, adding those variables to the regression analysis does little to explain why Vietnamese 15-year-olds perform much better than do those in the other three countries. The school variables do explain a small proportion (3-4%) of the gap between Vietnam and Ethiopia, but the fail to explain the gap between Vietnam and India and the gap between Vietnam and Peru.

One primary school variable in Table 3 was not used in Tables 7 and 8 because it was not collected in Ethiopia: teachers' latent pedagogical ability. Yet it can be used if Ethiopia is excluded from the sample and Vietnam is compared to only India and Peru. This is done in Tables 9 and 10, which are analogous to Tables 7 and 8 except that Ethiopia is no longer in the sample. The first thing to note is that in both Tables 9 and 10, math teachers' latent pedagogical ability has strong explanatory power in the expected positive direction. Comparing Vietnam with India, Table 9 (Table 10) shows that the gap between the two countries drops from 1.116 to 1.016 (0.968 to 0.919), a drop of 9% (5%), when the child and household variables are added, which is similar to the results in Table 7 and 8. More interesting is that when the school variables are added, including math teachers' pedagogical skills, the gap declines further, to 0.887 (0.829), so that together the child, household and school variables explain 21% (14%) of the gap between Vietnam and India. Overall, teachers' pedagogical skills have some explanatory power, but it is still the case that the Young Lives data can explain at most only one fifth of the gap in test scores between Vietnam and India.

In contrast, adding primary school math teachers' pedagogical skills has very strong explanatory power for explaining the gap between students' math performance between Vietnam and Peru. In Table 9 (Table 10), the gap between the two countries increases slightly, from 0.555 to 0.586, (from 0.613 to 0.659) when the child and household variables are added. But when the school variables, including primary school math teachers' pedagogical skills, are added the gap falls dramatically, to only 0.091 (0.214), so that together the child, household and school variables explain 84% (65%) of the gap between Vietnam and India.

To summarize the results of this section, the Young Lives data appear to explain 40-44% of the very large gap in mathematics skills between 15-year-olds in Vietnam and Ethiopia. Almost all of this gap is due to differences in child and household characteristics, and school characteristics explain very little (3-4%). Yet this lack of explanatory power for school characteristics may reflect that no data were collected on Ethiopian primary school math teachers' pedagogical skills; if such data were available, school variables may be able to explain a much larger proportion of the gap. In contrast, the Young Lives data explain only 14-21% of the large gap in the math scores of 15-year-olds between India and Vietnam, although it is interesting that most of this explanatory power is from Vietnamese primary school teachers' higher mathematics pedagogy skills. Finally, the Young Lives data appear to explain most of the gap in math skills between 15-year-olds in Vietnam and Peru, and almost all of the explanatory power reflects the very large difference in primary school math teachers' pedagogical skills in those two countries.

V. Oaxaca-Blinder Decompositions

The methodology in Section IV assumes that the impacts of both observed and unobserved variables are the same for all four countries. Yet this assumption may not hold. For example, the impact of years of parental education on a child's learning could vary across countries due to variation in school quality across countries. In theory this can be modeled by interacting years of parental schooling with indicators of school quality (for the years when the parents were in school), but such data may not be available. This section presents an analysis of the learning gaps between Vietnam and the three other Young Lives countries that allows the impacts of the observed variables to vary by country.

A. The Oaxaca-Blinder Decomposition. A more flexible approach to analyze the learning gaps between Vietnam and the other three Young Lives countries is to use the Oaxaca-Blinder decomposition. This can be done separately for each of the other Young Lives countries. To see how this is done, consider a comparison of Vietnam with Ethiopia, and allow the impacts of the observed variables on test scores to vary by country:

$$S_{i,vn} = \boldsymbol{\beta}_{vn}'\mathbf{x}_{i,vn} + u_{i,vn} \quad \text{(Vietnam)} \qquad (6)$$

$$S_{i,e} = \boldsymbol{\beta}_{e}'\mathbf{x}_{i,e} + u_{i,e} \quad \text{(Ethiopia)} \qquad (7)$$

Taking the mean of both sides of these two regression equations gives the following:

$$\bar{S}_{vn} = \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} \qquad (8)$$

$$\bar{S}_{e} = \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{e} \qquad (9)$$

The standard Oaxaca-Blinder decomposition uses these two equations to express the difference in the mean test scores between Vietnam and Ethiopia in the Young Lives data as follows: [15]

$$\bar{S}_{vn} - \bar{S}_{e} = \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{e} \qquad (10)$$

$$= \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{e} + \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{vn}$$

$$= \boldsymbol{\beta}_{e}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}) + (\boldsymbol{\beta}_{vn} - \boldsymbol{\beta}_{e})'\bar{\mathbf{x}}_{vn}$$

One critique of this decomposition is that the differences in ($\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}$) are "weighted" by the coefficients for Ethiopia, and the differences in ($\boldsymbol{\beta}_{vn} - \boldsymbol{\beta}_{3}$) are weighted by the means of Vietnam's **x** variables. Intuitively, both sets of "weights" should be based on data from both countries.

This shortcoming can be avoided by using the following decomposition:

$$\bar{S}_{vn} - \bar{S}_{e} = \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{e} \qquad (11)$$

$$= \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_{e}'\bar{\mathbf{x}}_{e} + \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}) - \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e})$$

$$= \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}) + [(\boldsymbol{\beta}_{vn} - \bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_{e})'\bar{\mathbf{x}}_{e}]$$

[15] An alternative decomposition is: $\bar{S}_{vn} - \bar{S}_{e} = \boldsymbol{\beta}_{vn}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}) + (\boldsymbol{\beta}_{vn} - \boldsymbol{\beta}_{e})'\bar{\mathbf{x}}_{e}$. This decomposition suffers from the same criticisms, discussed below, as the decomposition in equation (10).

where $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta}_{vn} + \boldsymbol{\beta}_e)/2$.[16] Intuitively, the first term weights the differences in the $\mathbf{x}$ variables by the simple average of the two $\boldsymbol{\beta}$ vectors, and the second splits the difference between $\boldsymbol{\beta}_{vn}$ and $\boldsymbol{\beta}_e$ into the difference between $\boldsymbol{\beta}_{vn}$ and $\bar{\boldsymbol{\beta}}$, weighted by $\bar{\mathbf{x}}_{vn}$ and that between $\boldsymbol{\beta}_e$ and $\bar{\boldsymbol{\beta}}$ weighted by $\bar{\mathbf{x}}_e$.

In principle, equation (11) can be used to decompose the impacts not only of child and household variables but also ofprimary and secondary school characteristics. Unfortunately, it cannot be applied to school variables' site-level means because each country has only 20 sites, yet the data contain site-level means for 14 primary school and 23 secondary school variables. Thus, in effect there are only 19 degrees of freedom (taking the constant term into consideration) to estimate the effect of these 37 school variables. Moreover, given that this decomposition uses separate regressions for each country, and the standard errors are clustered at the site level, the wild bootstrap is needed to obtain correct p-values given that there are at most only 20 clusters for each country; see Cameron, Gelbach and Miller (2008). Preliminary estimates that used a small number of school variables and the wild bootstrap to obtain accurate p-values yielded statistically insignificant estimates of those school variables' impacts.[17]

Yet the Oaxaca-Blinder decomposition can still be applied to the child and household level variables, using site-level fixed effects to control for variation in school characteristics across the sites. To see how this works, consider equations (6) and (7) for Ethiopia and Vietnam, separating the child and household variables from the school variables:

$$S_{i,vn} = \boldsymbol{\beta}_{vn}'\mathbf{x}_{i,vn} + \boldsymbol{\gamma}_{vn}'\mathbf{s}_{c,vn} + u_{i,vn} \quad \text{(Vietnam)} \quad (6')$$

[16] Equation (11) holds for *any* definition of $\bar{\boldsymbol{\beta}}$; the one used here is the most "natural" for decomposition purposes.
[17] Country-level regressions were estimated using only the three variables that were consistently significant at the 1% level in Tables 7-10: teacher absence at the primary school level, teacher helps students at the secondary school level, and primary school math teachers' pedagogical skills (not available for Ethiopia). The bootstrapped p-values for primary school teacher absence exceeded 0.10 for all but one country: the exception was 0.08 for India. Those for secondary-level teacher helps students exceeded 0.10 for all but one country: the exception was 0.04 for Peru. Those for primary math teachers' pedagogical skills exceeded 0.01 for all three countries with data for that variable.

$$S_{i,e} = \boldsymbol{\beta}_e'\mathbf{x}_{i,e} + \boldsymbol{\gamma}_e'\mathbf{s}_{c,e} + u_{i,e} \quad \text{(Ethiopia)} \qquad (7')$$

In these regressions, the $\mathbf{x}_i$ variables are now the child and household variables only, and the $\mathbf{s}_c$ variables are the school variables, where the c subscript indicates community (i.e. site) level means. For each community in these equations, the impacts of the school variables can be summarized by a community (site) level fixed effect, so one can rewrite these equations as:

$$S_{i,vn} = \boldsymbol{\beta}_{vn}'\mathbf{x}_{i,vn} + \gamma_{1,vn}d_{1,vn} + \ldots + \gamma_{20,vn}d_{20,vn} + u_{i,vn} \quad \text{(Vietnam)} \qquad (6'')$$

$$S_{i,e} = \boldsymbol{\beta}_e'\mathbf{x}_{i,e} + \gamma_{1,e}d_{1,e} + \ldots + \gamma_{20,e}d_{20,e} + u_{i,e} \quad \text{(Ethiopia)} \qquad (7'')$$

where the γ parameters are the community fixed effects for the 20 sites in each country, and the d variables are the corresponding dummy variables for each community.

The Oaxaca-Blinder decomposition can be applied to equations (6″) and (7″) only for the $\boldsymbol{\beta}$ coefficients and the $\mathbf{x}$ variables, since the site dummy variables are not comparable across countries. To do this, define the overall impact of the school variables in each country as:

$$S_{i,vn} = \boldsymbol{\beta}_{vn}'\mathbf{x}_{i,vn} + \boldsymbol{\gamma}_{c,vn} + u_{i,vn} \quad \text{(Vietnam)} \qquad (6''')$$

$$S_{i,e} = \boldsymbol{\beta}_e'\mathbf{x}_{i,e} + \boldsymbol{\gamma}_{c,e} + u_{i,e} \quad \text{(Ethiopia)} \qquad (7''')$$

where $\boldsymbol{\gamma}_{c,vn} = \gamma_{1,vn}d_{1,vn} + \ldots + \gamma_{20,vn}d_{20,vn}$ and $\boldsymbol{\gamma}_{c,e} = \gamma_{1,e}d_{1,e} + \ldots + \gamma_{20,e}d_{20,e}$.

Taking the means of both sides of these two equations yields:

$$\overline{S}_{vn} = \boldsymbol{\beta}_{vn}'\overline{\mathbf{x}}_{vn} + \overline{\boldsymbol{\gamma}}_{vn} \qquad (8')$$

$$\bar{S}_e = \boldsymbol{\beta}_e'\bar{\mathbf{x}}_e + \bar{\boldsymbol{\gamma}}_e \qquad (9')$$

The $\bar{\boldsymbol{\gamma}}$ terms are essentially constant terms in equations (8′) and (9′) and they represent the average impact of all (observed and unobserved) school characteristics in each country. They also represent the (mean) impacts of any unobserved child and household characteristics.[18]

The standard Oaxaca-Blinder decomposition can then be expressed as:

$$\bar{S}_{vn} - \bar{S}_e = \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_e'\bar{\mathbf{x}}_e + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e) \quad (10')$$

$$= \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_e'\bar{\mathbf{x}}_e + \boldsymbol{\beta}_e'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_e'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e)$$

$$= \boldsymbol{\beta}_e'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_e) + (\boldsymbol{\beta}_{vn} - \boldsymbol{\beta}_e)'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e)$$

The first term measures the contribution of differences in (observed) child and household characteristics to the test score gap between the two countries, and the second measures the contribution of the differences in the impacts of those variables. The $(\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e)$ term measures the impact of the differences in the school variables, both differences in the school characteristics and any differences in the impacts of those characteristics. It also measures differences in all *unobserved* child and household characteristics, differences both in their means and in their impacts.

This can also be extended to the more preferred decomposition, as follows:

$$\bar{S}_{vn} - \bar{S}_e = \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_e'\bar{\mathbf{x}}_e + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e) \quad (11')$$

$$= \boldsymbol{\beta}_{vn}'\bar{\mathbf{x}}_{vn} - \boldsymbol{\beta}_e'\bar{\mathbf{x}}_e + \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_e) - \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_e) + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_e)$$

[18] The $\bar{\boldsymbol{\gamma}}$ terms could also represent the constant term for each country, but if all variables that affect test scores are observed, then they would explain all variation in test scores across the two countries and the constant term would be the same for both countries, and so would be differenced out in equation (10′).

$$= \bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn} - \bar{\mathbf{x}}_{e}) + [(\boldsymbol{\beta}_{vn} - \bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_{e})'\bar{\mathbf{x}}_{e}] + (\bar{\boldsymbol{\gamma}}_{vn} - \bar{\boldsymbol{\gamma}}_{e})$$

where $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta}_{vn} + \boldsymbol{\beta}_{e})/2$.

In summary, while the Oaxaca-Blinder decomposition cannot be applied to the individual school variables, one can apply it to the child and household variables. It can be used to calculate the overall (aggregate) contribution of (observed) child and household variables to the gap in test scores between Vietnam and the other three countries, as well as the overall contribution of the school variables (which includes any overall contribution from unobserved child and household variables) to that gap. The contributions of the child and household variables can be further decomposed into the contribution due to differences in their means and the contribution due to differences in their impacts, both for these variables as a whole and for each variable separately.

B. Results. The first country to compare with Vietnam using the Oaxaca-Blinder decomposition is Ethiopia. The results are shown in Table 11 (standardized score on 23 math questions) and Table 12 (latent math skill from IRT analysis). Turning to Table 11, there is a very large gap in the (normalized) test scores between the two countries: 1.41σ. The overall decomposition of this gap into the differences in the mean values of the observed variables, the differences in the $\boldsymbol{\beta}$ coefficients of those variables, and the differences in the school (and unobserved child and household) variables is shown in the last two lines of the table. The largest component is the difference in the school (and unobserved child and household) variables, which at 0.52σ accounts for a little more than a third (37%) of the overall test score gap between the two countries. The overall contribution from the differences in the means of the (observed) child and household variables across the two countries is 0.49σ, which is slightly more than one third

(35%) of the overall gap. Finally, the contribution of the differences in the estimated **β**'s of those variables is 0.40, which is a little more than one fourth (28%) of the overall gap.

The explanatory power of the differences in means and the differences in the estimated **β**'s of the observed child and household variables can be further decomposed into the contributions of each of these variables. These are shown in the last two columns of Table 11. The variables that contribute most to the overall contribution of the differences in the variable means (0.49σ) are differences in wealth (0.12σ), mother's schooling (0.07σ), hours studying at home at ages 12 and 15 (both contribute 0.06σ), and spending on private tutoring at age 15 (0.12σ). Regarding the contributions in the differences of the **β**'s, by far the largest effect is the greater "efficiency" of hours per day in school at age 15, which accounts for almost all (0.33σ out of 0.40σ) of this component. Unfortunately, it is difficult to explain what is behind this higher "efficiency"; to the extent that it reflects greater unobserved school quality it could also be considered part of the contribution of unobserved school quality.

The results using the latent math skill variable in Table 12 are similar, although there are some differences. The overall contribution from the differences in the means of the (observed) child and household variables across the two countries is 0.51σ, which is close to that in Table 11 and again explains a little more than a third (38%) of the overall gap. As in Table 11, the variables contributing the most to this component are the wealth index, mother's education, hours studying at home, and private tutoring costs at age 15. The biggest single contribution of the three aggregate components is again the difference in the school (and unobserved child and household) variables; at 0.91σ it accounts for two thirds (67%) of the overall gap in the test scores between the two countries. Finally, there is essentially no contribution of the differences in the estimated **β**'s of the observed child and household variables: the estimated effect of -0.07

is statistically insignificant. Although several individual components are statistically significant, they tend to cancel each other out; indeed, the largest such canceling is hours per day at ages 12 and 15, which is rather puzzling. One possible reason why hours per day are more effective at age 15 in Vietnam than at age 8 and 12 is that most of the younger cohort in Vietnam (54.8%) were in Grade 10, and about half of the others (22.5%) were in Grade 9. Vietnamese students take an exam at the end of Grade 9 that determines whether they can enroll in Grade 10, so students, parents and teachers likely exert extra effort during school hours in Grade 9. Moreover, students who are able to enroll in Grade 10 (about 70% of Grade 9 students) are "above average" students, and their (upper secondary) schools may be of higher quality than the primary and lower secondary schools (which Vietnamese children are enrolled in at ages 8 and 12, respectively) in which they were previously enrolled. Yet it is unclear why Ethiopian students' hours in school were very productive at age 12.

Turn next to Tables 13 and 14, the Oaxaca-Blinder decomposition results that compare Vietnam to India. The overall gap of 1.12σ in Table 12 is almost entirely accounted for by the differences in school (and unobserved child and household) variables. Indeed, that contribution equals 1.19σ, and the overall contributions from differences in means of the observed child and household variables and differences in those variables estimated **β**'s across the two countries are both statistically insignificant. While differences in the means of several observed variables are significant and appear to explain some of the gap, such as household wealth, mother's education and time spent studying at home, their effects are essentially reversed by India's longer school day. Regarding differences in the **β** coefficients, only two of the fourteen are statistically significant, and they appear to cancel each other out: the "efficiency" of an hour in school in India appears to be higher than an hour in school in Vietnam at age 12, but the opposite is the case at age 15. Again, the use of exams to admit 15-year-olds to Grade 10 in Vietnam may explain why hours in school are unusually effective at that age in Vietnam, but it is unclear why hours in school appear particularly productive in India at age 12. One benefit of the general lack of differences in the **β**'s is that it supports the assumption in Section IV of no such differences.

The results in Table 14 are similar, although there is a "negative" contribution of the differences in the **β**'s for explaining the gap; this primarily reflects "more efficient" hours per day in school in India than in Vietnam at age 12. Again, observed differences in the child and household variables do not, as a group, explain the gap in 15-year-olds' math skills between Vietnam and India, which implies that differences in schools (or possibly in unobserved child and household variables) are the most likely explanation for this gap. Given that the observed school variables did little to explain this gap in Section IV, perhaps other, unobserved school variables may be the explanation.

Tables 15 and 16 show the Oaxaca-Blinder decomposition results that compare Vietnam with Peru. The gap between the two countries in Table 15 is smaller than was the case for Ethiopia and India, only about 0.63σ for the (normalized) score based on the 23 questions used in all four countries. The overall decomposition shows that virtually all of this gap is due to differences in school (and possibly unobserved child and household) characteristics. Indeed, the contribution of 0.86σ exceeds somewhat the gap to be explained. In contrast, the overall contribution of the differences in the means of the observed variables is essentially zero (-0.03σ, and statistically insignificant). While Vietnamese students have advantages in terms of time spent studying at home and spending on private tutoring, Peruvian students seem to benefit from longer school days. Finally, turning to the contribution of the differences in the **β**'s of the observed variables, the overall contribution is a negative one of -0.20σ. This is primarily driven by the single significantly different β coefficient: hours in school at age 12 appear much more effective in Peru than in Vietnam. There are no other significant differences in the **β**'s for the observed variables between Peru and Vietnam, which again supports the assumption in Section IV that the impacts of these variables vary little across the four countries.

Finally, consider the decompositions for Peru and Vietnam based on the latent math skill obtained using IRT methods, as shown in Table 16. The results are very similar to those in Table 15. Again, the entire gap is due to differences in school (and possibly unobserved child and household) variables. While the Oaxaca-Blinder decomposition cannot uncover which variable it may be, the results in Section IV suggest that teachers' pedagogical skills may be a key factor.

VI. Conclusion

Vietnam's economic achievements since 1990 have attracted much attention, yet its accomplishments in education, especially its performance on the 2012 and 2015 PISA assess-ments, have also generated international interest. This paper investigates the underlying determinants of Vietnam's apparent exceptional performance using the Young Lives data collected from Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam.

While the Young Lives data are unusually rich, they cannot completely explain why Vietnamese 15-year-olds have much higher mathematics skills than 15-year-olds in Ethiopia, India and Peru. If one assumes that the impacts of the child, household and school variables are the same for all four countries, then the observed child, household and school variables explain 40-43% of the very wide gap (about 1.4 standard deviations) in the mathematics test scores between 15-year-olds in Vietnam and Ethiopia. Oaxaca-Blinder decompositions indicate that almost all of this reflects differences in observed child and household characteristics, in particular differences in household wealth, mother's education, hours studying at home, and spending on private tutors. Observed school variables explain only 3-4% of the gap. However, a key school variable, primary school mathematics teachers' pedagogical skills, was not collected in Ethiopia, and it may be that this can account for much of the gap that remains unexplained.[19]

The very wide (1.0-1.1 standard deviations) gap in mathematics test scores between 15-year-olds in Vietnam and India is more difficult to explain.[20] If one assumes that the impacts of the child, household and school variables are the same for all four countries, then the observed child and household variables explain at most only 4-9% of the gap. Adding observed school variables, especially primary school math teachers' pedagogical skills, increases the overall explained portion to 14-21% of the gap. Oaxaca-Blinder decompositions suggest that the rest of the gap is explained by unobserved differences in school (and possibly child and household) characteristics. Those decompositions also show very few differences in the impacts of the child and household variables across the two countries, which supports the assumption of no differences in those impacts.

The gap in the mathematics skills between 15-year-olds in Vietnam and Peru, at about 0.7 standard deviations, is somewhat smaller but still very large. Assuming that the impacts of the child, household and school variables are the same for all four countries, the child and household variables as a group explain none of this gap. However, when school variables are added, including primary school math teachers' pedagogical skills, 65-84% of the gap is explained, and almost all of this explanatory power is from math teachers' pedagogical skills, which are 1.5 standard deviations higher in Vietnam than in Peru. Part of the explanatory power of these

[19] Ethiopian math teachers were tested on their mathematics knowledge in the primary school survey, but this test was very different from those given in the other three countries, and it did not assess their pedagogical skills.
[20] For an earlier analysis of differences in student performance between Vietnam and India using the first three rounds of the Young Lives data, see Rolleston and James (2015).

pedagogical skills could reflect unobserved school and teacher characteristics that are correlated with these skills, yet one should also recall that the regressions in Tables 9 and 10 included eight other primary school variables and ten secondary school variables. Oaxaca-Blinder decompose-tions found few differences in the impacts of the child and household variables between Vietnam and Peru, further supporting the assumption of no differences in those impacts.

Overall, the strong performance of Vietnam's 15-year-olds relative to their counterparts in Ethiopia, India and Peru is only partially explained by this analysis of the Young Lives data. About 40-44% of the gap between Vietnam and Ethiopia, and only 14-21% of the gap between Vietnam and India, are explained. In contrast, most of the gap (65-84%) is explained for Peru, and almost all of it reflects the gap in primary school math teachers' pedagogical skills. These results strongly suggest that future research to understand student performance in developing, and developed, countries should focus on measuring teachers' pedagogical skills, and more generally on teacher characteristics and behavior, and how they can be influenced by education policies.

References

Alderman, Harold, Jere Behrman, Lia Fernald, Paul Glewwe and Susan Walker. 2017. "Evidence of Impact of Interventions on Growth and Development during Early and Middle Childhood", in Bundy, D., et al. eds., *Child and Adolescent Health and Development* (Volume 8 of *Disease Control Priorities, 3rd edition*). The World Bank.

Aurino, Elisabetta, and Zoe James. 2014. "Young Lives School Survey, Data User Guide: Ethiopia School Survey (2012-13)". Young Lives Study. Oxford University.

Cameron, Colin, Jonah Gelbach and Douglas Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90(3):414-427.

Cueto, Santiago, and Juan Léon. 2012. "Psychometric Characteristics of Cognitive Development and Achievement Tests in Round 3 of Young Lives". Technical Note 25. Young Lives Study. Oxford University.

Dang, Hai-Anh, Paul Glewwe, Jongwook Lee and Khoa Vu. 2019. "What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012 and 2015 PISA Data". Department of Applied Economics, University of Minnesota.

Escobal, Javier. and Eva Flores (2008). "An Assessment of the Young Lives Sampling Approach in Peru". Technical Note 3, Oxford: Young Lives.

Fryer, Roland, and Steven Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School". *Review of Economics and Statistics* 86(2):447-464.

Glewwe, Paul, Qihui Chen and Bhagyashree Katare. 2015. "What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data". *Asia & the Pacific Policy Studies* 2(3):494-516.

Glewwe, Paul, and Karthik Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps and Policy Implications," in E. Hanushek, S. Machin and L. Woessmann, eds., *Handbook of the Economics of Education: Volume 5*. North Holland.

James, Zoe. 2013a. "Young Lives School Survey, India 2010/2011: Data User Guide". Young Lives Study. Oxford University.

James, Zoe. 2013b. "Young Lives School Survey, Data User Guide: Vietnam School Survey, Round 1 (2011-2012)". Young Lives Study. Oxford University.

Kumra, Neha. 2008. "An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India". Technical Note 2, Oxford: Young Lives.

OECD. 2014. *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science, Volume I (Revised Edition).* Paris: Organization for Economic Cooperation and Development.

Outes-Leon, Ingo, and Alan Sanchez. 2008. "An Assessment of the Young Lives Sampling Approach in Ethiopia". Technical Note 3, Oxford: Young Lives.

Rolleston, Caine, and Zoe James. 2015. "After access: Divergent learning profiles in Vietnam and India" *Prospects* 45(3):285-303.

World Bank and Ministry of Planning and Investment of Vietnam. 2016. *Vietnam 2035: Toward Prosperity, Creativity, Equity, and Democracy*. Washington, DC: World Bank.

World Bank. 2017. *World Development Indicators Online Database*. The World Bank, Washington, DC.

Figure 1. Mean Age 15 Math Scores in 2012 PISA, by 2010 Log Real GDP/capita

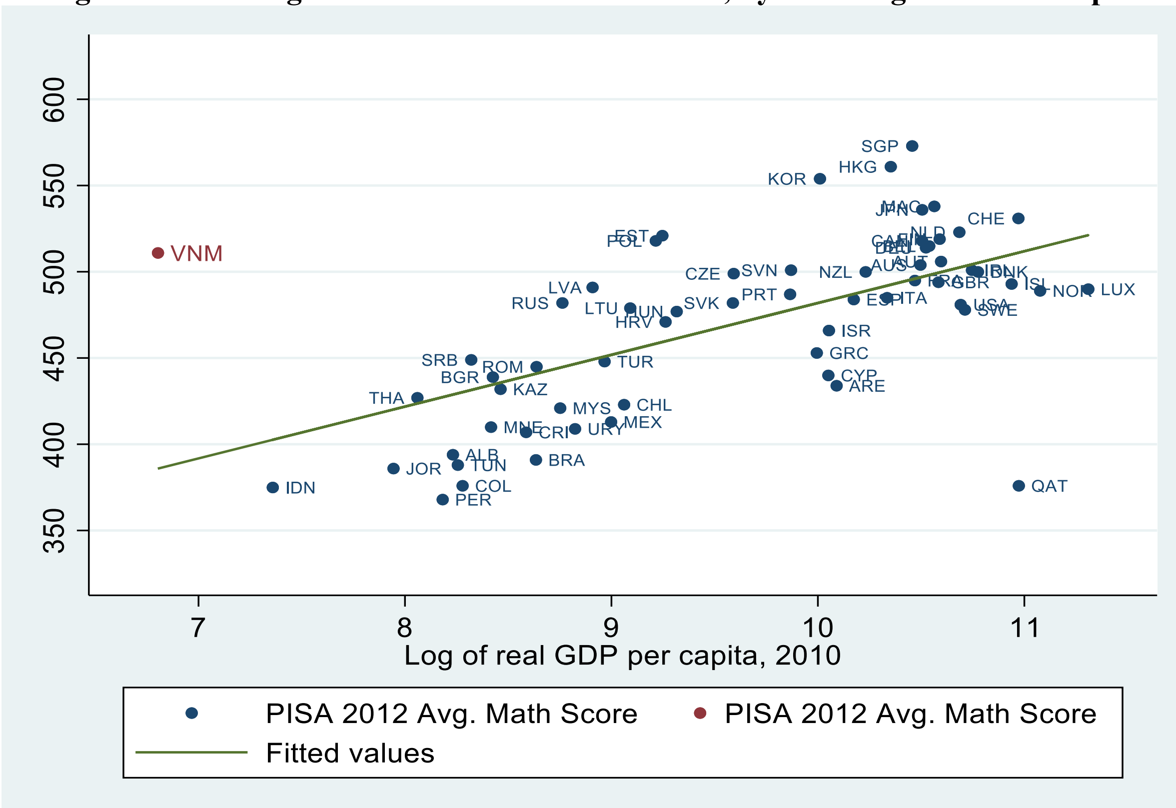

Figure 2. Mean Age 15 Reading Scores in 2012 PISA, by 2010 Log Real GDP/capita

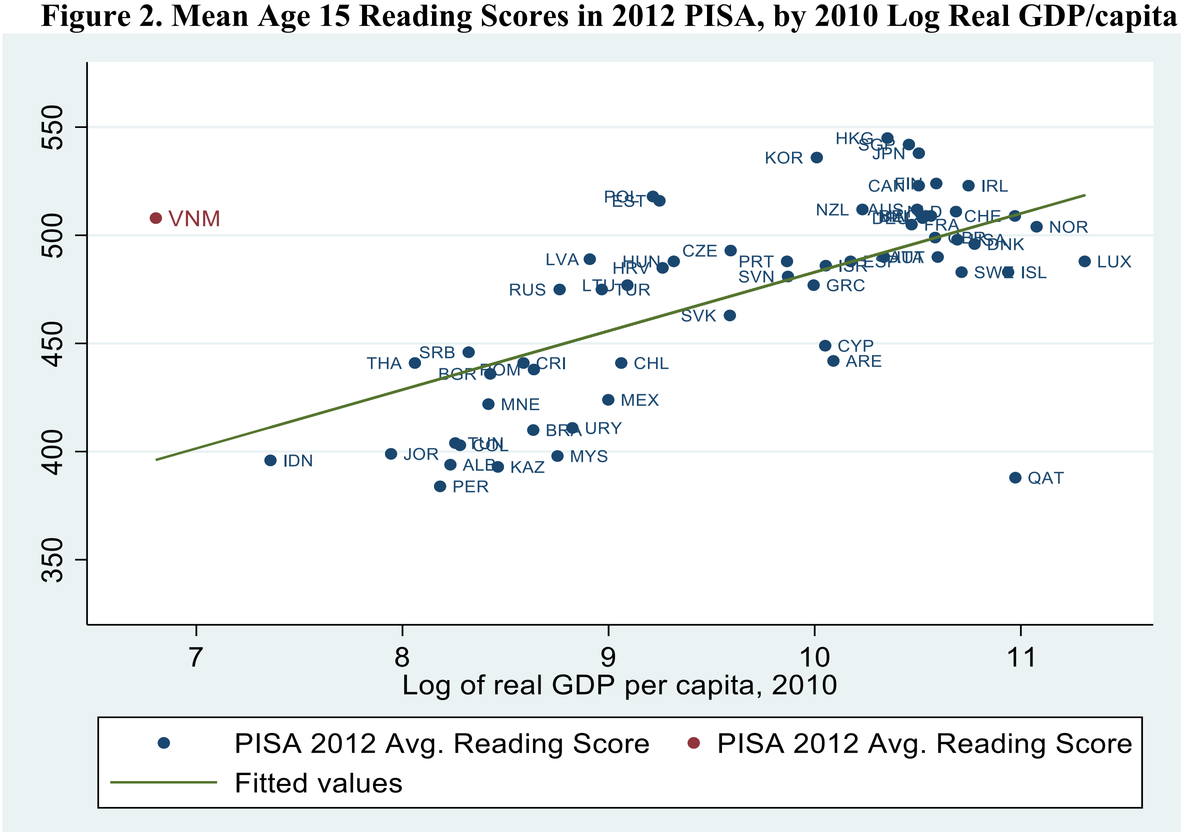

Figure 3: Kernel Density Estimates of Mathematics Test Scores in All Four Countries

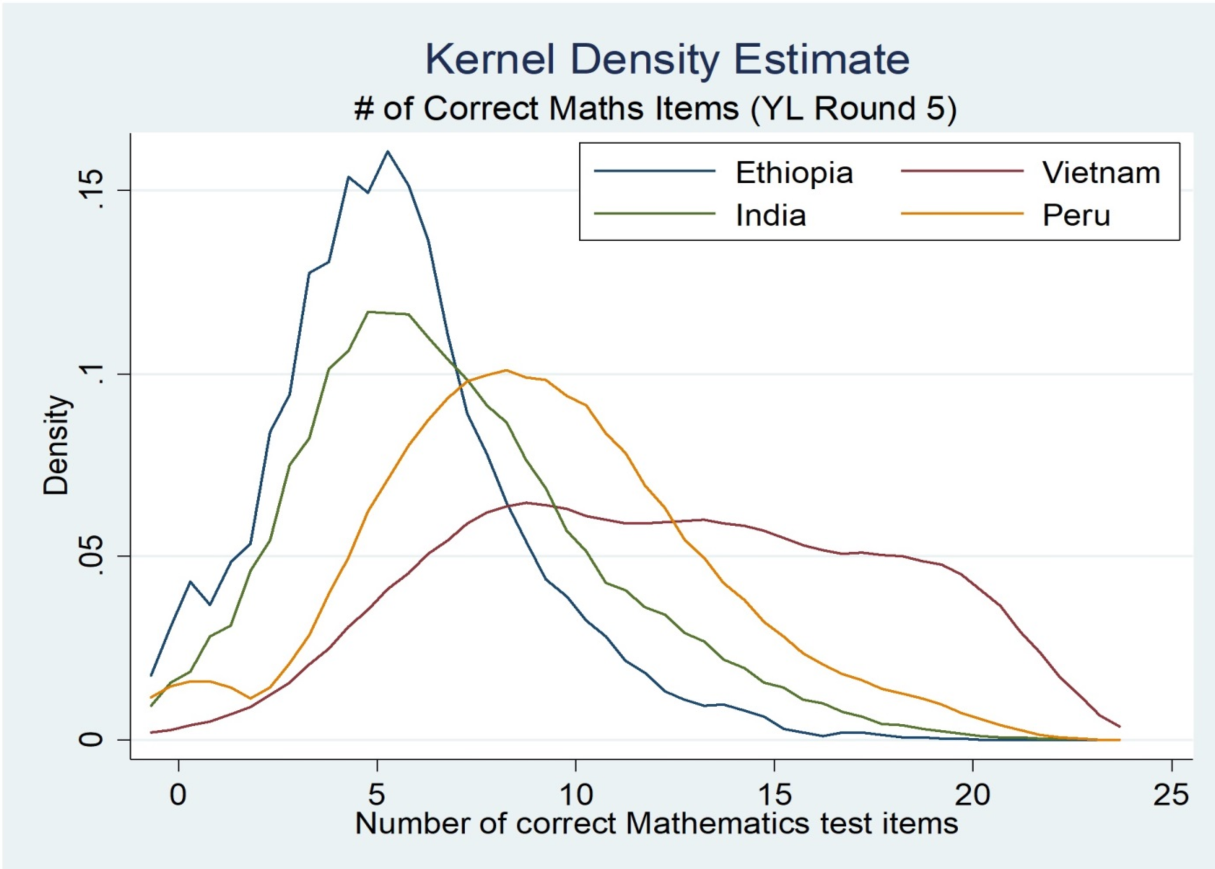

Figure 4: Sample Question of Test of Mathematics Teachers Pedagogical Skills

2. Sometimes pupils make mistakes with multiplication. For example, in the following exercise, Ngoc makes a mistake:

$$\begin{array}{r} 13 \\ \times\ \underline{15} \\ 65 \\ +\ \underline{13} \\ 78 \end{array}$$

Which reason would best explain Ngoc's mistake? (Put an X in one box)

2		**Put an X in one box**
A	She is confused between addition and multiplication	
B	She doesn't understand anything about multiplication	
C	She forgets to write 13 one column to the left	
D	She forgets to write 65 one column to the left	

Table 1. Number of Math Questions Correct and Estimated Latent Skill in All 4 Countries (includes all 15-year old children, including those not in school)

Country	Sample Size	Number of Questions Correct (out of 23)	Standardized Score	Latent Skill (IRT)	GDP/Capita in 2015 Unadjusted	GDP/Capita in 2015 PPP-adjusted
Ethiopia	1,709	5.5***	1.13***	-0.590***	$641	$1,658
		(3.0)	(0.62)	(0.696)		
India	1,840	6.9***	1.42***	-0.211***	$1,606	$5,465
		(3.7)	(0.77)	(0.748)		
Peru	1,869	9.2***	1.87***	0.033***	$6,229	$11,572
		(4.2)	(0.84)	(0.736)		
Vietnam	1,888	12.3	2.54	0.760	$2,085	$6,103
		(5.2)	(1.07)	(0.946)		

Notes: The standardized score divides the "raw" score by 4.8618, which is the standard deviation of the "raw" score over all four countries. PPP adjusted GDP/capita figures are adjusted for purchasing power parity, and are calculated by the World Bank. Number in parentheses are standard deviations. Statistical significance relative to Vietnam at the 1% significance level is denoted by ***.

Table 2. Possible Explanations for Performance of Vietnam's 15-year-olds: Child Variables

	Ethiopia	India	Peru	Vietnam
Nutritional Status				
Average height-for-age Z-score, age 5	-1.45	-1.65**	-1.54	-1.35
Percent of children who are stunted (Z-score < -2), age 5	31.3%	35.7%**	33.2%	25.3%
Family Size and Wealth				
Number of siblings, age 8	3.0***	1.5**	1.7**	1.3
Wealth index (when child was 12 years old)	0.32***	0.52***	0.62	0.63
Parental Education and Support to Education				
Average father's years of education	3.5***	4.7***	8.9**	7.0
Average mother's years of education	2.4***	3.1***	7.7*	6.2
Mother or father helps child with homework:				
Age 12	14.3%**	15.6%*	34.9%***	21.6%
Age 15	10.3%***	9.6%***	14.6%***	4.3%
Hours Devoted to Education				
Hours/day in school (includes travel time):				
Age 8	4.9	7.7***	6.0***	4.9
Age 12	5.6	8.0***	6.1***	5.4
Age 15	5.3	7.8***	6.9***	5.0
Hours/day studying at home:				
Age 8	1.0***	1.8***	2.0***	2.9
Age 12	1.5***	1.9***	1.8***	2.6
Age 15	1.8***	2.1**	2.1**	2.6
Spending per Year on Private Tutors (PPP $)				
Age 8 (% tutored in parentheses)	8.4*** (7.8)	13.0***(22.7)	5.8***(6.2)	87.0 (58.9)
Age 12	6.6*** (7.4)	9.5***(10.1)	12.5***(8.6)	213.7 (58.5)
Age 15	5.8*** (7.1)	13.2***(7.3)	14.0***(6.0)	369.8 (62.8)
Parental Aspirations (when child was 5 years old)				
Finish university or other post-sec. educ.	71.8%	57.9%***	87.2%**	78.6%
% of parents who think child will attain this	90.8%***	88.7%***	91.4%***	78.9%
Sample size range (varies by variable)	1,633-1,999	1,893-2,011	1,586-2,052	1,895-2,000

Notes. The wealth index, which ranges from 0 to 1, is a simple average of three sub-indices on housing, access to services and consumer durables. The housing sub-index is a simple average of four variables: three dummy variables indicating whether the walls are made of "good-quality" material, the roof is "sturdy" and the floors are made of finished material, plus the ratio of rooms to household size (rescaled to be between 0 and 1). The access to services subindex is a simple average of dummy variables indicating access to electricity, safe drinking water source, sanitary toilet and use of a clean cooking fuel. The consumer goods index, which also ranges from 0 to 1, indicates ownership of seven durable goods. The sample size lower bounds for Ethiopia and Peru reflect missing data on father's education; excluding that variable increases the lower bounds to 1,808 for Ethiopia and 1,830 for Peru. Statistical significance relative to Vietnam at the 10%, 5% and 1% levels is indicated by *, ** and ***, respectively.

Table 3. Possible Explanations for Performance of Vietnam's 15-year-olds: School Variables

	Ethiopia	India	Peru	Vietnam
Primary School Level Variables (site averages)				
Percent tchrs w/ general (non-ed.) univ. degree	5.4%***	78.9%***	84.2%***	94.9%
Teacher years of experience	10.6***	6.6***	17.3	17.1
Teacher is female	48.3%***	48.2%***	62.9%*	73.6%
Contract (non-tenured) teacher	9.0%**	51.5%***	--	0.8%
Teacher thinks tchrs can influence studs (1-4)	3.14	--	--	3.12
Principal's years of experience as a principal	4.0***	6.3***	12.7	10.2
Rate of teacher absence (days per month)	0.63***	0.71***	0.48*	0.23
School has electricity	52.8%***	86.0%**	94.1%	95.9%
School has a library	62.4%	21.2%**	43.9%***	74.0%
School has computers for student use	20.1%	30.0%	58.2%**	33.9%
Private school	11.2%**	48.6%***	15.6%***	0.0%
Student-teacher ratio	30.5	25.2	20.4***	28.1
Assign students to classes by ability (tracking)	13.9%	14.7%	0.8%**	10.0%
Teacher mean score on math pedagogy test:				
Latent ability estimated by IRT	--	0.021***	-0.786***	0.734
Sample size	1,999	2,011	1,443	2,000
Secondary School Level Variables (site averages)				
Principal's years of experience as a principal	8.5	6.2	8.8	--
Principal's years exper. principal in this school	3.0***	4.3*	5.2	6.4
Principal has masters or higher degree	0.0%***	70.5%***	48.1%*	34.7%
Principal helps teachers (scale of 1 to 4)	3.1	3.2**	2.9**	3.0
School has electricity	72.5%***	70.1%***	100.0%	99.5%
School has a library	75.0%*	80.3%*	64.5%**	92.2%
Computers per student	0.007***	0.025***	--	0.084
School has internet connection	18.5%***	52.6%***	86.1%	96.7%
Private school	6.5%	28.3%*	29.6%**	7.7%
Class size	42.0	29.9***	24.8***	39.2
Assign students to classes by ability (tracking)	14.3%	10.0%	--	17.1%
Minutes per week in math classes	210***	325***	--	182
Teacher years of experience	9.6***	11.8*	16.9**	14.0
Female teacher	10.1%***	29.3%***	22.2%***	49.5%
Teachers w/ general (non-ed.) master's degree	0.0%***	61.2%***	25.9%	22.5%
Teachers w/ bachelor teacher training degree	39.6%***	82.4%***	24.5%***	96.5%
Teacher self-assessed motivation (0 to 100)	74.7*	85.9	--	86.4
Contract (non-tenured) teacher	2.5%*	30.9%***	57.2%***	11.3%
Teacher wanted to be teacher	86.4%	57.0%***	--	82.1%
Teacher thinks tchrs can influence studs (1-4)	3.15***	3.11***	--	2.90
Teacher thinks tchrs can't influence stud (1-4)	2.58	2.37**	--	2.50
Students report that teacher helps them (0-1)	0.79***	0.80***	0.81***	0.75
Students report teacher checks homework (0-2)	--	1.39**	1.21	1.17
Sample size	1,999	2,011	1,338	2,000

Table 4. Regressions of Standardized Math Scores on Country Dummy Variables and Household/Child Variables

Ethiopia dummy variable	1.131*** [-1.406]	(0.069)	0.876*** [-0.915]	(0.049)	0.148* [-0.881]	(0.085)
India dummy variable	1.416*** [-1.120]	(0.055)	0.946*** [-0.845]	(0.061)	0.014 [-1.015]	(0.128)
Peru dummy variable	1.865*** [-0.671]	(0.063)	1.109*** [-0.682]	(0.077)	0.441*** [-0.588]	(0.110)
Vietnam dummy variable	2.537*** [0.000]	(0.103)	1.791*** [0.000]	(0.093)	1.029*** [0.000]	(0.099)
Wealth index	--		0.890***	(0.111)	0.548***	(0.083)
Mother's years of education	--		0.046***	(0.004)	0.028***	(0.004)
Number of siblings	--		-0.017**	(0.008)	-0.011*	(0.006)
Height-for-age Z-score, age 5	--		0.072***	(0.012)	0.052***	(0.011)
Hours/day study at home, age 8	--		--		0.021	(0.015)
Hours/day study at home, age 12	--		--		0.055***	(0.012)
Hours/day study at home, age 15	--		--		0.076***	(0.009)
Hours/day in school, age 8	--		--		0.013	(0.009)
Hours/day in school, age 12	--		--		0.029***	(0.009)
Hours/day in school, age 15	--		--		0.052***	(0.005)
Spending/year, priv. tutoring (PPP $), age 8	--		--		0.081	(0.082)
Spending/year, priv. tutoring (PPP $), age 12	--		--		0.012	(0.050)
Spending/year, priv. tutoring (PPP $), age 15	--		--		0.231***	(0.049)
Hope child will go to university	--		--		0.079***	(0.026)
Observations	7,297		7,008		6,933	
R-squared	0.824		0.854		0.870	

Standard errors, clustered at site level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Relative to Vietnam dummy in brackets.

Table 5. Regressions of Math Skill (IRT Estimates) on Country Dummy Variables and Household/Child Variables

Ethiopia dummy variable	-0.590*** [-1.350]	(0.086)	-0.846*** [-0.839]	(0.053)	-1.725*** [-0.828]	(0.080)
India dummy variable	-0.211*** [-0.971]	(0.061)	-0.703*** [-0.696]	(0.066)	-1.847*** [-0.950]	(0.129)
Peru dummy variable	0.033 [-0.727]	(0.058)	-0.742*** [-0.735]	(0.074)	-1.570*** [-0.673]	(0.107)
Vietnam dummy variable	0.760*** [0.000]	(0.095)	-0.007 [0.000]	(0.087)	-0.897*** [0.000]	(0.098)
Wealth index	--		0.961***	(0.108)	0.598***	(0.082)
Mother's years of education	--		0.044***	(0.003)	0.026***	(0.003)
Number of siblings	--		-0.022***	(0.007)	-0.022***	(0.007)
Height-for-age Z-score, age 5	--		0.072***	(0.012)	0.072***	(0.012)
Hours/day study at home, age 8	--		--		0.023	(0.014)
Hours/day study at home, age 12	--		--		0.063***	(0.010)
Hours/day study at home, age 15	--		--		0.080***	(0.007)
Hours/day in school, age 8	--		--		0.028***	(0.010)
Hours/day in school, age 12	--		--		0.039***	(0.009)
Hours/day in school, age 15	--		--		0.053***	(0.005)
Spending/year, priv. tutoring (PPP $), age 8	--		--		0.059	(0.064)
Spending/year, priv. tutoring (PPP $), age 12	--		--		0.000	(0.043)
Spending/year, priv. tutoring (PPP $), age 15	--		--		0.175***	(0.044)
Hope child will go to university	--		--		0.074***	(0.023)
Observations	7,297		7,008		6,933	
R-squared	0.280		0.425		0.503	

Standard errors, clustered at site level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Relative to Vietnam dummy in brackets.

Table 6. Comparing Primary School Parameter Estimates: School-level vs. Site Means

Dependent variable	***Standardize Math Score***		***Math Skill (IRT Estimate)***	
Primary School Variable Aggregation	School level	Site level	School level	Site level
Primary School Variables				
Teacher years of experience	-0.003	0.001	-0.003	-0.002
	(0.003)	(0.008)	(0.003)	(0.007)
Teacher is female	-0.037	-0.153*	-0.019	-0.104
	(0.035)	(0.086)	(0.035)	(0.091)
Principal years of experience	0.003	0.010*	0.004	0.010*
	(0.002)	(0.006)	(0.002)	(0.006)
Teacher absence rate (days per month)	-0.021***	-0.116***	-0.023***	-0.127***
	(0.007)	(0.028)	(0.007)	(0.031)
School has a library	0.125***	0.302***	0.111***	0.251**
	(0.036)	(0.110)	(0.034)	(0.103)
School has computers for students	0.066	0.022	0.071*	0.026
	(0.045)	(0.083)	(0.0408)	(0.067)
Private school	-0.033	-0.039	-0.004	0.145
	(0.069)	(0.155)	(0.061)	(0.163)
Student-teacher ratio	0.000	0.003	-0.001	0.003
	(0.001)	(0.003)	(0.001)	(0.002)
Secondary School Variables				
Principal’s yrs. exp. as principal in this school	-0.005	0.005	-0.006	0.002
	(0.009)	(0.008)	(0.008)	(0.008)
Principal helps teachers (scale of 1 to 4)	0.116***	0.101**	0.107**	0.096**
	(0.043)	(0.042)	(0.049)	(0.043)
School has electricity	-0.123	-0.162	-0.066	-0.101
	(0.088)	(0.102)	(0.091)	(0.103)
School has library	-0.080	-0.097	-0.003	-0.029
	(0.070)	(0.061)	(0.069)	(0.060)
Private school	0.106	-0.041	0.122	-0.083
	(0.120)	(0.116)	(0.119)	(0.118)
Class size	0.010***	0.005	0.008**	0.005
	(0.003)	(0.004)	(0.003)	(0.004)
Teacher years of experience	-0.002	-0.002	0.004	0.005
	(0.006)	(0.006)	(0.006)	(0.006)
Teacher is female	0.146	0.221*	0.138	0.216*
	(0.128)	(0.112)	(0.128)	(0.116)
Teacher has Master’s degree	0.510***	0.407***	0.511***	0.435***
	(0.133)	(0.117)	(0.121)	(0.108)
Students report teacher helps them	1.711***	2.000***	1.741***	2.285***
	(0.572)	(0.631)	(0.606)	(0.614)
Observations/R-squared	4,200/0.877	4,200/0.878	4,200/0.553	4,200/0.556

Note: These regressions also include all of the variables used in the regressions in Tables 4 and 5; they are not shown to avoid unnecessary clutter. Standard errors, clustered at site level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Regressions of Standardized Math Scores on Country, Household/Child & School Variables

Ethiopia dummy variable	1.122***	[-1.413]	0.140	[-0.888]	-2.171***	[-0.843]
	(0.068)		(0.088)		(0.537)	
India dummy variable	1.419***	[-1.116]	0.012	[-1.016]	-2.424***	[-1.096]
	(0.055)		(0.134)		(0.557)	
Peru dummy variable	1.980***	[-0.555]	0.466***	[-0.562]	-1.835***	[-0.507]
	(0.083)		(0.120)		(0.511)	
Vietnam dummy variable	2.535***	[0.000]	1.028**	[0.000]	-1.328**	[0.000]
	(0.102)		(0.103)		(0.511)	
Primary School Variables						
Teacher years of experience	--		--		-0.010	
					(0.007	
Teacher is female	--		adds child		-0.194**	
					(0.075)	
Principal years of experience	--		and		0.016***	
					(0.005)	
Teacher absence rate (days per month)	--		household		-0.093***	
					(0.024)	
School has a library	--		variables		0.260**	
					(0.107)	
School has computers for students	--		--		0.085	
					(0.078)	
Private school	--		--		0.0305	
					(0.136)	
Student-teacher ratio	--		--		0.002	
					(0.002)	
Secondary School Variables						
Principal's years exper. in this school	--		--		0.013	
					(0.008)	
Principal helps teachers (scale of 1 to 4)	--		--		0.120***	
					(0.040)	
School has electricity	--		--		-0.201**	
					(0.085)	
School has library	--		--		-0.061	
					(0.062)	
Private school	--		--		-0.213**	
					(0.105)	
Class size	--		--		0.005	
					(0.003)	
Teacher years of experience	--		--		-0.000	
					(0.005)	
Teacher is female	--		--		0.212*	
					(0.119)	
Teacher has master's degree	--		--		0.267**	
					(0.107)	
Students report teacher helps them	--		--		2.316***	
					(0.582)	
Observations/R-squared	6,186/	0.826	6,186/	0.870	6,186/	0.874

Coefficients not shown for household/child variables. Std. errors, clustered at the site level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1. Country dummy variables relative to Vietnam in brackets.

Table 8. Regressions of Math Skill (IRT Estimates) on Country, Household/Child & School Variables

Ethiopia dummy variable	-0.599***	[-1.358]	-1.733***	[-0.827]	-4.011***	[-0.767]
	(0.085)		(0.083)		(0.528)	
India dummy variable	-0.209***	[-0.968]	-1.856***	[-0.950]	-4.312***	[-1.068]
	(0.061)		(0.131)		(0.549)	
Peru dummy variable	0.146*	[-0.613]	-1.559***	[-0.653]	-3.889***	[-0.645]
	(0.074)		(0.119)		(0.506)	
Vietnam dummy variable	0.759***	[0.000]	-0.906**	[0.000]	-3.244***	[0.000]
	(0.094)		(0.113)		(0.502)	
Primary School Variables						
Teacher years of experience	--		--		-0.008	
					(0.007)	
Teacher is female	--		adds child		-0.165**	
					(0.077)	
Principal years of experience	--		and		0.014**	
					(0.005)	
Teacher absence rate (days per month)	--		household		-0.107***	
					(0.028)	
School has a library	--		variables		0.218**	
					(0.100)	
School has computers for students	--		--		0.095	
					(0.066)	
Private school	--		--		0.143	
					(0.144)	
Student-teacher ratio	--		--		0.001	
					(0.002)	
Secondary School Variables						
Principal’s years exper. in this school	--		--		0.007	
					(0.009)	
Principal helps teachers (scale of 1 to 4)	--		--		0.106**	
					(0.044)	
School has electricity	--		--		-0.128	
					(0.091)	
School has library	--		--		-0.051	
					(0.064)	
Private school	--		--		-0.219*	
					(0.110)	
Class size	--		--		0.005	
					(0.003)	
Teacher years of experience	--		--		0.004	
					(0.005)	
Teacher is female	--		--		0.187	
					(0.125)	
Teacher has master’s degree	--		--		0.317***	
					(0.102)	
Students report teacher helps them	--		--		2.298***	
					(0.570)	
Observations/R-squared	6,186/	0.309	6,186/	0.521	6,186/	0.536

Coefficients not shown for household/child variables. Std. errors, clustered at the site level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1. Country dummy variables relative to Vietnam in brackets.

Table 9. Regressions of Standardized Math Scores, Adding Teacher Latent Math Ability

India dummy variable	1.419***	[-1.116]	-0.011	[-1.016]	-3.170**	[-0.887]
	(0.055)		(0.171)		(1.211)	
Peru dummy variable	1.980***	[-0.555]	0.419***	[-0.586]	-2.374**	[-0.091]
	(0.084)		(0.152)		(1.119)	
Vietnam dummy variable	2.535***	[0.000]	1.005**	[0.000]	-2.283**	[0.000]
	(0.102)		(0.122)		(1.122)	
Primary School Variables						
Teacher years of experience	--		--		-0.001	
					(0.008)	
Teacher is female	--		adds child		-0.120	
					(0.102)	
Principal years of experience	--		and		0.009	
					(0.006)	
Teacher absence rate (days per month)	--		household		-0.149***	
					(0.040)	
School has a library	--		variables		0.044	
					(0.148)	
School has computers for students	--		--		0.035	
					(0.065)	
Private school	--		--		0.146	
					(0.151)	
Student-teacher ratio	--		--		0.0104*	
					(0.005)	
Math teacher latent pedagogical skill (IRT)	--		--		0.268***	
					(0.074)	
Secondary School Variables						
Principal's years exper. in this school	--		--		0.014*	
					(0.008)	
Principal helps teachers (scale of 1 to 4)	--		--		0.077	
					(0.106)	
School has electricity	--		--		-0.007	
					(0.106)	
School has library	--		--		-0.172**	
					(0.074)	
Private school	--		--		-0.285***	
					(0.101)	
Class size	--		--		0.009**	
					(0.005)	
Teacher years of experience	--		--		-0.001	
					(0.007)	
Teacher is female	--		--		0.287*	
					(0.160)	
Teacher has master's degree	--		--		0.244**	
					(0.105)	
Students report teacher helps them	--		--		2.911***	
					(1.058)	
Observations/R-squared	4,536/	0.833	4,536/	0.877	4,536/	0.882

Coefficients not shown for household and child variables. Std. errors, clustered at the site level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Country dummy variables relative to Vietnam in brackets.

Table 10. Regressions of Math Skill (IRT Estimates), Adding Teacher Latent Math Ability

India dummy variable	-0.209***	[-0.968]	-1.693***	[-0.919]	-5.057***	[-0.829]
	(0.061)		(0.169)		(1.107)	
Peru dummy variable	0.146*	[-0.613]	-1.433***	[-0.659]	-4.442***	[-0.214]
	(0.074)		(0.145)		(1.052)	
Vietnam dummy variable	0.759***	[0.000]	-0.774***	[0.000]	-4.228***	[0.000]
	(0.095)		(0.123)		(1.048)	
Primary School Variables						
Teacher years of experience	--		--		-0.002	
					(0.007)	
Teacher is female	--		adds child		-0.148	
					(0.096)	
Principal years of experience	--		and		0.007	
					(0.006)	
Teacher absence rate (days per month)	--		household		-0.161***	
					(0.036)	
School has a library	--		variables		0.038	
					(0.135)	
School has computers for students	--		--		0.064	
					(0.059)	
Private school	--		--		0.154	
					(0.155)	
Student-teacher ratio	--		--		0.011**	
					(0.005)	
Math teacher latent pedagogical skill (IRT)	--		--		0.253***	
					(0.075)	
Secondary School Variables						
Principal's years exper. in this school	--		--		0.012	
					(0.008)	
Principal helps teachers (scale of 1 to 4)	--		--		0.098	
					(0.103)	
School has electricity	--		--		0.081	
					(0.112)	
School has library	--		--		-0.096	
					(0.073)	
Private school	--		--		-0.259**	
					(0.104)	
Class size	--		--		0.009**	
					(0.004)	
Teacher years of experience	--		--		0.003	
					(0.006)	
Teacher is female	--		--		0.223	
					(0.155)	
Teacher has master's degree	--		--		0.274***	
					(0.100)	
Students report teacher helps them	--		--		2.948***	
					(0.991)	
Observations/R-squared	4,536/	0.270	4,536/	0.490	4,536/	0.512

Coefficients not shown for household and child variables. Std. errors, clustered at the site level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Country dummy variables relative to Vietnam in brackets.

Table 11: Oaxaca-Blinder Decomposition for Normalized Math Score, Age 15, Vietnam and Ethiopia (diff = 2.535– 1.122 = 1.413)

Variable	$\boldsymbol{\beta}_{vn}$	$\bar{\mathbf{x}}_{vn}$	$\boldsymbol{\beta}_{e}$	$\bar{\mathbf{x}}_{e}$	$\bar{\boldsymbol{\beta}}$	$\bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn}-\bar{\mathbf{x}}_{e})$	$(\boldsymbol{\beta}_{vn}-\bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_{e})'\bar{\mathbf{x}}_{e}$
Wealth index (adjusted), age 12	0.312	0.635	0.453***	0.324	0.318	0.119***	-0.068
Mom years schooling	0.032***	6.332	0.003	2.415	0.018	0.068***	0.126***
Number of siblings, age 8	0.008	1.292	0.006	3.032	0.007	-0.012	0.005
Height-for-age Z-score, age 5	0.048*	4.663	0.028	4.585	0.038	0.003	0.089
Hours study at home, age 8	0.002	2.922	0.035**	1.019	0.019	0.036*	-0.064
Hours study at home, age 12	0.056***	2.688	0.055***	1.533	0.056	0.064***	0.002
Hours study at home, age 15	0.086***	2.608	0.072***	1.915	0.079	0.055***	0.032
Hours/day in school, age 8	0.025	4.945	-0.001	5.038	0.012	-0.001	0.130
Hours/day in school, age 12	0.005	5.491	0.026**	5.699	-0.011	-0.003	-0.120
Hours/day in school, age 15	0.076***	5.139	0.013	5.464	0.045	-0.014	0.332***
Spending/yr. priv. tutoring, age 8	0.112	0.085	0.959**	0.008	0.536	0.041**	-0.039**
Spending/yr. priv. tutoring, age 12	-0.036	0.222	0.124	0.006	0.044	0.010	-0.018
Spending/yr. priv. tutoring, age 15	0.124***	0.375	0.548***	0.006	0.336	0.124***	-0.081**
Parents hope child go to university	0.114*	0.809	0.022	0.761	0.785	0.003	0.073
Total contributions of observed child and household variables						0.491***	0.398***
Total contribution of school (and unobserved child and household) variables						0.524***	

Notes: The height-for-age z-score was rescaled by adding 6 so that the minimal value is slightly above zero. Spending on private tutoring was obtained from the parent questionnaire, in terms of the local currency; for comparability across countries it is converted to U.S. dollars using purchasing power parity (PPP) exchange rates.

Table 12: Oaxaca-Blinder Decomposition for Math Skill (IRT Estimates), Age 15, Vietnam and Ethiopia (diff = 0.759 – (-0.599) = 1.358)

Variable	β_{vn}	$\bar{x}_{vn}$	β_{e}	$\bar{x}_{e}$	$\bar{\beta}$	$\bar{\beta}'(\bar{x}_{vn}-\bar{x}_{e})$	$(\beta_{vn}-\bar{\beta})'\bar{x}_{vn} + (\bar{\beta}-\beta_{e})'\bar{x}_{e}$
Wealth index (adjusted), age 12	0.228	0.635	0.572***	0.324	0.400	0.124***	-0.164**
Mom years schooling	0.030***	6.332	0.002	2.415	0.016	0.062***	0.122***
Number of siblings, age 8	0.002	1.292	0.009	3.032	0.006	-0.009	-0.015
Height-for-age Z-score, age 5	0.058**	4.663	0.030*	4.585	0.044	0.003	0.130
Hours study at home, age 8	0.006	2.922	0.043*	1.019	0.025	0.047**	-0.072
Hours study at home, age 12	0.050***	2.688	0.091***	1.533	0.071	0.081***	-0.087**
Hours study at home, age 15	0.079***	2.608	0.090***	1.915	0.085	0.058***	-0.024
Hours/day in school, age 8	0.018	4.945	0.010	5.038	0.014	-0.001	0.039
Hours/day in school, age 12	0.012	5.491	0.040***	5.699	0.026	-0.005	-0.157**
Hours/day in school, age 15	0.067***	5.139	0.014	5.464	0.041	-0.013	0.280***
Spending/yr. priv. tutoring, age 8	0.054	0.085	0.906***	0.008	0.480	0.037**	-0.039**
Spending/yr. priv. tutoring, age 12	-0.023	0.222	0.268	0.006	0.123	0.026	-0.033
Spending/yr. priv. tutoring, age 15	0.102***	0.375	0.420**	0.006	0.261	0.096***	-0.061**
Parents hope child go to university	0.069	0.809	0.048	0.761	0.059	0.003	0.016
Total contributions of observed child and household variables						0.510***	-0.066
Total contribution of school (and unobserved child and household) variables						0.913***	

See notes to Table 11.

Table 13: Oaxaca-Blinder Decomposition for Normalized Math Score, Age 15, Vietnam and India (diff = 2.535– 1.419 = 1.116)

Variable	$\boldsymbol{\beta}_{vn}$	$\bar{\mathbf{x}}_{vn}$	$\boldsymbol{\beta}_{i}$	$\bar{\mathbf{x}}_{i}$	$\bar{\boldsymbol{\beta}}$	$\bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn}-\bar{\mathbf{x}}_{i})$	$(\boldsymbol{\beta}_{vn}-\bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_{i})'\bar{\mathbf{x}}_{i}$
Wealth index (adjusted), age 12	0.312	0.635	0.446***	0.525	0.379	0.042***	-0.078
Mom years schooling	0.032***	6.332	0.031***	3.211	0.032	0.098***	0.005
Number of siblings, age 8	0.008	1.292	-0.028	1.532	-0.01	0.002	0.051
Height-for-age Z-score, age 5	0.048*	4.663	0.055***	4.356	0.055	0.016**	-0.034
Hours study at home, age 8	0.002	2.922	0.028*	1.855	0.015	0.016	-0.061
Hours study at home, age 12	0.056***	2.688	0.041*	1.959	0.049	0.035***	0.035
Hours study at home, age 15	0.086***	2.608	0.075***	2.181	0.081	0.034**	0.025
Hours/day in school, age 8	0.025	4.945	0.026	7.732	0.026	-0.071*	-0.012
Hours/day in school, age 12	0.005	5.491	0.047***	8.116	0.026	-0.067***	-0.288***
Hours/day in school, age 15	0.076***	5.139	0.038***	8.047	0.057	-0.165***	0.252***
Spending/yr. priv. tutoring, age 8	0.112	0.085	0.147	0.014	0.130	0.009	-0.002
Spending/yr. priv. tutoring, age 12	-0.036	0.222	-0.207	0.010	-0.122	-0.026	0.020
Spending/yr. priv. tutoring, age 15	0.124***	0.375	-0.237	0.014	-0.057	-0.020	0.070
Parents hope child go to university	0.114*	0.809	0.092***	0.612	0.103	0.020***	0.016
Total contributions of observed child and household variables						-0.076	-0.002
Total contribution of school (and unobserved child and household) variables						1.194***	

See notes to Table 11.

Table 14: Oaxaca-Blinder Decomposition for Math Skill (IRT Estimates), Age 15, Vietnam and India (diff = 0.759 – (-0.209) = 0.968)

Variable	$\boldsymbol{\beta}_{vn}$	$\bar{\mathbf{x}}_{vn}$	$\boldsymbol{\beta}_{i}$	$\bar{\mathbf{x}}_{i}$	$\bar{\boldsymbol{\beta}}$	$\bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn}-\bar{\mathbf{x}}_{i})$	$(\boldsymbol{\beta}_{vn}-\bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_{i})'\bar{\mathbf{x}}_{i}$
Wealth index (adjusted), age 12	0.228	0.635	0.392***	0.525	0.310	0.034***	-0.095
Mom years schooling	0.030***	6.332	0.031***	3.211	0.031	0.095***	-0.005
Number of siblings, age 8	0.002	1.292	-0.021	1.532	-0.010	0.002	0.032
Height-for-age Z-score, age 5	0.058**	4.663	0.049**	4.356	0.054	0.017**	0.040
Hours study at home, age 8	0.006	2.922	0.042***	1.855	0.024	0.026**	-0.085**
Hours study at home, age 12	0.050***	2.688	0.054***	1.959	0.052	0.038***	-0.010
Hours study at home, age 15	0.079***	2.608	0.080***	2.181	0.080	0.034**	-0.002
Hours/day in school, age 8	0.018	4.945	0.038*	7.732	0.028	-0.078*	-0.127
Hours/day in school, age 12	0.012	5.491	0.051***	8.116	0.032	-0.083***	-0.272**
Hours/day in school, age 15	0.067***	5.139	0.048***	8.047	0.058	-0.167***	0.123**
Spending/yr. priv. tutoring, age 8	0.054	0.085	-0.034	0.014	0.010	0.001	0.004
Spending/yr. priv. tutoring, age 12	-0.023	0.222	-0.028	0.010	-0.026	-0.005	0.001
Spending/yr. priv. tutoring, age 15	0.102***	0.375	-0.184	0.014	-0.041	-0.015	0.056
Parents hope child go to university	0.069	0.809	0.089***	0.612	0.079	0.016**	-0.014
Total contributions of observed child and household variables						-0.086	-0.355***
Total contribution of school (and unobserved child and household) variables						1.409***	

See notes to Table 11.

Table 15: Oaxaca-Blinder Decomposition for Normalized Math Score, Age 15, Vietnam and Peru (diff = 2.535– 1.904 = 0.631)

Variable	$\boldsymbol{\beta}_{vn}$	$\bar{\mathbf{x}}_{vn}$	$\boldsymbol{\beta}_{p}$	$\bar{\mathbf{x}}_{p}$	$\bar{\boldsymbol{\beta}}$	$\bar{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_{vn}-\bar{\mathbf{x}}_{p})$	$(\boldsymbol{\beta}_{vn}-\bar{\boldsymbol{\beta}})'\bar{\mathbf{x}}_{vn} + (\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_{i})'\bar{\mathbf{x}}_{p}$
Wealth index (adjusted), age 12	0.312	0.635	0.513***	0.623	0.413	0.005	-0.127
Mom years schooling	0.032***	6.332	0.035***	7.826	0.034	-0.050*	-0.024
Number of siblings, age 8	0.008	1.292	-0.008	1.718	0.000	0.000	0.025
Height-for-age Z-score, age 5	0.048*	4.663	0.051***	4.485	0.051	0.009	-0.017
Hours study at home, age 8	0.002	2.922	0.014	2.034	0.008	0.007	-0.029
Hours study at home, age 12	0.056***	2.688	0.037**	1.852	0.047	0.039***	0.041
Hours study at home, age 15	0.086***	2.608	0.057***	2.079	0.072	0.038***	0.068
Hours/day in school, age 8	0.025	4.945	0.003	6.015	0.014	-0.015	0.120
Hours/day in school, age 12	0.005	5.491	0.072***	6.062	0.039	-0.022***	-0.390***
Hours/day in school, age 15	0.076***	5.139	0.065***	6.912	0.071	-0.125***	0.063
Spending/yr. priv. tutoring, age 8	0.112	0.085	-0.544	0.006	-0.216	-0.017	0.030*
Spending/yr. priv. tutoring, age 12	-0.036	0.222	0.657*	0.013	0.311	0.065	-0.081
Spending/yr. priv. tutoring, age 15	0.124***	0.375	0.107	0.014	0.116	0.042**	0.003
Parents hope child go to university	0.114*	0.809	-0.021	0.915	0.047	-0.005	0.117*
Total contributions of observed child and household variables						-0.029	-0.201***
Total contribution of school (and unobserved child and household) variables						0.861***	

See notes to Table 11.

Table 16: Oaxaca-Blinder Decomposition for Math Skill (IRT Estimates), Age 15, Vietnam and Peru (diff = 0.759 – 0.073 = 0.686)

Variable	β_{vn}	$\bar{x}_{vn}$	β_p	$\bar{x}_p$	$\bar{\beta}$	$\bar{\beta}'(\bar{x}_{vn}-\bar{x}_p)$	$(\beta_{vn}-\bar{\beta})'\bar{x}_{vn} + (\bar{\beta}-\beta_p)'\bar{x}_p$
Wealth index (adjusted), age 12	0.228	0.635	0.421***	0.623	0.325	0.004	-0.121
Mom years schooling	0.030***	6.332	0.032***	7.826	0.031	-0.046*	-0.012
Number of siblings, age 8	0.002	1.292	-0.004	1.718	-0.001	0.001	0.010
Height-for-age Z-score, age 5	0.058**	4.663	0.042**	4.485	0.050	0.009	0.076
Hours study at home, age 8	0.006	2.922	0.012	2.034	0.009	0.008	-0.015
Hours study at home, age 12	0.050***	2.688	0.028*	1.852	0.039	0.032***	0.050
Hours study at home, age 15	0.079***	2.608	0.066***	2.079	0.073	0.038***	0.030
Hours/day in school, age 8	0.018	4.945	0.013	6.015	0.016	-0.016	0.029
Hours/day in school, age 12	0.012	5.491	0.049***	6.062	0.031	-0.017***	-0.214**
Hours/day in school, age 15	0.067***	5.139	0.061***	6.912	0.064	-0.113***	0.032
Spending/yr. priv. tutoring, age 8	0.054	0.085	-0.351	0.006	-0.149	-0.012	0.018
Spending/yr. priv. tutoring, age 12	-0.023	0.222	0.459	0.013	0.218	0.046	-0.056
Spending/yr. priv. tutoring, age 15	0.102***	0.375	0.145*	0.014	0.124	0.045***	-0.008
Parents hope child go to university	0.069	0.809	-0.002	0.915	0.034	-0.004	0.061
Total contributions of observed child and household variables						-0.026	-0.121***
Total contribution of school (and unobserved child and household) variables						0.832***	

See notes to Table 11.

Appendix A. Matching Younger Cohort Children to Their Primary Schools

This appendix explains how, for all four Young Lives countries, students for whom there are mathematics test scores from Round 5 were matched to the primary schools that were surveyed in the Primary School Surveys between Round 3 and Round 4 in each country.

Ethiopia

According to the school survey document (Aurino and James, 2014), the Ethiopia Primary School Survey was conducted during the 2012-13 school year, and all primary schools located within the geographic boundaries of the 20 sampling sites were to be included. Thus, any children enrolled in primary schools outside of the sampling sites would not have their schools included. Also, the survey focused on Grades 4 and 5, so no data were collected from Young Lives children in the surveyed schools who were in other grades, such as Grade 3 or Grade 6.

The data on the school attended in the Round 4 education history data set include the region, zone, woreda (district) and kebele/PA where the school is located. A site consists of one or two kebeles. If 100 or more households with children of younger cohort age were found in a selected kebele in Round 1, then the site consists of that kebele only, but if 100 such households could not be found, then a neighboring kebele was added to the site in order to draw a sample up to 100 households with children of younger cohort age. The first two columns of Table A.1 show how many kebele belonged to each site; site "names" are fictitious to ensure households' privacy.[21]

The primary school survey collected data from 92 schools, plus 2 satellite schools that were attached to the same "main" school that was one of the 92 schools. Of these 92 schools, 66 are matched to one of the 20 original sites. The other 26 (28 if the satellite schools are counted as separate schools) are in "new" sites that pertain to a separate data collection effort involving schools in Ethiopia's Afar and Somali regions; these schools have no relation to the original Young Lives sample of children, which did not include these two regions. The 66 schools in the original 20 sites of have school codes ranging from 53 to 516, as well as codes 542, 543 and 544. The schools in "new" sites have school codes from 517 to 541, plus code 545.

To get an idea of which schools should have been included in the Primary School Survey, because they were physically located in the kebeles that constituted the 20 sampling sites, the Round 4 education history data were used. These data contain school codes for the 2012-13 school year for 1,744 of the 1,874 children in that data set. These 1,744 children provided 320 different school codes, so at least 254 of these children cannot be matched to a school in the Primary School Survey since that survey collected data from only 66 schools in the 20 sampling sites. Fortunately, the Round 4 education history data contain the name or code number for almost all of the kebeles in which these 320 schools are located (the main exception is that this information is missing for the 3 sites located in Addis Ababa). These data were used to approximate the number of the kebeles associated with each site for sites 4-20, as reported in parentheses in the second column of Table 2. Based on these reports of kebeles where these schools were located (from the Round 4 education history data), the third column of Table A.1

[21] We thank Rayyan Mobarak for assisting in this exercise, which was very tedious due to many different spellings of kebele names in the Round 4 education history data.

reports which of the 320 schools were expected to be in the Primary School Survey (since the Round 4 data indicate that they are located in a specific site), and the fourth column reports which schools were not expected to be in the survey because the Round 4 data indicated that they were not in one of the 20 sites.

Table A.1: Ethiopia Primary School Survey, Schools Located In and Outside of the 20 Sites

Site	Kebeles (anonymized name, with number of kebeles in parentheses)	Location of Schools in or out of Kebele, as indicated in Round 4 Education History		Schools found in Primary School Survey
		In	**Out**	
1-3	No information available	Unknown	Unknown	Not applicable
4	Kok (1)	164, 177, 192, 806, 817	160, 498-754, 807, 818-1360	177
5	Muz (1)	153, 154, 836	177-823, 840-1217	153, 154
6	Enkoy (2)	136, 138, 174, 175, 257, 542, 826	142-168, 805-815, 831-852	137, 144, 168, 174, 438
7	Tach-Meret (1)	180, 189	492-12 47	158, 180, 189, 502, 503
8	Leki (1)	209	201, 202, 554-997	209
9	Lomi (1)	243	224,228, 253-940	213, 243
10	Ananas (1)	177, 237, 238, 239, 240, 949, 998	631-941, 953-993	238, 240, 241
11	Dinich (1)	215, 249	218, 226, 824-1362	249, 252
12	Timatim (1)	261, 270, 330, 369	930 - 1099	261, 369
13	Shenkurt (1)	276, 277	295-1151	276, 504
14	Leku (2)	269, 301, 353, 512, 514, 516, 544, 1039, 1052, 1117, 1142, 1349	511, 759-977, 1047, 1055-1112, 1124, 1147-1317	269, 301, 353, 510-514, 516, 542, 543, 544
15	Buna (1)	327, 1155, 1342	1069-1154,1156-1162	327
16	Weyn (2)	264, 289, 293, 341	304, 1014-1149	289, 293, 341
17	Zeytuni (1)	385, 388, 390, 429, 500	380, 1199-1241	385, 390, 429, 500, 501
18	Selata (2)	388, 395, 405, 416, 491	393, 400, 1201-1236	388, 491, 499
19	Gomen (1)	391, 392, 393	400 – 1240	391, 392, 393
20	Beles (1)	375	400-1239	375, 415

Note: No Young Lives children were enrolled in schools 806, 836, 998 and 1117 in the 2012-13 school year, although Young Lives children were enrolled in them in the 2013-14 school year.

For sites 4-20, the younger cohort children in the Round 4 education history report attending 230 different schools in the 2012-13 school year, of which 67 schools were reported to be located in the kebeles assigned to those sites (these are the schools in column 3 of Table A.1). Of these 67 schools, 36 are found in the school survey data. (It is possible that 4 of the 67 schools were not operating in the 2012-13 schools year, as explained in the note to Table A.1, in which case one would not expect these four schools to be included in the school survey.) While this may seem to be a low “success” rate in terms of covering the schools that should have been covered, note that of the 1,477 children from the education history in Round 4 in sites 4-20 who report a school code for the school year 2012-13, only 1,166 report attending a school that (based on the algorithm used above) is within the geographic boundaries those sites. Of these 1,166, the school survey collected data for 924, which is 79.2% of the younger cohort children whose schools were supposed to be in the school survey. If the comparison is limited to children who report being in grade 4 or 5 in the 2012-13 school year, as reported in the Round 4 education history data, there are 561 (out of the 1,166), of whom 432 (77.0%) were matched to a school in the school survey.

To summarize for sites 4-20, the Round 4 education history contains school codes for the 2012-13 school year for 1,477 of the younger cohort children in those sites. The school survey was planned to collect data only for the schools that were physically within the site boundaries. Based on our reconstruction of which schools were in those sites (using kebele codes from Round 4), 1,166 of these 1,477 children were enrolled in schools located in those sites in the 2012-13 schools year. Of these 1,116 children, the primary school survey collected data from the schools of 924 (79.2%) of them.

Next, consider how many children who have mathematics test scores in Round 5 in Ethiopia can be matched to a primary school for which data were collected in the 2012-13 Primary School Survey. Note that this is for all 20 sites, not just sites 4-20. The Round 5 Ethiopia data contain 1,887 children, of whom 1,709 have math test scores. Of these, 1,682 (all but 27) have a school code found in the Round 4 education history data, and of these 1,682 there are 528 who were included in the pupil roster of the Primary School Survey (because they were in Grade 4 or 5 in their schools) and thus have a “guaranteed” school code from that source (“guaranteed” because by definition they are matched to a school in the Primary School Survey).

The following approach was taken to create a new school id variable (called “newschid”) to use for matching school data to student data. First, for the 528 children with school codes directly from the pupil roster of the Primary School Survey, the school code from that survey was used. Then, for children without a school code from the pupil roster of the Primary School Survey, the school codes from Round 4 education history for the 2012-13 school year were used. This increased the number of younger cohort children with school codes by 1,154, to 1,682 (in part by incorporating children in grades other than 4 or 5, who were excluded from the Primary School Survey pupil roster). Thus, of the 1,709 younger cohort children for whom there are mathematics tests from Round 5, there are school codes for 1,682 (98.4%). However, since many of the Round 4 school codes do not match any codes from the school survey (in part due to children attending schools outside of the site boundaries), the actual match rate to school data is much lower. More specifically, of the 1,682 younger cohort children for whom there are school codes, those codes actually match to schools in the school survey for only 1,156. Thus, of the 1,709 younger cohort children for whom there are mathematics tests in Round 5, it was possible to match only 1,156 (67.6%) to a school from the Primary School Survey.

Finally, note that there is serious concerns about the accuracy of the matching of students to schools using the school codes from the Round 4 education history. The school codes in the

Pupil Roster of the School Survey should be for 2012-13 school year. The school codes from the Round 4 Education History can also be extracted for the 2012-13 school year. These two data sources can be merged by child ID codes. When this is done, and no effort is made to "replace" missing school codes in the Round 4 Education History by looking at later school years (which would be misleading because, for this exercise, we are focused on where the children were in the 2012-13 school year), there are 542 younger cohort children who can be matched and have school codes from both sources. Of these, only 420 (77.5%) have the same school codes from the two sources. This suggests possible measurement error in the school data because of matching errors.

India

According to the school survey document (James, 2013a), the younger cohort children attended 807 schools in 2009 (when Round 3 data were collected), and for 538 schools there was only one younger cohort child in each of those schools. A decision was made to collect data from about 300 schools, and to do so in a way that collected sufficient data from each of six school types (urban private recognized schools, rural private recognized schools, urban private unrecognized schools, rural private unrecognized schools, urban public schools, and rural public schools). This was done by using the Round 3 data to randomly draw children within each of these six strata, which yielded a sample of 1,111 children who attended 299 different schools. About 20% of these children had switched schools between Round 3 (2009) and the months that the school survey was implemented (December 2010 to March 2011), and any such "new" schools (which were not attended by any sample children in 2009) were not surveyed. In the end, data were collected from 249 schools in 2010-11, which were attended by 953 of the 1,111 sampled children (the other 158 had switched schools). Note that the school codes in the Primary School Survey for these 249 schools range from 1 to 839.

The Round 5 India data contain 1,905 children, of whom 1,840 have math test scores. Of these, 1,836 (all but 4) are also found in the Round 4 education history data, and 3 more are found in the Primary School Survey roster file but are not in the Round 4 education history file, yielding 1,839 for whom there should be school codes. Of these 1,839, there are school codes for 1,631, but 994 of these have school codes >839, so they do not match school survey data.

So the approach taken is as follows. First, school codes were obtained from pupil roster of Primary School Survey. This yields school codes for 920 children. Then, the school codes from the Round 3 education were examined; although the codes are not Young Lives School Codes, Young Lives School Codes could be assigned to almost all of them based on matching Round 3 education history child id codes to the child id codes in the pupil roster of the Primary School Survey. Doing this increased the number of younger cohort children with school codes by 153, to 1,073. Then school codes from Round 4 education history were added as long those codes were ≤ 839. This increased the number of younger cohort children with school codes by 83, to 1,156. Finally, the Round 3 education history also has an "expected" school code, the school that the child was expected to be in. Adding these increases the number of children with school codes by 46, to 1,202. Thus, of the 1,839 for whom we hoped to obtain school id codes, we were able to obtain such a code for 1,202, which is 65.4% rate. However, a few of the school codes from the Round 4 education history, and perhaps some from the expected codes from Round 3 education history, do not match the school codes in the Primary School Survey data; this is expected since only a sample of the schools were drawn for the Primary School Survey. In the end, only 1,158 of the 1,839 children could be matched, which is a match rate of 63.0%.

Finally, note that there are serious concerns about the accuracy of the matching. The school codes in the pupil roster of the School Survey should be for 2010-11 school year. The school codes from the Round 4 Education History can also be extracted for the 2010-11 school year. These two data sources can be merged by child ID codes. When this is done, and no effort is made to "replace" missing school codes in the Round 4 Education History by looking at later school years, there are 653 children who have school codes from both sources. Of these, only 497 (76.1%) have the same school codes from the two sources. If missing 2010-11 school codes are replaced by using school codes from more recent years in the Round 4 education history, there are 823 younger cohort children who have school codes from both source. Of these, only 510 (62.0%) have the same school codes from both sources. Part of the reason for this discrepancy could be recall errors in the education history for the 2010-11 school year, which were collected in late 2013 and early 2014.

Peru

According to the Peru school sample documentation (Guerrero et al., 2012), the 1,721 "eligible" sample of Young Lives children attending primary school in Round 3 were enrolled in 591 different schools, which was too large of a sample given the budget available. A subsample of these schools was randomly selected, with stratification, which led to a sample of 120 schools spread across 14 regions (departments) of Peru. Another sampling procedure was then used to reduce the number of regions, which led to a sample of 131 schools in 9 regions; these schools include 662 of the younger cohort children. The five regions that were dropped (Amazonas, Apurimac, Piura, Puno and Tumbes) imply that six sites were dropped 1, 2, 3, 4, 18 and 20. The data were collected in October and November of 2011.

The Round 5 Peru data contain 1,860 children, and all 1,860 have math test scores. Of these, 1,857 (all but 3) are also found in the Round 4 education history data. Of these 1,857 we have school codes for 1,848 (ignoring 3 with school code of 9999999) from the Round 4 education history data.

So the approach taken was as follows. First, school codes were obtained from pupil roster of Primary School Survey. This yields school codes for 552 of these children. Then, school codes were obtained from the education history data (combined for Round 3, 4 and 5, sent by GRADE); although the codes are not the same codes used in the School Survey, we can assign School Survey school codes to many of them based on matching child id codes with child id codes in the Pupil Roster of the Primary School Survey. Doing this increases the number of younger cohort children with school codes by 90, to 642. Thus, of the 1,848 for whom we hoped to obtain school id codes, a matching school code could be obtained for only 642, which is match rate of only 34.7%. However, recall that the school questionnaire was not administered in sites, 1, 2, 3, 4, 18 and 20. Ignoring these sites, there are 1,284 younger cohort children in the other 14 sites for whom we hoped to match to school data. This was done for 642, which is a match rate of 50.0%. While one would have liked this rate to be higher, recall that the schools surveyed were only a sample of the schools attended by the younger cohort children.

Finally, note that (unlike in Ethiopia, India and Vietnam) there appear to be no serious concerns about the accuracy of the matching. The school codes in the Pupil Roster of the School Survey should be for 2011 school year. The school codes from the Round 4 Education History were extracted for the 2011 school year. These two data sources can be merged by child ID codes. When this is done, there are 545 children who can be matched and who have school codes from both sources. Of these, almost all (539, a rate of 98.9%) have the same school codes from the two sources.

Vietnam

According to the Vietnam school sample documentation (James, 2013b), the primary school survey was conducted during the 2011-12 school year. It was designed to include all younger cohort children in grade 5 in that school year (more specifically in September of 2011). The plan was to include only schools that were physically within the geographic boundaries of the 20 sampling sites. The survey document says that 1,138 younger cohort children were included in the survey, but the pupil roster data include 1,323 children, all of whom are presumably younger cohort children since they all have a younger cohort ID codes.

While these 1,323 children can be matched to the household survey data from any of the rounds by using the child ID codes, there are also younger cohort children in other grades (e.g. grade 4) who attended these schools, as well as children in lower secondary grades (grades 6 and higher) that attended these primary schools in the past. To match these younger cohort children to the schools that were sampled, the Round 4 education history data were used, which contain school codes for the 2011-12 school year (and other years) for children who were enrolled in school in that school year. One problem with this approach is that the school codes in the primary school survey are different from the school codes in the Round 4 education history. Fortunately, one of the author team found an Excel file that linked these two sets of codes. Using this Excel file, it was possible to match 1,871 out of the 1,931 younger cohort children in Round 4 to a school code, a match rate of 96.9%. But not all of these are matched to a school in the school survey data, as explained below.

The Round 5 Vietnam data contain 1,940 children, and all 1,888 have math test scores. Of these, 1,864 are also found in the Round 4 education history data (2011-12 school year). Of these 1,864, we have school codes for 1,806, and we can match 1,662 to the primary school data. Thus the overall match rate for Round 5 children is 85.7% (1,662/1,940).

Finally, note that there are some concerns about the accuracy of the matching. In particular, when the school codes in the Round 4 education history for the 2011-12 school year are converted, using the file from Caine, into the school codes in the school survey, and those children are matched with the children in the pupil roster of the school survey, there are 1,021 matched children who have school codes from both the Round 4 education history and the school survey. Of these 1,021 younger cohort children, the school codes match for only 882 (86.4%).

Part of the reason for this discrepancy could be recall errors in the education history for the 2011-12 school year, which were collected in late 2013 and early 2014.

Appendix B: Econometrics of Using Site-Level Averages of School Characteristics

This appendix provides the econometric results used in the paper regarding regressions of student test scores on site-level averages of school characteristics.

The data contain students in schools, grouped by sites, with multiple schools in each site. Let y be the dependent variable, in this case student test scores: y_{isc} is the test score of student i in school s in site c ("c" is for community, that is a site, but subscript s is reserved for schools).

Consider a regression of students' test scores on their school characteristics. For simplicity, the focus is on one school characteristic, and all child- and household-level variables are ignored.

The key to understanding whether regressing student test scores on site-level averages of school characteristics yields biased estimates is to distinguish between the variance of school characteristics and the variance of site averages of school characteristics. Without loss of generality, let the school characteristic of school s in site c, denoted by D_{sc}, be the sum of a common site-level component, denoted by D_c, and the deviation from this component for school s in site c, which can be denoted as $D_{s|c}$:

$$D_{sc} = D_c + D_{s|c}, \qquad \text{where } E[D_{s|c} = 0]$$

Since $D_{s|c}$ is the deviation from the site-level component, its expected value given D_c can be set to 0: $E[D_{s|c}|\ D_c] = 0$. This is similar to a random effects (error components model) regression, except in this case D_c will generally have a positive expected value, rather than a zero expected value.

The next step is to compare $Var(D_{sc})$ and $Var(D_c)$. This is straightforward and is similar to the setup for a random effects regression. For simplicity, assume that $Var(D_{s|c})$ is the same for all sites, and note that $D_{s|c}$ is uncorrelated with D_c since $E[D_{s|c}|\ D_c] = 0$. Then:

$$Var(D_{sc}) = Var(D_c) + Var(D_{s|c})$$

Consider a simple regression of y_{isc} on a constant term and the school characteristic, D_{sc}. The regression equation, which shows the causal impact of D_{sc} on y_{isc}, is:

$$y_{isc} = \alpha + \beta D_{sc} + u_{isc}$$

Assume that u_{isc}, which represents all factors other than D_{sc} that determine y_{isc}, is uncorrelated with D_{sc}. As is well known, the OLS estimate of β, which can be denoted by $\hat{\beta}_{OLS}$ is equal to β:

$$\hat{\beta}_{OLS} = \frac{Cov(y_{isc}, D_{sc})}{Var(D_{sc})}$$

$$= \frac{Cov(\alpha+\beta D_{sc}+u_{isc}, D_{sc})}{Var(D_{sc})}$$

$$= \frac{\mathrm{Cov}(\alpha, D_{sc}) + \beta \mathrm{Cov}(D_{sc}, D_{sc}) + \mathrm{Cov}(u_{isc}, D_{sc})}{\mathrm{Var}(D_{sc})}$$

$$= \frac{\beta \mathrm{Var}(D_{sc})}{\mathrm{Var}(D_{sc})}$$

$$= \beta$$

What happens if y_{isc} is regressed on site-level means? Note that the site-level mean is *not* D_c, although its expected value is D_c. For simplicity, assume that there are S schools in each site; then the site level mean, which can be denoted as $\overline{D}_c$, is:

$$\overline{D}_c = (1/S)(D_c + D_{1|c} + D_c + D_{2|c} + \ldots + D_c + D_{S|c})$$

$$= D_c + (1/S)(D_{1|c} + D_{2|c} + \ldots + D_{S|c})$$

As long as we assume that both components of D_{sc}, D_c and $D_{s|c}$, are also uncorrelated with u_{isc}, and that $\mathrm{Cov}(D_{s|c}, D_{s'|c}) = 0$ for $s' \neq s$,[22] then OLS will estimate β as the covariance of y_{isc} and the right-hand side variable ($\overline{D}_c$) over the variance of that same right-hand side variable:

$$\hat{\beta}_{\text{OLS (site-level means)}} = \frac{\mathrm{Cov}(y_{isc}, \overline{D}_c)}{\mathrm{Var}(\overline{D}_c)}$$

$$= \frac{\mathrm{Cov}(\alpha + \beta\overline{D}_c + \beta D_{s|c} + u_{isc}, D_c + (1/S)(D_{1|c} + D_{2|c} + \ldots + D_{S|c}))}{\mathrm{Var}(D_c + (1/S)(D_{1|c} + D_{2|c} + \ldots + D_{S|c}))}$$

$$= \frac{\beta \mathrm{Cov}(D_c, D_c) + \beta(1/S)\mathrm{Cov}\left(D_{s|c}, D_{s|c}\right)}{\mathrm{Var}(D_c) + (1/S^2)\mathrm{Var}(D_{1|c} + D_{2|c} + \ldots + D_{S|c})}$$

$$= \frac{\beta \mathrm{Var}(D_c) + \beta(1/S)\mathrm{Var}\left(D_{s|c}\right)}{\mathrm{Var}(D_c) + (1/S^2)S\mathrm{Var}(D_{s|c})}$$

$$= \beta$$

Initially, this result may seem counterintuitive. If the site-level means deviate from the characteristics of students' school in a random way, one would expect this to lead to attenuation bias, underestimating β. The intuition for this result is that these deviations are not random; instead, the deviations, $D_{s|c}$, are correlated with the site-level mean, $\overline{D}_c$.

To see this, let e represent the deviation of $\overline{D}_c$, the site-level average, from D_{sc}, the value of the school characteristic for school s in site c, so that $\overline{D}_c = D_{sc} + e$. The deviation, e, is:

$$e = \overline{D}_c - D_{sc}$$

[22] That is, all correlation of D_{sc} for two schools in the same site is due to the common component D_c, which is the standard random effects regression assumption.

$$= D_c + (1/S)(D_{1|c} + \ldots + D_{s|c} + \ldots + D_{S|c}) - D_c - D_{s|c}$$

$$= (1/S)(D_{1|c} + \ldots + D_{s-1|c} + D_{s+1|c} + \ldots + D_{S|c}) - (\tfrac{S-1}{S})D_{s|c}$$

This deviation (measurement error), e, is clearly correlated with the school characteristics, D_{sc}, via the component $D_{s|c}$. More specifically:

$$Cov(D_{sc}, e) = Cov(D_c + D_{s|c}, (1/S)(D_{1|c} + \ldots + D_{s-1|c} + D_{s+1|c} + \ldots + D_{S|c}) - (\tfrac{S-1}{S})D_{s|c})$$

$$= Cov(D_{s|c}, -(\tfrac{S-1}{S})D_{s|c})$$

$$= (\tfrac{1-S}{S})Var(D_{s|c})$$

Since S > 1 when there is more than one school in a site, this covariance is < 0.

It is straightforward to show that the effective error term in a simple regression of y_{isc} on the site-level average is uncorrelated with that average, and thus yields an unbiased estimate of β. The regression equation in this case is:

$$y_{isc} = \alpha + \beta D_{sc} + u$$

$$= \alpha + \beta(\overline{D}_c - e) + u$$

$$= \alpha + \beta\overline{D}_c + (u - \beta e)$$

If $\overline{D}_c$ is uncorrelated with the "effective" error term, (u – βe), than OLS regression of y_{isc} on $\overline{D}_c$ yields an unbiased estimate of β. This lack of correlation is easily shown, continuing to assume that u is uncorrelated with both $\overline{D}_c$ and e:

$$Cov(\overline{D}_c, u - \beta e)$$

$$= Cov(D_{sc} + e, u - \beta e)$$

$$= -\beta Cov(D_{sc} + e, e)$$

$$= -\beta\{Cov(D_{sc}, e) + Var(e)\}$$

$$= -\beta\{(\tfrac{1-S}{S})Var(D_{s|c}) + Var(\tfrac{1}{S}(D_{1|c} + \ldots + D_{s-1|c} + D_{s+1|c} + \ldots + D_{S|c}) - (\tfrac{S-1}{S})D_{s|c})\}$$

$$= -\beta\{(\tfrac{1-S}{S})Var(D_{s|c}) + (\tfrac{S-1}{S^2})Var(D_{s|c}) + (\tfrac{(S-1)^2}{S^2})Var(D_{s|c})\}$$

$$= -\beta\{(\tfrac{1-S}{S})Var(D_{s|c}) + (\tfrac{S(S-1)}{S^2})Var(D_{s|c})\}$$

$$= 0$$

Random measurement error in D_{sc}. Finally, consider the situation where there is random measurement error in the school characteristic, D_{sc}. For each school, the *observed* value of D_{sc}, which can be denoted as $D_{sc,o}$, is the sum of the "true" value and a random measurement error, denoted by ε_{sc}:

$$D_{sc.o} = D_{sc} + \varepsilon_{sc}$$

Assume that all measurement errors are drawn from a distribution with zero mean and variance Var(ε), and that these measurement errors are uncorrelated with each other, even within the same site, and are uncorrelated with all other observed and unobserved variables.

Consider the bias when y_{isc} is regressed on the observed school-level variable; the estimated β is denoted by $\hat{\beta}_{OLS,me}$.

$$\hat{\beta}_{OLS,me} = \frac{\mathrm{Cov}(y_{isc}, D_{sc,o})}{\mathrm{Var}(D_{sc,o})}$$

$$= \frac{\mathrm{Cov}\left(\alpha+\beta D_{sc}+u_{isc}, D_{sc,o}\right)}{\mathrm{Var}\left(D_{sc,o}\right)}$$

$$= \frac{\mathrm{Cov}(\alpha+\beta D_{sc}+u_{isc}, D_{sc}+\varepsilon_{sc})}{\mathrm{Var}(D_{sc}+\varepsilon_{sc})}$$

$$= \frac{\mathrm{Cov}(\alpha, D_{sc}) + \beta\mathrm{Cov}(D_{sc}, D_{sc}) + \mathrm{Cov}(u_{isc}, D_{sc})}{\mathrm{Var}(D_{sc})+\mathrm{Var}(\varepsilon_{sc})} \quad \text{(since } \varepsilon_{sc} \text{ is uncorrelated with everything)}$$

$$= \frac{\beta\mathrm{Var}(D_{sc})}{\mathrm{Var}(D_{sc})+\mathrm{Var}(\varepsilon)}$$

$$< \beta$$

This is the standard attenuation bias due to random measurement error.

Next, what happens when y_{isc} is regressed on site-level means of D_{sc}? In this case, the observed site-level mean, denoted by $\overline{D}_{c,o}$ is:

$$\overline{D}_{c,o} = D_c + (1/S)(D_{1|c} + D_{2|c} + \ldots + D_{S|c}) + \bar{\varepsilon}_c,$$

where $\bar{\varepsilon}_c$ is the average of ε_{sc} in site c.

For future reference, the assumption that ε is uncorrelated within any given site implies that:

$$\mathrm{Var}(\bar{\varepsilon}_c) = (1/S)\mathrm{Var}(\varepsilon)$$

where S is the number of schools in the community, which for simplicity is assumed to be the same for all communities.

OLS regression of y_{isc} on $\overline{D}_{c,o}$ will yield an estimate of β, denoted by $\hat{\beta}_{OLS,me\ (site\text{-}level\ means)}$, that is the covariance of y_{isc} and $\overline{D}_{c,o}$ over the variance of $\overline{D}_{c,o}$:

$$\hat{\beta}_{OLS,me\ (site\text{-}level\ means)} = \frac{\mathrm{Cov}(y_{isc}, \overline{D}_{c,o})}{\mathrm{Var}(\overline{D}_{c,o})}$$

$$= \frac{\mathrm{Cov}(\alpha+\beta D_{sc}+u_{isc}, D_c + (1/S)(D_{1|c} + D_{2|c} + \dots + D_{S|c}) + \bar{\varepsilon}_c)}{\mathrm{Var}(D_c + (1/S)(D_{1|c} + D_{2|c} + \dots + D_{S|c}) + \bar{\varepsilon}_c)}$$

$$= \frac{\beta \mathrm{Cov}(D_c + D_{s|c}, D_c + (1/S)(D_{1|c} + D_{2|c} + \dots + D_{S|c}) + \bar{\varepsilon}_c)}{\mathrm{Var}(D_c) + (1/S^2)\mathrm{Var}(D_{1|c} + D_{2|c} + \dots + D_{S|c}) + \mathrm{Var}(\bar{\varepsilon}_c)}$$

$$= \frac{\beta[\mathrm{Cov}(D_c, D_c) + (1/S)\mathrm{Cov}(D_{s|c}, D_{s|c})]}{\mathrm{Var}(D_c) + (1/S)\mathrm{Var}(D_{s|c}) + (1/S)\mathrm{Var}(\varepsilon)}$$

$$= \frac{\beta[\mathrm{Var}(D_c) + (1/S)\mathrm{Var}(D_{s|c})]}{\mathrm{Var}(D_c) + (1/S)\mathrm{Var}(D_{s|c}) + (1/S)\mathrm{Var}(\varepsilon)}$$

$$< \beta$$

This regression also underestimates β.

Which regression has a larger bias? To see, compare the two terms multiplied by β.

School level bias – Site level

$$= \frac{\mathrm{Var}(D_{sc})}{\mathrm{Var}(D_{sc})+\mathrm{Var}(\varepsilon)} - \frac{\mathrm{Var}(D_c) + (1/S)\mathrm{Var}(D_{s|c})}{\mathrm{Var}(D_c) + (1/S)\mathrm{Var}(D_{s|c}) + (1/S)\mathrm{Var}(\varepsilon)}$$

$$= \frac{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c})}{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c}) + \mathrm{Var}(\varepsilon)} - \frac{\mathrm{SVar}(D_c)+\mathrm{Var}(D_{s|c})]}{\mathrm{SVar}(D_c) + \mathrm{Var}(D_{s|c}) + \mathrm{Var}(\varepsilon)}$$

$$= \frac{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c})}{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c}) + \mathrm{Var}(\varepsilon)} - \frac{(S-1)\mathrm{Var}(D_c)+\mathrm{Var}(D_c)+\mathrm{Var}(D_{s|c})]}{(S-1)\mathrm{Var}(D_c) + \mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c}) + \mathrm{Var}(\varepsilon)}$$

$$< 0$$

The last term is < 0 because $\frac{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c})}{\mathrm{Var}(D_c) + \mathrm{Var}(D_{s|c}) + \mathrm{Var}(\varepsilon)}$ is clearly < 1, and adding (S-1)Var(D_c) to both the numerator and the denominator moves it closer to 1, so the site-level bias term is closer to 1 than the school-level bias term.

Thus, with random measurement error, the school-level regression underestimates β more than the site-level regression. The intuition for this is that the site mean of school characteristics includes the mean of the measurement errors, which tend to cancel each other out.