Do Schools Reinforce or Reduce Learning Gaps between Advantaged and Disadvantaged Students? Evidence from Vietnam and Peru

PAUL GLEWWE
University of Minnesota

SOFYA KRUTIKOVA
Institute for Fiscal Studies, London

CAINE ROLLESTON
Institute of Education, University College London

I. Introduction

Investing in education increases economic growth and, at the individual level, raises incomes and living standards (Krueger and Lindahl 2001). Developing countries have achieved substantial increases in primary and secondary enrolment rates in the last 25 years (Glewwe et al. 2013), yet students in many developing countries learn much less than students in developed countries (Glewwe and Kremer 2006). Equally disturbing is that, within developing countries, some children learn much more than others, and in some countries schools appear to widen learning gaps between advantaged and disadvantaged students (Banerjee and Duflo 2011). While eliminating inequality in student learning is probably impossible, there may be scope for reducing within-country learning gaps by implementing education policies that disproportionately benefit disadvantaged students.

The existing literature on learning gaps focuses on whether disadvantaged students attend lower quality schools (McEwan 2004; McEwan and Trowbridge 2007). Yet recent evidence (reviewed below) finds large learning differences between children of different backgrounds even within the same schools. This article uses panel data from Vietnam and Peru to investigate intraschool differences in learning. It uses six different advantage indicators to assess whether advantaged students learn more than disadvantaged students when they attend the same school. Comparison of Vietnam and Peru is of particular interest given their starkly different scores on the 2012 Program for International Student Assessment (PISA; OECD 2013); of the 65 countries assessed, Viet-

Electronically published May 24, 2017

nam ranked seventeenth in mathematics and nineteenth in reading (ahead of the United States and the United Kingdom), while Peru ranked last (sixty-fifth) in both subjects. This difference is particularly striking given that Peru's income per capita in 2012 ($6,060) was four times larger than Vietnam's ($1,550; World Bank 2013).

More specifically, this article estimates the combined impacts of all school and teacher characteristics on student learning by estimating separate fixed effects for advantaged and disadvantaged students in each school. For each country, the relative sizes of the two fixed effects are compared for six types of advantage (being wealthy, having higher skills at age 5, being male, being from the ethnic majority group [Peru only], higher maternal education, and better nutrition in early childhood) to examine whether advantaged children learn more than otherwise comparable disadvantaged children in the same schools.

This article advances knowledge on student learning in developing countries in three ways. First, the rich data set, which links detailed individual-level panel data from early childhood to very detailed school surveys, reduces omitted-variable bias. Second, separate school fixed effects for advantaged and disadvantaged children allow for within-school heterogeneity in the impact of school and teacher variables on learning, which is usually ignored in the literature but may be an important driver of inequality. Third, using very similar data for two very different countries allows for cross-country heterogeneity in whether schools raise or lower student learning gaps.

The article is organized as follows. Section II reviews the relevant literature. Section III describes the Vietnam and Peru school systems, and Section IV presents the methodology. The next three sections describe the data, discuss estimation issues, and present results, respectively. Section VIII provides robustness checks and an extension to the analysis. Section IX concludes.

II. Literature Review

Gaps in education outcomes between children from more and less socioeconomically advantaged backgrounds are widely documented. Filmer and Pritchett (1999) use data from 35 developing countries to show that most shortfalls in basic education indicators (e.g., enrollment and dropout) are due to children in the bottom 40% of the wealth distribution. Indeed, the link between socioeconomic background and schooling has been found in both countries examined in this article: Vietnam (Glewwe and Jacoby 2004) and Peru (Jacoby 1994).

A large literature considers why children from advantaged backgrounds have better educational outcomes. This literature focuses mainly on differences at home and in school, which can affect learning through many mechanisms, including the availability of good quality schools, household demand for and attitudes toward education, and children's learning capacity, which may be reduced by their exposure to adversity (Duncan and Murnane 2011).

Home factors that seem to matter include income, parental education, and early childhood nutrition (Glewwe and Miguel 2008; Behrman 2010; Alderman and Bleakley 2013). There is evidence that school and teacher quality affect education outcomes in many contexts (Aaronson, Barrow, and Sander 2007; Altonji and Mansfield 2011; Lai, Sadoulet, and de Janvry 2011), yet the available evidence is inconclusive on exactly what constitutes school quality. Glewwe et al. (2013), reviewing 20 years of studies on developing countries, conclude that much of what matters is unlikely to be easy to observe. More recent work suggests that teacher behaviors collectively described as "responsive teaching" (Hamre et al. 2014) are strongly associated with increased learning (Araujo et al. 2016).

Whatever school quality is, the literature finds that access to it varies by children's socioeconomic background, and this explains much of the differences in their educational outcomes. For example, UK and US studies find that learning gaps between more and less advantaged children widen during their time in school and that differences in school quality explain part of this trend (Currie and Thomas 2001; Fryer and Levitt 2004). There is also developing country evidence: two recent studies find that a significant portion (e.g., 50%–69% in Guatemala) of the learning gap between indigenous and nonindigenous children is due to differences in school quality in Latin American countries (McEwan 2004; McEwan and Trowbridge 2007).

Much of the literature focuses on differences between schools attended by advantaged and disadvantaged children. Our article adds to a smaller literature on within-school differences in learning between children of different backgrounds. Banerjee and Duflo (2011) suggest that this leads to inequality in educational outcomes in many developing countries. We know of no studies that directly examine the effect of within-school differences in the educational experiences of children from different backgrounds on learning gaps, yet there is evidence of within-school differences in treatment of children of different backgrounds that is consistent with such an effect.

For example, Hanna and Linden (2012) randomly assigned child characteristics (gender, age, and caste) to cover sheets of pupil exams in India. Teachers recruited to grade those exams, who could see the characteristics from the sheets, gave low-caste children lower scores. Botelho, Madeira, and Rangel (2015) found similar discrimination by race in Brazil. In the United States, Dee (2005) showed that racial and ethnic distance between teachers and pupils affects teachers' perceptions of pupil performance, especially of disadvantaged students. Other studies find that discrimination may lower pupils' performance via "stereotype" threat. Hoff and Pandey (2006) show that a caste gap emerges in India when school-age task participants' caste is revealed. This supports the emphasis on identity by Akerlof and Kranton (2010), who argue that within a school the dominant group's norms can reduce the performance of "outsiders." These findings motivate our focus on within-school differences in learning among children from advantaged and disadvantaged backgrounds.

III. Overview of the School Systems in Vietnam and Peru

The education systems of Vietnam and Peru provide an interesting contrast. Enrollment has increased in both countries, but Vietnam's expansion has been much more equitable.

Primary school has long been compulsory in both countries. In Vietnam, the 1991 Law on Universal Primary Education made primary education (grades 1–5) compulsory for 6–14-year-old children. Universal primary education was essentially achieved by 2010, with a net enrollment rate of 98% (World Bank 2012). Peru's 2003 General Education Law also stipulated compulsory basic education, and its net enrollment rate was 95% in 2010 (World Bank 2012).

Education quality is a focus in both counties. Vietnam has invested heavily in primary school facilities, teacher training, and curricular and textbook reform and raised hours per day in primary school, which were relatively low. Vietnam's high rank in the 2012 PISA tests suggests that its efforts succeeded in raising school quality; Peru's PISA results indicate less success.

Evidence also suggests large inequities in educational provision in both countries. Vietnam has introduced several new programs to ensure "minimum standards" of quality to all, especially in disadvantaged areas (World Bank 2004; UNESCO 2011). Although concern remains in Vietnam about growing inequity (World Bank 2011), comparisons with other developing countries suggest that Vietnam's educational outcomes are quite equitably distributed (Holsinger 2005). In contrast, in Peru there are large gaps in pupils' educational outcomes by socioeconomic group, ethnicity, and location (Murray 2012). Indeed, Peru had the largest ratio of performance at the 95th percentile over performance at the 5th percentile of all 43 countries participating in the 2000 PISA assessment (Crouch, Gustafsson, and Lavado 2009).

These differences between Vietnam and Peru suggest that their education systems have different effects on learning gaps between advantaged and disadvantaged students. We use comparable data to examine these effects for a relatively high-performing, and equitable, education system (Vietnam) and a more unequal, less well performing system (Peru).

IV. Methodology

This article examines whether the impact of school characteristics on learning differs within schools between advantaged and disadvantaged pupils. This section presents the equations to be estimated. Data and estimation strategy are then discussed in Sections V and VI.

To begin, consider a linear approximation of a general production function for cognitive skills (C) of primary school students:

$$\begin{aligned} C = \beta_0 + \beta_1 C_p + \beta_2 N_p + \beta_3 N + \beta_4 \text{PE} + \beta_5 \text{PT} \\ + \beta_6 \text{EI}_3 + \beta_7 \text{IA} + \boldsymbol{\gamma}_1' \mathbf{SC} + \boldsymbol{\gamma}_2' \mathbf{TC} + u. \end{aligned} \tag{1}$$

In this equation, C_p is “preschool” cognitive skills (skills when starting primary school at age 5); N_p is “preschool” nutritional status (nutrition from birth to age 5); N is current nutritional status; PE is parental education; PT is parental time spent with the child during primary school on activities that raise cognitive skills; EI is purchased educational inputs such as textbooks, school supplies, and tutoring; IA is the child’s innate ability; and **SC** and **TC** are vectors of school and teacher characteristics, repectively. The residual, u, accounts for errors due to the linear approximation and measurement errors in S and is assumed to be uncorrelated with the explanatory variables.

Equation (1) assumes linear separability between household/child characteristics and teacher/school characteristics; interactions between household and child variables and school and teacher variables are assumed to be negligible. This is a strong assumption, yet studies of student learning in developed and developing countries often invoke it (e.g., Rockoff 2004; Chay, McEwan, and Urquiola 2005; Rivkin, Hanushek, and Kain 2005; Rothstein 2010). We will relax this assumption by allowing the effects of teacher/school variables to differ across advantaged and disadvantaged students.

Glewwe et al. (2013) explain how difficult it is to estimate the impacts of school and teacher characteristics on student learning. Fortunately, for the main objective of this article—to assess whether schools reinforce or reduce learning gaps between advantaged and disadvantaged students—one can simply use a summary measure of school quality that captures both easy and difficult (if not impossible) to measure school and teacher characteristics. More specifically, one can combine the impacts of school and teacher variables in equation (1) into school fixed effects:

$$\begin{aligned} C = \beta_1 C_p + \beta_2 N_p + \beta_3 N + \beta_4 \text{PE} + \beta_5 \text{PT} \\ + \beta_6 \text{EI} + \beta_7 \text{IA} + \sum_{s=1}^{S} \delta_s D_s + u, \end{aligned} \tag{2}$$

where $\delta_s = \beta_0 + \boldsymbol{\gamma}_1'\mathbf{SC}_s + \boldsymbol{\gamma}_2'\mathbf{TC}_s$, and D_s is a dummy variable for school s. Note that each δ_s includes β_0; one can estimate only the relative, not the absolute, impact of each school on student learning. Two notable features of equation (2) are that (i) no restrictions were imposed on (1) to obtain (2)—the latter simply converts each school's $\boldsymbol{\gamma}_1'\mathbf{SC} + \boldsymbol{\gamma}_2'\mathbf{TC}$ term into a school fixed effect—and (ii) these fixed effects include all possible interactions between the **SC** and **TC** variables.

The disadvantage of estimating (2) instead of (1) is that (2) does not reveal which school characteristics increase student learning. Yet the advantage is that the fixed effect captures both observed and unobserved school variables that affect learning and so can be used to assess the role of school (and teacher) characteristics in explaining that outcome. This method has been used to study both school and teacher quality (Rockoff 2004; Rivkin et al. 2005; Aaronson et al. 2007).[1]

To date, studies that use school fixed effects assume that there are no interactions between (observed or unobserved) school variables and student and household variables. Thus, this method assumes that within schools all students benefit equally from what schools offer regardless of student characteristics. To our knowledge, this assumption has never been relaxed. Yet the evidence discussed in Section II of differential performance among students in the same schools because of caste or race discrimination, or the effects of stereotype threat, belies this assumption.

One way to allow the impacts of school and teacher characteristics to differ among students in the same school is to divide students in each school into two or more groups and allow each group to have a distinct school fixed effect. In particular, if students can be classified as "advantaged" and "disadvantaged" one can rewrite (2) as

$$\mathrm{C}=\boldsymbol{\beta}'\mathbf{X} + \sum_{s=1}^{S}\kappa_s D_s + \sum_{s=1}^{S}\theta_s D_s A + u_3, \tag{3}$$

where A is a dummy variable indicating "advantaged" students and, to reduce clutter, $\boldsymbol{\beta}'\mathbf{X}$ denotes $\beta_1 C_p + \beta_2 N_p + \beta_3 N + \beta_4\mathrm{PE} + \beta_5\mathrm{PT} + \beta_6\mathrm{EI} + \beta_7\mathrm{IA}$. The κ_s term measures the impact of school s on disadvantaged students, and $\kappa_s + \theta_s$ measures its impact on advantaged students. If schools affect the learning of both types of students equally, then $\theta_s = 0$. The goal of this article is to test whether the value of θ_s, averaged across the schools in the sample, is equal to zero.

[1] Note that, unlike Aaronson et al. (2007), our goal is not to estimate the variance of the distribution of school fixed effects; instead, it is to estimate the mean of a particular type of school fixed effect.

We use six different definitions of advantage: being wealthy, having high skills at school starting age, being male,[2] belonging to the dominant ethnic group,[3] having an educated mother, and adequate nutrition in early childhood. How they are measured is discussed in the next section.

To clarify what we intend to estimate, consider the ideal “experiment” that would yield such an estimate. Assume that students are randomly “labeled” as advantaged or disadvantaged conditional on their values of **X**. These labels may be correlated with some **X** variables—for example, wealthy parents may help their children more with schoolwork (an **X** variable)—but conditional on **X** the labels are randomly assigned and so do not predict student performance unless teachers or schools, intentionally or unintentionally, are more effective for students with a particular label. If no differential effectiveness exists, estimated school fixed effects conditional on **X** will not differ between students randomly labeled as advantaged or disadvantaged: θ_s should be 0. Yet if differential effectiveness does exist, estimated school fixed effects conditional on **X** will differ across advantaged and disadvantaged students: $\theta_s \neq 0$.

This “experiment” closely resembles that of Hanna and Linden (2012), yet our data are not from a randomized controlled trial. Hanna and Linden were able to randomize only favoritism in grading exams, not favoritism, or more generally differential effectiveness, in daily classroom interactions with students, and it would be very difficult to randomize labeling to detect this more general differential effectiveness. While we cannot use a randomized controlled trial to investigate such general favoritism, an advantage we have is unusually detailed data that allow us to control for a wide variety of student and household variables, including cognitive skills at age 5 (before starting school), which could account for many difficult to measure factors such as innate ability (IA) and many types of parental time (PT) and educational inputs (EI).

V. Data

This section presents the data we use and discusses the definitions of advantage.

A. Survey and Sampling

This article uses data from Young Lives, a multicountry longitudinal study of child poverty in four developing countries: Ethiopia, India, Peru, and Viet-

[2] While some evidence suggests that girls outperform boys in reading (e.g., Guiso et al. 2008), other studies (e.g., Fryer and Levitt 2010) find a significant, persistent gender gap favoring boys in math in both developed and developing countries. However, there is little evidence, at least in the United States, of a gender gap before starting school.

[3] We use this definition only for Peru, as only 8% of the Vietnam school survey subsample are ethnic minorities.

nam. Each country collects data on two cohorts of children: 1,000 children born in 1994–95 and 2,000 born in 2000–2001. Our study uses data only from the latter cohort, known as the "younger cohort." In each country, the 2,000 younger cohort children were sampled by randomly selecting 100 children who were 6–18 months old in 2001–2 from each of 20 sites. The data are statistically representative of the site level populations but are not nationally representative. However, the 20 sites in each country were purposively selected to represent each country's socioeconomic and geographic diversity (except for the wealthiest areas).

Three rounds of household-level data were collected in 2002, 2006–7, and 2009. School-level data were collected in 2011 for subsamples of the younger cohort children.[4] The 2011 school surveys differ somewhat by country, reflecting differences in school systems and policies. This study uses both the household and school survey data, focusing on Peru and Vietnam because of the availability of recently collected school survey data.[5]

The Vietnam school survey was conducted in October 2011, in all 20 sites. The school survey subsample, the focus of this article, consists of all younger cohort children enrolled in grade 5 (the appropriate grade for their age) in the 2011–12 school year; all schools attended by these children were surveyed. This subsample consists of 1,138 children in 92 primary schools.

The Peru school survey was conducted from October to December 2011. The sampling was stratified, with four school types as strata so that each type—private urban, public urban, public rural bilingual, and public rural nonbilingual (Spanish only)—is represented. A random subsample was drawn of younger cohort children in each of these four school types. The school sample consists of these children's schools; all younger cohort children in them were surveyed. This subsample has 572 children in 132 schools. Most were in grade 4 (59%), while 32% were in grade 5, and 9% in grade 3. After dropping a few observations with missing data, the final samples are 547 children in 132 schools in Peru and 1,129 children in 90 schools in Vietnam.

B. Testing

Child learning is measured using scores on curriculum-based math and reading comprehension tests administered as part of the school surveys. In Viet-

[4] In 2011, these children were about 10 years old. They typically started primary school in 2006. Note that a fourth round of data was collected in 2013 and has just become available; this round was not available when this article was written.

[5] School data were not available for Ethiopia when this analysis was done; school data were available for India but because of the large number of small schools attended by the younger cohort children, the data are not well suited to identifying school fixed effects.

nam, each test had 30 multiple-choice items that reflected the grade 5 curriculum; they are similar to the Ministry of Education's grade 5 assessment tests (World Bank 2011). In Peru, grade-specific math and reading tests with 30–35 items were developed by GRADE (Grupo de Análisis para el Desarrollo), a Peruvian research institute, to reflect the curricula in grades 3–5.[6] The tests were designed to reflect each country's curriculum and so are not comparable across countries.[7]

We also use two measures of cognitive ability from assessments administered when the children were age 5. These are the Cognitive Development Assessment (CDA) and the Peabody Picture Vocabulary Test (PPVT). Both tests have been validated in many contexts and are highly correlated with broad-based IQ measures (Cueto et al. 2009).

The CDA was developed by the International Evaluation Association to assess the cognitive development of 4-year-old children. The subscale administered to the Young Lives children assesses perceptions of quantity by testing children's understanding of concepts such as few, most, half, many, equal, and pair with statements such as "Point to the plate that has *few* cupcakes." The subscale has 15 items, and each correct answer is scored as 1 point, so that the minimum number of points a child can get is 0 and the maximum is 15.

The PPVT is a test of child receptive vocabulary, which measures vocabulary acquisition.[8] The test taker is asked to select the picture that best represents the meaning of a stimulus word presented orally by the examiner. Test items were arranged in order of increasing difficulty, and only the items within the critical range of the specific child were administered to each child, as determined by a well-defined procedure followed by the interviewer. The test score is calculated as the difference between the ceiling item (e.g., word number 78) and the total number of errors. Official test manuals include tables for standardization of the raw scores. However, these standardization procedures were not followed within the Young Lives study, as the standardization samples have characteristics different from the Young Lives samples.[9]

[6] The English versions of the complete set of tests and questionnaires used in the Vietnam and Peru school surveys are available on the Young Lives website: http://www.younglives.org.uk/content/school-survey.

[7] The curriculum itself can induce gaps between advantaged and disadvantaged students; we assess whether the entire school system generates such gaps and leave to future research which aspects of the systems produce any gaps we find.

[8] Specifically, the versions administered in Peru and Vietnam were the Spanish version of the PPVT-R (Dunn et al. 1986) and a translation into Vietnamese of the PPVT-III (Dunn and Dunn 1997), respectively. Both versions measure the same underlying construct and follow the same principle, although the PPVT-R has fewer items than the PPVT-III (125 and 204, respectively).

[9] For a more detailed description of the CDA and PPVT tests administered within Young Lives, see Cueto et al. (2009).

The scores for all of the tests that we use in the analysis were generated using Item Response Theory (IRT) models. IRT has a long history in education and psychometrics. For instance, it is applied to generate internationally comparable scores for tests such as PISA and TIMSS (Trends in International Mathematics and Science Study), but it is still not widely applied in economics. The main reason that we use IRT in this article is that it offers a less arbitrary way to construct the skill measure than, for instance, summing up correct responses. Further, it offers a way to construct comparable measures across grades.[10] IRT studies tests at the item (test question) level, modeling the probability that an individuals with given ability will correctly answer an item. The overall estimate of this ability is generated by analyzing an individual's response to many different items, each defined by their own characteristics. The estimate of the ability parameter maximizes the posterior probability of a person's whole set of responses. See Das and Zajonc (2010) for further details.

C. *Advantage*

The first of the six advantage categories used in this article is being wealthy, defined as living in a household in the top two wealth quintiles of a wealth index that combines measures of housing quality, ownership of consumer durables, and access to key services.[11] Having high skills at school starting age, the second advantage category, is defined as performing in the top 40% on the CDA test, which measures children's basic quantitative skills at age 5 (the school starting age in both countries is 6). Robustness checks in Section VIII show that the main findings hold if other cutoff points are used for the wealth and skill advantage classifications.

Being male is the third advantage type. The fourth, ethnicity, is used only for Peru; only 8% of the sample in the Vietnam school survey are ethnic mi-

[10] This is particularly relevant for Peru, where the school survey sample includes children from multiple grades. The tests for children in different grades were designed to have common items that are used to create scores that are comparable across grades using IRT.

[11] We use the wealth index constructed and publicly archived by the Young Lives team. It draws on work undertaken by the World Bank and Macro International. It is a simple average of three separate indexes that range from 0 to 1: housing quality, consumer durables, and access to services. The housing quality index is a mean of (1) rooms per person (number of rooms divided by number of household members), set to take a maximum value of 1; (2) floor quality (a dummy variable that takes the value of 1 if the floor is made of finished material); and (3) roof quality (a dummy variable that takes the value of 1 if the roof is made of iron, concrete tiles, or slate. The consumer durables index is the proportion of durables a household owns from a list of seven (radio, motorbike/scooter, bicycle, TV, motorized vehicle or truck, landline telephone, and modern bed or a table). The services index is the proportion of key services that a household has access to: electricity, piped water, own pit latrine/flush toilet, and modern cooking fuel (gas, kerosene, or electricity).

norities.[12] Ethnicity is a sensitive issue in Peru. Thus, in the Young Lives sample the great majority of mothers identify their children as mestizo.[13] For the purpose of identifying disadvantage, a more accurate definition of ethnicity is the mother's first language; it is an ethnicity marker and so is associated with exclusion in Peru (Escobal, Saavedra, and Vakis 2012). Thus, children are classed as being an ethnic minority if their mother's first language is not Spanish (but rather Quechua, Aymara, or another indigenous language).

The last two definitions of advantage are having an educated mother and adequate nutrition in infancy. The former is defined as having a mother who completed at least primary school. The latter is defined as not being stunted, that is, having a height-for-age z-score greater than -2, at age 1.

While these types of advantage are correlated, they overlap only partially (see appendix tables A1 and A2). Table 1 shows that they represent advantage in terms of skill acquisition; for all six types, the advantaged group almost always has significantly higher scores on cognitive tests (the CDA and the PPVT, both described above) taken at age 5, before the children started school, suggesting that advantaged children enjoy favorable conditions for acquiring skills before starting primary school. Of course, these are unconditional means, and so some of these differences may not persist after conditioning on other variables.

Equation (3) can be estimated only for schools with at least one advantaged child and at least one disadvantaged child in our sample. Thus, the estimation samples vary by the definition of advantage; table A3 shows the number of observations and schools in each of these estimation samples.[14] In the Vietnam sample, of the 90 schools, 82 are attended by two or more sample children. Of these 82, 48 have sample children from both the poorer and the richer groups, 61 include both more able and less able children, 77 have both boys and girls, 61 have children with both more and less educated mothers, and 63 have children who were and were not malnourished at age 1. Of the 132 schools in Peru, 80 have two or more sample children. Of these 80, 36 have both richer and poorer children, 47 have both more able and less able children, 63 have both boys and girls, 25 have both ethnic minority and ethnic majority children, 48 have children with both more educated and less educated mothers, and 54

[12] Ethnicity is not used as an advantage category for Vietnam, yet a variable indicating an ethnic majority group child is used as a control in the estimations. It is defined as being Kinh (i.e., not H'mong, Dao, Tay, or other).

[13] Meaning a person of combined European and indigenous descent.

[14] Perhaps the precision of estimates for the nonschool variables would be higher, and bias in the school effects lower, if we used the whole sample (not just the "overlap" subsample of schools that have at least one advantaged child and one disadvantaged child), still estimating the school effects only for the "overlap" subsample. Yet the estimates of differences in school effects change little when the whole sample is used (results available on request).

TABLE 1

DIFFERENCE IN PRESCHOOL TEST SCORES BETWEEN MORE AND LESS ADVANTAGED CHILDREN (ADVANTAGED − DISADVANTAGED)

	Vietnam		Peru	
Advantage Category	PPVT (1)	CDA (2)	PPVT (3)	CDA (4)
Being richer (top 2 wealth quintiles)	21.89***	13.29***	24.80***	30.94***
Being more able (top 2 CDA quintiles)	30.06***	77.35***	11.23***	76.39***
Being male	5.99**	4.13*	−2.29	−1.92
Being an ethnic majority	NA	NA	21.17***	15.19***
Having a mother with at least completed primary schooling	16.68***	6.33*	22.81***	18.17***
Having not been malnourished in infancy	10.74***	2.82	13.91***	−.53

Note. The Peabody Picture Vocabulary Test (PPVT) is a measure of receptive vocabulary. The Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. The difference in CDA scores between more and less able children is there by construction since being more able is defined as having a CDA score in the top two quintiles of the distribution at age 5. Being richer: wealth quintiles are constructed using a wealth index that is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). Being an ethnic majority: Peru—mother's first language is Spanish (i.e., not Quechua, Aymara, or other). Not being malnourished in infancy: defined using anthropometric indicator of adequate long-term growth (height-for-age *z*-scores > −2).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

have children who were and were not stunted at age 1. Limiting the analysis to schools with both advantaged and disadvantaged sample children raises issues of selection bias in the estimates of θ_s; we return to this issue in Section VIII.

VI. Estimation

Equation (3) is estimated using Young Lives' child, parent, and household characteristics data. Skills acquired before entering primary school (C_p) are measured by scores on the CDA and PPVT tests in round 2, when the children were 5 years old. Both tests have been validated in many contexts and are highly correlated with broad-based IQ measures, which suggests that they can also serve as controls for innate ability (IA). Nutritional status before entering school (N_p) is measured by maternal assessments of the child's size at birth (five-point scale, very small to very large) and of the child's health at age 1 relative to that of other children, as well as the height-for-age *z*-score measured at age 5 (in round 2). Current nutritional status (N) could, in principle, be measured by the height-for-age *z*-score in round 3, but it is highly correlated with round 2 height for age (0.80 for Vietnam, 0.79 for Peru), so it is excluded.

Parental education (PE) in equation (3) is measured by dummy variables for different levels of education, separately for mothers and fathers, with no education as the base category. Parental time with the child (PT) is measured by a dummy variable indicating whether someone outside the household cared for the child (including attending a crèche) in round 1 (when the child was about 1 year old), the mother's mental health in round 1,[15] a dummy variable for whether the child is an only child, and household size. Educational inputs (EI) are measured by time in preschool and several asset variables (the wealth index, land owned, and per capita consumption) that should be highly correlated with unobserved educational inputs.[16]

A few control variables are also added to equation (3): age when the test was taken and dummy variables for being in grades 3 and 4 (Peru only), for ethnic minority, and for being male. Finally, school identifiers from the school survey are used to generate school fixed effects; these are interacted with the relevant advantage dummy variable to estimate θ_s.

Table 2 presents summary statistics for all these variables. The children were about 10 years old when they took the math and verbal tests used as outcome measures. Almost all test scores were transformed using a three-parameter IRT model, with means and standard deviations standardized to 500 and 100, respectively.[17] The exception is that the CDA and PPVT scores were standardized (on the full Young Lives sample) to have means of 300 and standard deviations of 50. Finally, child nutrition is measured by height-for-age *z*-scores using WHO standards.

A number of estimation issues require further discussion. First, recall from Section IV the "thought experiment" underlying our analysis. It is straightforward when disadvantage is defined as ethnicity or gender: conditional on $\mathbf{X}$, ethnicity and gender should be considered as randomly assigned labels that have no predictive power for educational outcomes unless teachers or schools, intentionally or unintentionally, are differentially effective for students with those labels. Even if girls tend to have better language skills, and boys better math skills, after controlling for skills at age 5 there should be little difference in the learning of boys and girls.[18]

[15] Maternal mental health is measured using the Self-Reporting Questionnaire; see the table 2 note for details.

[16] This means that (2) and (3) are no longer "pure" production functions as other explanatory variables, such as child ability, may also affect educational input purchases. But this will have little effect on estimated school fixed effects.

[17] The scores and standard deviations in table 2 are not exactly 500 and 100, respectively, as the standardization was done for the whole school survey sample, including school peers.

[18] In fact, differences in cognitive skills by gender at age 5 are small in Vietnam and insignificant in Peru; see table 1. For Peru, once we condition on income, nutritional status, parental education, and preschool attendance, we find no significant differences in test scores by ethnicity at age 5.

TABLE 2
SAMPLE DESCRIPTIVE STATISTICS

	Vietnam		Peru	
	Mean	SD	Mean	SD
From school survey:				
Mathematics test score (IRT transformed; whole sample mean = 500, SD = 100)	499.19	96.48	508.60	102.65
Language test score (IRT transformed; whole sample mean = 500, SD = 100)	497.44	96.69	505.77	99.69
Age in months at the time of the test	123.24	2.71	122.16	3.87
From household survey (time-invariant variables):				
Male (dummy)	.51	.50	.48	.50
Ethnic minority (dummy)	.08	.27	.36	.48
Father: years of schooling	7.69	3.53	9.23	3.77
Mother: years of schooling	7.17	3.30	7.84	3.11
Measured during infancy (round 1):				
Birth size (maternal assessment: 1 = very large to 5 = very small)	3.08	.66	3.11	.96
Health better than other children (dummy, maternal assessment)	.26	.44	.41	.49
Childcare: cared for by people outside the household /in crèche (dummy)	.42	.49	.21	.41
Maternal mental health	4.4	4.02	5.78	4.26
Wealth index	.47	.20	.43	.19
Measured at age 5 (round 2):				
Height-for-age *z*-score	−1.30	.97	−1.53	1.02
PPVT Score (IRT transformed; whole sample mean = 300, SD = 50)	305.33	44.33	301.34	44.84
CDA Score (IRT transformed; whole sample mean = 300, SD = 50)	306.80	46.56	299.35	46.39
Time spent in preschool (hours per day in a "normal" week)	5.79	2.45	3.44	1.83
Only child (dummy)	.22	.42	.19	.40
Household size	4.57	1.39	5.52	2.23
Log per capita real consumption (in local currency)	5.77	.54	5.02	.67
Area of land owned (in hectares)	.42	1.07	1.09	3.34
Number of observations	1,129		547	

Note. Ethnicity minority status: Peru—mother's first language is not Spanish (i.e., is Quechua, Aymara, or other); Vietnam—child reported by mother as not belonging to the Kinh group (i.e., is H'mong, Dao, Tay, other). Maternal mental health is measured using the Self-Reporting Questionnaire (SRQ20), a screening tool for risk of common mental disorders (rCMD) developed by the World Health Organization for developing countries (Beusenberg and Orley 1994). It is a scale consisting of 20 statements with yes/no answer options. The score reflects the number of yes responses out of possible 20; a higher score reflects higher rCMD. Wealth index is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). The Peabody Picture Vocabulary Test (PPVT) measures receptive vocabulary. The Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. IRT = Item Response Theory.

The situation is a bit more complicated when advantage is defined by household wealth; it should not directly affect learning after conditioning on purchased educational inputs (EI),[19] and so it should be excluded from **X** in equation (3), but a complication is that it is a continuous variable. Teachers or schools may be more effective for wealthier students in two distinct ways: increasing effectiveness that steadily rises with wealth, or a "sharp" increase in effectiveness (which could reflect intentional favoritism) that "turns on" when wealth crosses a specific threshold. In the former case, the school fixed effects will vary continuously with wealth, so estimates of their difference when the sample is divided roughly in half will estimate the differential effectiveness between an average advantaged student and an average disadvantaged student. In the latter case, school fixed effects take only two values in any given school, and if the "jump" in differential effectiveness occurs near the middle of that variable's distribution one can estimate that jump. However, if the jump is in one of the distribution's tails it is difficult to estimate because for most schools the data will not have both advantaged and disadvantaged students.

A final complication is when the advantage variable is both continuous and directly affects learning, which occurs for three of our variables: skills at age 5, parental education, and nutritional status.[20] Such variables must be included in **X**. If differential effectiveness that benefits advantaged students gradually increases as the advantage variable rises, it would be impossible to estimate such differential effectiveness because it cannot be disentangled from that variable's direct effect on learning; the estimated school fixed effects would not differ across the two types of students. Yet if differential effectiveness takes the form of a discrete jump when a threshold is crossed, it can be identified if that threshold is roughly in the middle (say, middle two quartiles) of that variable's distribution and that variable's direct impact on learning is smooth (does not jump). To define advantage in terms of nutritional status, we use the standard definition for not being stunted: a height for age z-score above -2; for the other two variables, we define disadvantage so that about 60% of the sample is advantaged.[21]

A second estimation issue is omitted-variable bias; any child and household variables in equations (2) and (3) not in the data become part of the error term,

[19] Indeed we find that controlling for purchased educational inputs, and for individual, parental, and household factors in the production function, there is no statistically significant correlation between the wealth index and test scores at age 5 in either Peru or Vietnam.

[20] If advantage were a dummy variable that directly affected learning, it would be impossible to estimate differential effectiveness because that **X** variable would be a linear combination of the advantaged students' school fixed effects.

[21] For Peru, 60.8% are "more able," 61.8% have better educated mothers, and 68.7% are not stunted. The corresponding figures for Vietnam are 65.6%, 81.5%, and 78.1%.

likely making it correlated with the observed variables, which can bias estimates of all the coefficients, including θ_s. More specifically, estimates of θ_s could include not only school fixed effects that affect advantaged children only but also general nonschool advantages those children have after controlling for other **X** variables. Fortunately, the Young Lives data have detailed information dating back to the students' first year of life; those data capture many, if not most, of the nonschool advantages of being an advantaged child. In particular, **X** has a comprehensive array of child and household variables, including conditions in the first year of life and the following 4 years before starting school, and—most importantly—children's cognitive skills before starting primary school, which should account for the effect of omitted education inputs (EI) and innate ability (IA) on *C*.[22]

If students sort into schools on the basis of unobserved traits—such as studiousness or parental tastes for education—that directly affect learning and are correlated with advantage indicators, such unobserved differences can be mistakenly interpreted as differential effectiveness that benefits advantaged students even if no such differential effectiveness exists. There is little school choice in rural Vietnam, where most of the Vietnam sample children live; children are assigned to schools on the basis of their residence, and rural areas typically have only one primary school. Yet school choice is an option in Peru, where most of the sample children live in urban areas and can enroll in other public schools; also, and unlike Vietnam, many private schools are available. We cannot rule out selection bias completely, but we expect that our extensive set of controls, including cognitive skills at age 5, will capture most unobserved factors correlated with school choice and test scores.

Omitted-variable bias concerns are also reduced because this article focuses on differences between groups rather than level effects. Omitted variables will bias estimates of θ_s only if they vary systematically across advantaged and disadvantaged children. Thus, our analysis is more robust than it would if we estimated, say, the level effects of school inputs. We explore in Section VIII the sensitivity of our estimates to omitted variables by comparing them to estimates based on stronger assumptions (cf. n. 22).

A third issue is that θ_s can be identified only for advantaged children who attend schools with at least one disadvantaged child from the school survey

[22] The estimated model is thus similar to the combined cumulative inputs value-added model favored by Todd and Wolpin (2007). A pure cumulative inputs model includes only the history of inputs and so assumes that observed inputs fully account for all inputs. In contrast, a value-added model includes only current inputs and last period's test score; it assumes that the lagged score fully captures endowments and the effects of the complete history of inputs. In the mixed cumulative value-added model that we estimate, identification relies on a weaker assumption than either of these, that combined lagged test scores and observed current and lagged inputs fully control for all past inputs.

sample. This may lead to biased estimates of the school fixed effects due to unobservable selection into schools with both types of children. This possibility is also examined in Section VIII.

A fourth estimation concern is that while school fixed effects are allowed to vary across advantaged and disadvantaged children, coefficients on child and household variables are not. In fact, this appears to be plausible; tests of whether those variables' impacts varied by advantaged status rejected the null of no difference in only 4 of 22 cases (see table A4).

Finally, some of the explanatory variables could be measured with error, such as the C_p, N, and PT variables. In principle, we need instrumental variables (IVs) for them, but finding valid IVs is a challenge. For this reason and because we focus on the impact of school, not child or household, characteristics on student learning, IV methods are not used to estimate equations (2) and (3).

VII. Results

This section examines evidence on whether within-school differences in learning exist between advantaged and disadvantaged children. It begins by briefly discussing estimates of the effect of the nonschool determinants of math and reading skills and then turns to estimates of the average value of θ_s, first for Vietnam and then for Peru, for each definition of advantage.

A. Nonschool Determinants of Learning

Table 3 presents estimates of equation (2). It focuses on the estimates of nonschool determinants of school attainment (**X** variables), separately for math and reading. Each school has only one school fixed effect, which applies to all pupils.[23]

Columns 1 and 2 of table 3 show the results for Vietnam. The 1,129 students in 90 schools include all younger cohort children in grade 5 in the 2011–12 school year. The coefficients on both cognitive skill (at age 5) variables are positive and highly significant. Several other variables are significant. First, other things equal, boys scored lower on both tests; the negative impact of being male on math skills is modest, −0.14 standard deviations, but the negative effect on reading, −0.32 standard deviations, is fairly large. Larger negative effects are seen for ethnic minority children; their math and reading scores are 0.38 and 0.47 standard deviations lower, respectively. Last, children of more educated parents had significantly higher scores on both tests.

[23] Estimates for eq. (3) are not shown as that requires six tables, one for each type of advantage. Estimates for the six versions of eq. (3) yield very similar results for the nonschool (child and household) variables.

TABLE 3

NONSCHOOL DETERMINANTS OF COGNITIVE SKILLS MEASURED BY 2011 SCHOOL SURVEY IN VIETNAM AND PERU

	Vietnam		Peru	
	Math Score (1)	Vietnamese Score (2)	Math Score (3)	Spanish Score (4)
PPVT score (IRT transformed; whole sample mean = 300, SD = 50)	.292***	.401***	.185	.220***
	(.077)	(.089)	(.116)	(.083)
CDA score (IRT transformed; whole sample mean = 300, SD = 50)	.196***	.141**	.208*	.350***
	(.070)	(.063)	(.108)	(.107)
Age at the time of the test (months after 9 years old)	15.386	7.453	5.347	2.380
	(9.291)	(10.100)	(6.709)	(7.466)
Age^2 at time of the test (months after 9 years old)	−.491	−.299	−.117	−.091
	(.298)	(.315)	(.231)	(.266)
In grade 3 at the time of the test	. . .	. . .	−67.145***	−49.427**
			(20.442)	(21.127)
In grade 4 at the time of the test	. . .	. . .	−17.014	−12.007
			(12.716)	(12.716)
Ethnic minority (dummy)	−38.259***	−47.172***	−7.563	−12.705
	(12.003)	(14.335)	(13.648)	(12.239)
Male (dummy)	−13.620**	−31.796***	5.290	−8.140
	(5.994)	(4.434)	(8.384)	(8.353)
Father's schooling: 4–6 years	1.670	11.129	−3.214	2.415
	(10.350)	(9.047)	(18.839)	(17.371)
Father's schooling: 7–9 years	6.978	17.339	14.176	−15.180
	(8.210)	(10.915)	(20.816)	(19.545)
Father's schooling: 10+ years	22.141**	28.035**	19.520	13.971
	(10.322)	(12.147)	(21.274)	(17.805)
Mother's schooling: 4–6 years	12.039	16.818*	7.236	8.746
	(7.869)	(9.376)	(12.976)	(11.407)
Mother's schooling: 7–9 years	13.426*	20.135*	14.154	25.313**
	(7.695)	(10.641)	(13.110)	(10.217)
Mother's schooling: 10+ years	28.050**	35.351**	35.199***	40.638***
	(12.466)	(15.004)	(11.519)	(10.489)
Birth size (maternal assessment: 1 = very large to 5 = very small)	−1.608	1.639	2.513	−2.573
	(3.210)	(3.855)	(4.344)	(5.111)
Health better than other children's health (dummy, maternal assessment)	−1.699	−4.625	4.676	6.382
	(8.420)	(6.687)	(6.804)	(8.766)
Childcare: looked after by others/in crèche (dummy)	−9.501*	−10.996*	−11.404	−24.625***
	(5.681)	(5.576)	(7.829)	(8.718)
Maternal mental health	−1.276*	−1.065	−.057	−.018**
	(.666)	(.693)	(.035)	(.009)
Wealth index (at age 1)	−74.921	−12.927	59.262	334.004***
	(46.253)	(72.223)	(136.076)	(108.217)
Wealth index^2 (at age 1)	78.239	2.534	−39.833	−326.216***
	(51.615)	(88.090)	(135.841)	(117.557)
Height-for-age *z*-score	1.725	7.007**	7.067**	8.483**
	(3.480)	(3.117)	(3.443)	(3.556)

TABLE 3 (*Continued*)

	Vietnam		Peru	
	Math Score (1)	Vietnamese Score (2)	Math Score (3)	Spanish Score (4)
Time spent in preschool (hours per day in a "normal" week)	−.937	−.807	.575	−3.282
	(2.124)	(2.019)	(2.468)	(1.997)
Only child (dummy)	−7.804	2.601	2.364	2.783
	(6.765)	(4.683)	(8.130)	(8.863)
Household size	−1.119	−4.472***	−.439	1.489
	(1.869)	(1.652)	(1.949)	(1.990)
Log per capital real consumption (in local currency)	2.237	3.103	4.643	6.650
	(6.725)	(6.754)	(8.267)	(6.866)
Area of land owned (hectares)	−1.729	.126	1.960**	−.702
	(2.500)	(2.360)	(.900)	(1.039)
Constant	253.384***	301.134***	290.666***	235.264***
	(76.393)	(87.763)	(103.405)	(84.032)
School fixed effects	Yes	Yes	Yes	Yes
Number of observations	1,129	1,129	547	547

Note. Robust standard errors shown in parentheses; they are not clustered at the site level because there are only 20 sentinel sites and so not enough degrees of freedom and because school fixed effects will capture most of the correlation of the error terms within the sentinel sites. Ethnicity minority status: Peru—mother's first language is not Spanish (i.e., is Quechua, Aymara, other); Vietnam—child reported by mother as not belonging to the Kinh group (i.e., is H'mong, Dao, Tay, other). Grade at the time of the test: Vietnam—all children in grade 5, so no control; Peru—grade 5 is the omitted category. Father and mother's schooling: no schooling/incomplete primary is the omitted category. Maternal mental health is measured using the Self-Reporting Questionnaire (SRQ20), a screening tool for risk of common mental disorders (rCMD) developed by the World Health Organization for developing countries (Beusenberg and Orley 1994). It is a scale consisting of 20 statements with yes/no answer options. Score reflects the number of yes responses out of possible 20; a higher score reflects higher risk of rCMD. Wealth index is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). The Peabody Picture Vocabulary Test (PPVT) measures receptive vocabulary. The Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. IRT = Item Response Theory.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

Two of the five variables measured around age 1 have significant impacts.[24] First, enrollment in childcare at age 1 significantly lowers math and Vietnamese scores (10% significance level). This may reflect reduced parental contact in early childhood. Second, there is a marginally significant negative impact of maternal stress or depression on math scores.

[24] Some of these variables may be endogenous, which could bias the estimates of all variables, including school fixed effects. To check this we reestimated the main tables after dropping the two most likely to be endogenous: someone outside the household cared for the child and time in preschool. The results, available on request, are very similar.

As expected (see Glewwe and Miguel 2008; Alderman and Bleakley 2013), height for age at age 5 has positive impacts on both math and reading, although significantly so for the latter only. Household size at age 5 has a negative impact; this suggests that additional siblings reduce resources and parental attention for the child, but it is significant only for reading skills.

In Peru (table 3 cols. 3 and 4), the PPVT and CDA scores again have generally significantly positive effects on math and reading scores. Boys have higher math skills and lower reading skills, yet unlike for Vietnam neither effect is significant. Ethnic minority children have slightly lower math and reading scores but not significantly so. As in Vietnam, parental education matters even after controlling for preschool skills, but only mother's education is significant.

Peruvian children enrolled in childcare at age 1 scored lower on both tests, and the large impact on Spanish skills (-0.25 standard deviations) is significant. Unlike in Vietnam, the wealth index has a significantly positive but diminishing impact on Spanish scores but not math scores. As in Vietnam, nutritional status at age 5 has a significantly positive impact. Finally, land ownership has a significantly positive effect but only on math skills.

B. *Differences in School Impacts by Advantaged Status*

The θ_s parameters in equation (3) indicate whether the impacts of school and teacher characteristics on learning are, on average, larger for advantaged children. Columns 1 and 2 of table 4 present estimates of the average θ_s, the difference between the average school fixed effects of advantaged and disadvantaged children, for Vietnam for five definitions of advantage. The average θ_s is weighted by the number of younger cohort children in each school. The standard errors for these averages are calculated using the standard formula for the variance of a linear combination of estimated parameters from a regression (see, e.g., the discussion of eq. [5-17] in Greene 2012, 119).[25]

The estimated means of θ_s reveal little evidence that the average school impact in Vietnam varies between advantaged and disadvantaged children. For all but one definition, the estimated averages of θ_s are statistically insignificant. The exception is when advantage is defined as being male: Vietnamese schools convey a sizable *disadvantage* to being male. That is, the average impact of schooling on boys' math scores is 12.7 points (0.13 standard deviations) lower than for girls, a difference significant at the 5% level, and 29.9 points (0.30 stan-

[25] Note that the joint significance of the school fixed effects, separately for advantaged and disadvantaged children, is very high for all definitions of advantage, for both Vietnam and Peru (not shown in table 4); this indicates that schools vary substantially in their contribution to children's acquisition of cognitive skills.

TABLE 4

DIFFERENCES IN SCHOOL EFFECTIVENESS BY ADVANTAGE GROUP: MEAN INCREMENTAL EFFECT, $\bar{\theta}_s$

	Vietnam		Peru	
	Math Score (1)	Vietnamese Score (2)	Math Score (3)	Spanish Score (4)
Being richer (top 2 wealth quintiles)	−8.11	−3.55	10.31	1.87
	(7.42)	(6.89)	(10.79)	(12.29)
Being more able (top 2 CDA quintiles)	9.12	11.10	19.98	38.12**
	(9.63)	(9.72)	(14.76)	(17.04)
Being male	−12.69**	−29.88***	7.52	−5.47
	(5.09)	(4.88)	(6.67)	(7.19)
Being an ethnic majority			24.56**	22.05*
			(9.68)	(11.79)
Having a mother with at least primary education	−8.98	−14.22	−14.31	2.01
	(11.45)	(10.96)	(17.85)	(19.56)
Having not been malnourished in infancy	4.81	−6.76	1.33	14.36
	(9.11)	(7.72)	(9.14)	(11.76)

Note. Robust standard errors in parentheses. The mean incremental effect of school on test scores of advantaged children compared to disadvantaged children is $\bar{\theta}_s$, as shown in eq. (3). It is the mean value of the coefficients for the interaction of school fixed effects with each of the advantage categories. All models control for the full set of nonschool child, parent, and household characteristics shown in table 3. Math and verbal scores were transformed using Item Response Theory; whole sample mean = 500, standard deviation = 100. Being richer: wealth quintiles are constructed using a wealth index that is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). Being more able: the Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. Being an ethnic majority: Peru—mother's first language is Spanish (i.e., not Quechua, Aymara, other). Not being malnourished in infancy: defined using anthropometric indicator of adequate long-term growth (height-for-age *z*-scores > −2).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

dard deviations) lower on the Vietnamese test, a difference significant at the 1% level, even after controlling for age 5 test scores. This is surprising, because at age 5 boys outperform girls on both cognitive tests (table 1).[26] Figure 1*A* shows the distribution (over schools) of θ_s (on the left) as well as the (separate) distributions of boys' and girls' school fixed effects. The difference in the values of the fixed effects θ for math scores occurs mostly in the middle and upper range of those distributions, while for Vietnamese it is at all points in the distributions.

Overall, we find no evidence that schools in Vietnam are more effective for advantaged children; the means of θ_s are close to zero. Indeed, the only signif-

[26] We cannot directly control for any nonschool advantages of boys, yet if such an advantage is not captured by age 5 test scores and other controls the estimated difference in boys' and girls' school effects would be biased toward zero.

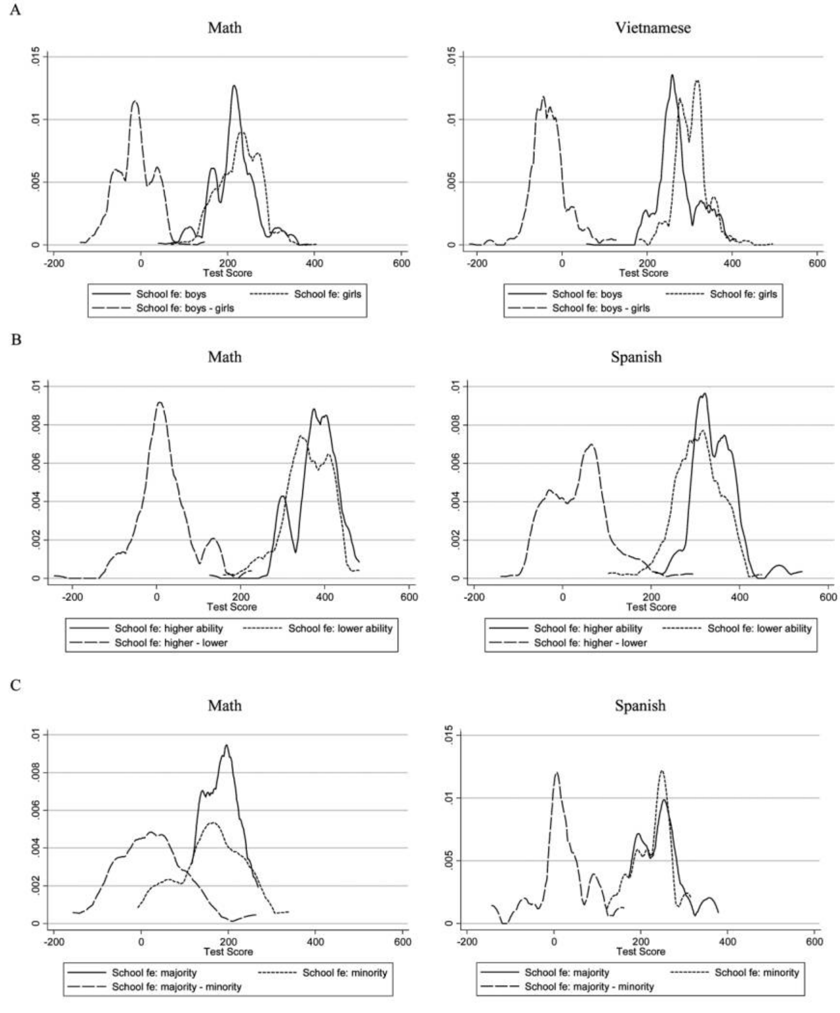

Figure 1. Distribution of estimated θ parameters for selected definitions of advantage. *A*, Vietnam by gender; *B*, Peru by being more able; *C*, Peru by ethnicity. Color version available as an online enhancement.

icant effect is that Vietnamese schools seem to be more effective at increasing the learning of girls, who are often considered disadvantaged, especially in math (e.g., Fryer and Levitt 2010).

In contrast, the results for Peru, in table 4 columns 3 and 4, show differences in school fixed effects between advantaged and disadvantaged children when advantaged is defined in terms of preschool skills or in terms of ethnicity. When advantage is defined in terms of children's preschool skills (CDA scores at age 5), schools in Peru seem to be more effective for better prepared children in their acquisition of Spanish skills, an impact of 38.1 (0.38 standard deviations).[27] The estimated average θ_s is also large (20.0), but not quite significant, for math. Figure 1*B* shows that the difference in estimated school fixed effects for math scores occurs mostly in the lower and middle of the distributions of those fixed effects, while for Spanish the distributions differ at almost all points.

Unlike in Vietnam, schools in Peru are not more effective for girls than for boys, or vice versa, in either math or reading. There is also no evidence of nonschool differences between boys and girls at age 5 (see table 1). Yet schools in Peru appear to contribute more to learning for ethnic majority children. Even after controlling for age 5 skills, nutritional status, and parental education (and all other variables in table 3), the impact of school and teacher characteristics on ethnic majority children's math skills is 24.6 points (0.25 standard deviations) higher than their impact on ethnic minority children; the analogous differential impact for Spanish skills is 22.1 (0.22 standard deviations). Both differences are statistically significant.[28] Note that this does not reflect some type of higher innate ability for ethnic majority children; as explained in footnote 18, after conditioning on income, nutritional status, parental education, and preschool attendance, there are no significant differences in test scores by ethnicity at age 5.

Figure 1*C* shows that the difference in the values of θ for math is manifested in a long lower tail for ethnic minority children but not for ethnic majority children; for Spanish there is a similar, although less clear, pattern. One could argue that unobserved child or household factors explain this difference (i.e., that these estimated differences in school effects are biased upward), yet it is difficult to imagine mechanisms for such bias.[29]

Finally, table 4 also shows, as in Vietnam, no significant differences in school impacts between children in Peru with less educated mothers and those with more educated mothers, as well as no significant difference by children's nutritional status in early childhood. Overall, even after controlling for a large

[27] This does not reflect that such children had better Spanish skills at age 5, as those skills are in the control variables (the PPVT score).

[28] A plausible alternative explanation for this difference could be that parents of children from ethnic majority backgrounds invest more into supporting their children through school, through both time and material investments, than parents of ethnic minority children. While we cannot rule out this mechanism, we do not find evidence to support it; specifically, when we add a control for household educational expenditure at age 10, the difference in school effectiveness between ethnic minority and majority children remains significant and, if anything, increases.

[29] Consistent with this, controlling for household, parent, and child characteristics, including preschool test scores, we find that there is no significant association between ethnicity and household educational expenditure at age 5.

number of child and household variables, when advantage is defined as having higher skills at age 5 and not being an ethnic minority, advantaged children in Peru learn more than disadvantaged children who attend the same schools. This is in stark contrast with Vietnam, where advantaged children never appear to learn more than disadvantaged children in the same school and for one advantage type (male) appear to learn less.

VIII. Robustness Checks and an Extension

This section checks the robustness of our results and presents some evidence on the role of parental investments in explaining the gap between advantaged and disadvantaged children.

A. Omitted-Variable Bias

As discussed in Section VI, despite using an extensive set of controls for nonschool inputs, including test scores at age 5, omitted-variable bias remains a concern. To explore this, we assess the sensitivity of the main estimates to different assumptions about omitted variables. We compare three sets of estimates: (1) main model (i.e., combining current and lagged inputs with lagged skill scores), (2) using only lagged skill scores to capture all lagged inputs and innate ability, and (3) controlling only for current and lagged observed inputs. If the main results are driven by omitted-variable bias, they are likely to be sensitive to different assumptions about omitted variables and so should vary noticeably across these three models.

Table 5 shows the three sets of estimates for Vietnam and Peru for the types of disadvantage with significant differences: gender for Vietnam and preschool cognitive skills and ethnicity in Peru. Estimates of $\bar{\theta}_s$ for Vietnam when advantage is defined as being male are very stable across the three specifications for both math and Vietnamese. Turning to Peru, when advantage is defined as skills at age 5, estimates of $\bar{\theta}_s$ across the three models are very stable for Spanish and reasonably stable for math. When advantage is defined as being from the ethnic majority group, estimates of $\bar{\theta}_s$ are very stable for math skills, although there is more fluctuation for Spanish scores (estimates of $\bar{\theta}_s$ range between 16.9 and 31.1). Overall, for all advantage categories in which we find significant differences in school effects, the consistency of the direction and in most cases size and statistical significance of estimates of $\bar{\theta}_s$ yield little evidence of omitted-variable bias.

B. Sample Restrictions

Another plausible source of bias is selection of the analysis subsamples. First, as explained in Section V, the school survey subsample for Vietnam includes

TABLE 5

SENSITIVITY OF THE ESTIMATED SIGNIFICANT DIFFERENCES IN SCHOOL EFFECTIVENESS BY ADVANTAGE GROUP: MEAN INCREMENTAL EFFECT, $\bar{\theta}_s$

	Vietnam		Peru	
	Math Score (1)	Vietnamese Score (2)	Math Score (3)	Spanish Score (4)
Model 1:				
Being male	−12.69**	−29.88***		
	(5.09)	(4.88)		
Being more able (top 2 CDA quintiles)			19.98	38.12**
			(14.76)	(17.04)
Being an ethnic majority			24.56**	22.05*
			(9.68)	(11.79)
Model 2:				
Being male	−13.63***	−31.39***		
	(5.05)	(4.95)		
Being more able (top 2 CDA quintiles)			29.10*	41.37***
			(15.48)	(15.90)
Being an ethnic majority			26.61***	31.07***
			(9.15)	(8.74)
Model 3:				
Being male	−9.98**	−26.61***		
	(5.09)	(4.95)		
Being more able (top 2 CDA quintiles)			19.95**	33.14***
			(8.40)	(9.75)
Being an ethnic majority			22.21**	16.88
			(9.73)	(11.36)

Note. Robust standard errors in parentheses. The mean incremental effect of school on test scores of advantaged children compared to disadvantaged is $\bar{\theta}_s$, as shown in eq. (3). Model 1: main model (estimates reproduced for convenience)—controlling for all nonschool child, parent, and household characteristics shown in table 3. Model 2: controls for age 5 cognitive test scores only. Model 3: controls for age 1 and 5 child, parent, and household characteristics, excluding age 5 cognitive test scores. Math and verbal scores were transformed using Item Response Theory; whole sample mean = 500, standard deviation = 100. Being more able: the Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. Being an ethnic majority: Peru—mother's first language is Spanish (i.e., not Quechua, Aymara, other).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

only those in grade 5 (the appropriate grade for their age) in the 2011–12 school year. In contrast, the Peru school survey sample is a subsample of the Young Lives children in grades 3, 4, and 5. Estimates using the Vietnam school survey subsample may be biased if the children excluded from that subsample are more likely to be disadvantaged (e.g., grade repeaters). Indeed, the data support this; unlike, Peru, the children included in the Vietnam school survey subsample are on average better off than those in the full Vietnam Young Lives sample (see table A5).

To check whether our main findings are driven by this selection, we reestimated them using the full sample of children and test scores from the last

TABLE 6
COMPARING MAIN RESULTS TO ESTIMATES USING HOUSEHOLD SURVEY TESTS AND SAMPLE IN VIETNAM: MEAN INCREMENTAL EFFECT, $\bar{\theta}_s$

	Math Score (School Survey) (1)	Math Score (Household Survey) (2)	Vietnamese Score (School Survey) (3)	Early Grade Reading Assessment Score (Household Survey) (4)
Being richer (top 2 wealth quintiles)	−8.11	−.38	−3.55	−.07
	(7.42)	(.77)	(6.89)	(.84)
Being more able (top 2 CDA quintiles)	9.12	−.80	11.10	−.31
	(9.63)	(1.03)	(9.72)	(1.11)
Being male	−12.69**	−.50	−29.88***	−2.30***
	(5.09)	(.58)	(4.88)	(.61)
Having a mother with at least primary education	−8.98	.86	−14.22	−2.56*
	(11.45)	(1.21)	(10.96)	(1.36)
Having not been malnourished in infancy	4.81	1.76**	−6.76	−.81
	(9.11)	(.82)	(7.72)	(.89)

Note. Robust standard errors in parentheses. The mean incremental effect of school on test scores of advantaged children compared to disadvantaged children is $\bar{\theta}_s$, as shown in eq. (3). All models control for full set of nonschool child, parent, and household characteristics shown in table 3. Columns 1 and 3: math and Vietnamese scores on tests administered to subset of children in grade 5 who were in the school survey ($N = 1{,}129$); test scores transformed using Item Response Theory; whole sample mean = 500, SD = 100. Columns 2 and 4: math and Early Grade Reading Assessment scores on tests administered to all children in round 3 of household survey ($N = 1{,}965$; not the same tests as the school survey tests); raw test score standardized to have a mean of 300 and a standard deviation of 15. Being richer: wealth quintiles are constructed using a wealth index that is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). Being more able: the Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. Not being malnourished in infancy: defined using anthropometric indicator of adequate long-term growth (height-for-age z-scores > -2).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

available round of the main Young Lives survey (round 3).[30] Table 6 reveals no systematic differences between the two sets of estimates.[31] For math, the signs are the same for three of the five types of advantage, and in the two cases with

[30] The reason why the main analysis uses the subsample of children who were in the school survey is that the school survey tests were carefully designed to reflect the school curriculum and thus are the most appropriate outcome measures for exploring the impact of schooling on student learning. In contrast, the round 3 tests were designed to reflect more general mathematics and verbal skills and so are not linked to specific school curricula.
[31] Note that the round 3 tests have a standard deviation of only 15, compared to a standard deviation of 100 for the school survey tests, so one would expect coefficients to be six to seven times smaller for the round 3 (household survey) tests.

different signs neither estimate is statistically significant. In the two cases in which one estimate is significant and the other is not, the signs are the same, and in one case the significant result is for the school survey subsample and in the other case it is for the household sample. For reading, the signs are again the same for three of the five definitions of advantage, and in the two cases with different signs both estimates are insignificant. For the sole case when both samples find a significant differential, they agree: boys appear to be at a disadvantage relative to girls. Thus, we conclude that excluding grade 3 and 4 students in the Vietnam school survey does not lead to systematic biases.

A second sample restriction is inclusion of only those schools with both advantaged and disadvantaged children in the school survey subsample. Selection into these schools may differ across these two types of children.[32] For example, among children from richer households, those in the same schools as children from poorer families, rather than elite schools, may be the less able or less motivated among the richer children, while the reverse may hold for students in those schools from poorer households. While possible, the insensitivity of our estimates to different assumptions about omitted-variable bias suggests that this is unlikely to drive the main results.

Third, the results may be sensitive to differences in the sample for different definitions of advantage, which occur because the estimates require at least one advantaged and one disadvantaged child in each school, so the analysis samples vary slightly for different types of advantage.[33] To address this for Vietnam, we constructed a "master sample" of 38 schools with 822 children that are in the sample for all definitions of advantage; estimates of differences in school effects between more and less advantaged children for this sample are very similar to the main estimates, although less precise. For Peru, the smaller school sample rules out a general "master sample" (it would have only 96 children in six schools), so we focused on a somewhat larger sample for which each school had (at least) one child with relatively low skills at age 5, one with relatively high age 5 skills, one with a relatively uneducated mother, and one with an educated mother. Again, we found similar, although very imprecise, results. Details for both countries are available on request.

Finally, sample restrictions may limit the external validity of our findings. If schools with both advantaged and disadvantaged children are atypical, our results may hold only for a small subset of schools in each country. To check this,

[32] Recall that all unobservables affecting both groups are differenced out, as θ_s is the difference in school fixed effects.

[33] For example, the subsample of children in schools that have both more and less wealthy children from the school survey sample is different from the subsample in schools that have both ethnic majority and minority children.

tables 7 and 8 show, by country, several school characteristics for the whole sample and the "overlap subsamples" (schools with both types of children), with asterisks indicating significant differences between whole sample and subsample means. For Vietnam, the results show almost no significant differences for all definitions of advantage. Yet some differences emerge for Peru; this is unsurprising since, on average, the Peru sample has fewer children per school than the Vietnam sample, so fewer Peru schools have both advantaged and dis-

TABLE 7
MAIN CHARACTERISTICS OF SCHOOLS IN VIETNAM IN THE WHOLE SAMPLE AND SCHOOLS IN THE ANALYSIS OVERLAP SUBSAMPLES

		Overlap Subsamples				
	All (1)	Wealth (2)	Ability (3)	Sex (4)	Maternal Education (5)	Malnutrition in Infancy (6)
School assets index	.00	.33	.25	.26	.29	.07
	(1.64)	(1.36)	(1.59)	(1.57)	(1.46)	(1.45)
School size (number of students)	481.73	535.81	533.77	502.63	501.75	445.71
	(347.04)	(387.28)	(387.38)	(367.53)	(375.76)	(196.72)
School offers free full-day schooling	.24	.17	.13	.25	.23	.26
	(.43)	(.38)	(.34)	(.44)	(.43)	(.44)
School offers free lunch	.88	.91	.83	.87	.88	.90
	(.33)	(.28)	(.38)	(.34)	(.04)	(.30)
Principal: years of experience	10.78	10.91	10.10	10.76	10.15	10.48
	(7.08)	(7.27)	(6.98)	(7.35)	(7.10)	(6.83)
Teacher: years of experience	17.58	17.10	16.68	17.81	17.88	17.49
	(6.37)	(6.26)	(5.39)	(6.33)	(6.45)	(6.72)
Teacher: score on Vietnamese pedagogy test	70.27	69.97	71.17	70.68	70.13	71.57
	(12.23)	(13.10)	(13.57)	(12.96)	(12.79)	(11.18)
Teacher: score on math pedagogy test	68.84	68.30	70.61	68.93	67.90	68.38
	(8.44)	(6.74)	(8.32)	(8.34)	(7.89)	(7.38)
Pupils: mean number of days absent (per month)	1.94	1.57	2.52	2.18	2.46	1.94
	(6.82)	(6.51)	(8.16)	(.84)	(8.18)	(7.53)
Pupils: proportion ethnic minority	.26	.09***	.12**	.19	.16*	.20
	(.37)	(.19)	(.24)	(.31)	(.28)	(.32)
Pupils: proportion of grade repeaters	.05	.04	.04	.05	.05	.05
	(.05)	(.04)	(.05)	(.06)	(.05)	(.05)
Number of schools	90	48	61	77	61	63

Note. Means, with standard deviations in parentheses and asterisks indicating significance level of *t*-test of difference in means for all schools. Column 1 shows mean for whole school survey sample. Overlap subsamples include schools that have both advantaged and disadvantaged children when advantage is defined as follows: col. 2, top 2 wealth quintiles; col. 3, top 2 ability quintiles; col. 4, boys; col. 5, children of more educated mothers; col. 6, children who were not malnourished in infancy. See table 4 note for details on how each category is measured. School asset index is the first factor from principal components analysis of school-level indicators for need for repairs, separate room for each grade 5 class, and availability of library, computers, electricity, and Internet in school.
* $p < .1$.
** $p < .05$.
*** $p < .01$.

TABLE 8

MAIN CHARACTERISTICS OF SCHOOLS IN PERU IN THE WHOLE SAMPLE AND SCHOOLS IN THE ANALYSIS OVERLAP SUBSAMPLES

		Overlap Subsamples					
	All (1)	Wealth (2)	Ability (3)	Sex (4)	Ethnicity (5)	Maternal Education (6)	Malnutrition in Infancy (7)
School assets index	.00	1.33***	.52	.28	.27	−.27	−.09
	(1.74)	(1.47)	(1.83)	(1.88)	(1.76)	(1.70)	(1.89)
School size (number of students)	378.0	785.9***	553.7***	491.1	557.7**	417.3	410.3
	(398.6)	(444.4)	(449.2)	(456.9)	(504.4)	(413.3)	(415.1)
School private	.14	.08	.11	.10	.04	.00***	.06*
	(.35)	(.28)	(.31)	(.30)	(.20)	(.00)	(.23)
School offers free lunch	.66	.58	.62	.67	.80	.71	.72
	(.48)	(.50)	(.49)	(.48)	(.41)	(.46)	(.45)
Principal: years of experience	12.14	13.6	11.66	11.76	9.92	11.57	11.81
	(8.67)	(9.37)	(8.06)	(8.34)	(7.40)	(8.40)	(8.12)
Teacher: years of experience	17.23	19.52	18.05	18.92	18.57	19.59*	17.81
	(7.69)	(6.40)	(7.22)	(7.05)	(6.32)	(6.34)	(7.38)
Teacher: score on math pedagogy test	7.74	8.23	7.62	7.57	7.99	7.25	7.67
	(2.02)	(1.44)	(2.15)	(2.08)	(1.42)	(2.35)	(2.38)
Pupils: mean number of days absent (per week)	.47	.30	.31	.43	.36	.50	.45
	(1.06)	(.47)	(.47)	(.76)	(.50)	(.86)	(.79)
Pupils: proportion ethnic minority	.19	.05**	.15	.17	.08	.21	.22
	(.32)	(.11)	(.29)	(.30)	(.17)	(.32)	(.34)
Pupils: proportion of grade repeaters	.26	.14***	.19	.22	.23	.24	.27
	(.29)	(.13)	(.18)	(.19)	(.19)	(.16)	(.21)
Number of schools	132	36	47	63	25	48	54

Note. Means, with standard deviations in parentheses and asterisks indicating significance level of *t*-test of difference in means for all schools. Column 1 shows mean for whole school survey sample. Overlap subsamples include schools that have both advantaged and disadvantaged children when advantage is defined as follows: col. 2, top 2 wealth quintiles; col. 3, top 2 ability quintiles; col. 4, boys; col. 5, ethnic majority children; col. 6, children of more educated mothers; col. 7, children who were not malnourished in infancy. See table 4 note for details on how each category is measured. School asset index is the first factor from principal components analysis of school-level indicators for need for repairs, separate room for each class, and availability of library, computers, electricity, and Internet in school.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

advantaged children. Indeed, schools in the overlap samples are larger than in the overall sample. The subsample most different from the main sample is the schools with both wealthier and poorer children. They are larger, with more assets and fewer ethnic minority pupils and grade repeaters. To the extent that these are "above average" schools, the external validity of the results for this type of advantage may be lower. Overall, however, other than school size there is little systematic selection of schools into the overlap samples, especially for the types of advantage for which we find significant within-school differences in school effectiveness.

C. Advantage Category Cutoffs

We also checked whether our results hold when different cutoffs (50th and 70th percentiles) are used to define advantage by wealth and preschool skills. The main results are unchanged (results available on request). In Vietnam, irrespective of the cutoff the results suggest that schools are equally effective at teaching children from more and less advantaged backgrounds. In Peru, the three different cutoffs (lowest 50%, 60%, and 70%) yield the persistent finding that schools are more effective for children with higher preschool skills.

D. Role of Parental Investment

This article focuses on the impacts of teachers and schools on learning and whether they differ for advantaged and disadvantaged children in the same school. Yet the data can also shed light on parental behavior. One approach is to remove from **X** all parental choice variables—child size at birth, child health (mother's assessment), use of childcare, maternal stress/depression, household wealth, height for age, skills at age 5, time in preschool, only-child status, household size, per capita consumption, and land owned—keeping only child age, sex, and ethnicity and parental education. Estimates of θ_s from this specification will be biased if parental behavior is correlated with a particular type of advantage. For example, parents of disadvantaged children may be discouraged and so not enroll their children in preschool, which would lead to overestimation of θ_s; alternatively, such parents may do more to help their children, leading to underestimates of θ_s. Of course, such bias could vary by the type of advantage.

Table 9 presents estimates of table 4 after removing parental choices from **X**. For most types of advantage, the average estimated within-school differential in learning between advantaged and disadvantaged students barely changes. Yet there are two exceptions. First, for both countries the average difference in school fixed effects (θ_s) increases sharply for advantage defined by nutrition in infancy. This implies that parents of well-nourished children take other actions to raise their children's learning, so not controlling for parental behavior will overestimate the difference in school fixed effects between advantaged and disadvantaged students.

Second, defining advantage in terms of skills at age 5, the insignificant difference in fixed effects between advantaged and disadvantaged children in Vietnam becomes large and significant. This suggests that Vietnamese parents make additional educational investments in their "high-ability" children, and

TABLE 9

MEAN INCREMENTAL EFFECTS WITH AND WITHOUT CONTROLS FOR PARENTAL INVESTMENTS: MEAN INCREMENTAL EFFECT, $\bar{\theta}_s$

	Vietnam		Peru	
	Math Score (1)	Vietnamese Score (2)	Math Score (3)	Spanish Score (4)
Model 1:				
Being richer (top 2 wealth quintiles)	−8.11	−3.55	10.60	2.47
	(7.42)	(6.89)	(10.63)	(12.44)
Being more able (top 2 CDA quintiles)	9.12	11.10	19.83	38.32**
	(9.63)	(9.72)	(14.80)	(17.12)
Being male	−12.69**	−29.88***	7.58	−4.39
	(5.09)	(4.88)	(6.75)	(7.07)
Being an ethnic majority			24.30**	20.39*
			(9.67)	(11.76)
Having a mother with at least completed primary schooling	−8.98	−14.22	−14.08	1.30
	(11.45)	(10.96)	(17.79)	(19.24)
Having not been malnourished in infancy	4.81	−6.76	1.40	13.55
	(9.11)	(7.72)	(9.24)	(11.79)
Model 2:				
Being richer (top 2 wealth quintiles)	−1.81	1.96	8.99	6.05
	(6.62)	(6.23)	(8.05)	(9.91)
Being more able (top 2 CDA quintiles)	21.18***	21.79***	19.64**	35.10***
	(6.22)	(5.92)	(8.09)	(9.95)
Being male	−9.94**	−26.32***	7.06	−6.50
	(5.01)	(4.97)	(6.77)	(7.14)
Being an ethnic majority			19.26**	14.88
			(9.36)	(15.63)
Having a mother with at least completed primary schooling	−18.01	−23.70**	−10.10	−7.33
	(11.18)	(10.97)	(17.58)	(20.62)
Having not been malnourished in infancy	7.84	8.52	9.01	23.38**
	(7.79)	(6.32)	(11.13)	(9.34)

Note. Robust standard errors in parentheses. The mean incremental effect of school on test scores of advantaged children compared to disadvantaged children is $\bar{\theta}_s$, as shown in eq. (3). Model 1: main model (reproduced for convenience). Model 2: excludes controls for parental inputs, includes controls for sex, age, ethnicity, parental education only (Peru only: grade at the time of the test). Math and verbal scores were transformed using Item Response Theory; whole sample mean = 500, standard deviation = 100. Being richer: wealth quintiles are constructed using a wealth index that is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). Being more able: the Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. Being an ethnic majority: Peru—mother's first language is Spanish (i.e., not Quechua, Aymara, other). Not being malnourished in infancy: defined using anthropometric indicator of adequate long-term growth (height-for-age z-scores > -2).

* $p < .1$.

** $p < .05$.

*** $p < .01$.

not controlling for parental behavior may lead one to find that Vietnamese teachers or schools are more effective for advantaged students.

IX. Conclusion

Students in many developing countries learn far less than their counterparts in developed countries, and within developing countries some children learn much more than others. This article considers whether schools in Vietnam and Peru reinforce or reduce gaps in learning between advantaged and disadvantaged students for six different types of advantage by estimating separate school fixed effects for advantaged and disadvantaged students. This focuses on intra- rather than interschool differences in schools' contribution to learning among different types of students.

We estimate school fixed effects for advantaged and disadvantaged children, controlling for a variety of child and household characteristics, including cognitive skill tests at age 5 (before starting primary school). We find no evidence that schools in Vietnam are more effective for advantaged students. Indeed, the one significant effect is that girls, who some consider disadvantaged, pull ahead of boys between age 5 and 10. In contrast, for two definitions of disadvantage schools in Peru appear to be more effective for advantaged students: those with higher skills at age 5 acquire more Spanish skills than do less prepared students, and ethnic majority students learn more math and Spanish than minority students, even after controlling for skills at age 5.

These findings conform with existing evidence on differences between the school systems of Vietnam and Peru. Vietnam's expansion in primary education in the last 2 decades included effective investment in education quality and a focus on equity, emphasizing that all pupils attain "minimum standards." In contrast, Peru's schools suffer from low average quality and high inequality in student learning, with evidence of gaps in access and learning outcomes by, for example, income and ethnicity. Our results show that such gaps exist even among students attending the same school; more advantaged pupils learn more than less advantaged pupils, even after controlling for many pupil characteristics, including skills at age 5.

This article does not explore the mechanisms behind these effects, yet the recent literature on within-school discrimination against disadvantaged groups and biases in school curriculum in favor of advantaged children in several contexts offers potential explanations. Classroom observations to see whether teachers discriminate against disadvantaged children should be a high priority. Yet one conclusion is clear: estimates that assume that schools have the same impacts on different types of students may overlook a major source of inequality in student learning in developing countries.

Appendix

TABLE A1

OVERLAP IN ADVANTAGE DEFINITIONS: VIETNAM

	Wealth		Ability		Gender		Maternal Education		Nutrition	
	Rich	Poor	More Able	Less Able	Male	Female	Mother More Educated	Mother Less Educated	Not Stunted in Infancy	Stunted in Infancy
Richer	100	0	38.0	62.0	51.1	48.9	94.5	5.5	86.8	13.2
Poorer	0	100	26.9	73.1	51.3	48.7	72.2	27.8	74.7	25.3
More able	50.7	49.3	100	0	53.2	46.8	83.5	16.5	81.0	19.0
Less able	38.0	62.0	0	100	50.3	49.7	80.7	19.3	79.2	20.8
Male	41.8	58.2	32.8	67.2	100	0	83.2	16.8	75.0	25.0
Female	42.2	57.8	30.3	69.7	0	100	79.9	20.1	84.8	15.2
Mother more educated	48.7	51.3	32.3	67.7	52.2	47.8	100	0	81.6	18.4
Mother less educated	12.5	87.5	28.4	71.6	46.6	53.4	0	100	71.6	28.4
Not malnourished in infancy	45.7	54.3	32.0	68.0	48.1	51.9	83.5	16.5	100	0
Malnourished in infancy	27.5	72.5	29.7	70.3	63.3	36.7	74.2	25.8	0	100

Note. Proportion of children in each of the row categories who are in each of the column categories. Proportions add up to 100 for each of the advantage/disadvantage column pairs. For example, row 1 shows that among the children in the "rich" category, 38% are in the "more able" category and 62% are in the "less able" category. See table 4 note for details on how each category is measured.

TABLE A2

OVERLAP IN ADVANTAGE DEFINITIONS: PERU

	Wealth		Ability		Gender		Ethnicity		Maternal Education		Nutrition	
	Rich	Poor	More Able	Less Able	Male	Female	Ethnic Majority	Ethnic Minority	Mother more Educated	Mother less Educated	Not stunted in Infancy	Stunted in Infancy
Richer	100	0	49.8	50.2	50.2	49.8	88.6	11.4	94.1	5.9	83.6	16.4
Poorer	0	100	21.0	79.0	46.7	53.4	47.3	52.7	58.5	41.5	69.2	30.8
More able	61.2	38.8	100	0	46.1	53.9	70.2	29.8	85.4	14.6	72.5	27.5
Less able	29.8	70.2	0	100	49.1	50.9	60.7	39.3	66.7	33.3	76.2	23.8
Male	41.8	58.2	31.2	68.8	100	0	65.8	34.2	75.3	24.7	70.3	29.7
Female	38.4	61.6	33.8	66.2	0	100	62.0	38.0	70.4	29.6	79.2	20.8
Ethnic majority	55.6	44.4	35.8	64.2	49.6	50.4	100	0	88.5	11.5	81.1	18.9
Ethnic minority	12.6	87.4	26.8	73.2	45.5	54.5	0	100	44.9	55.1	64.1	35.9
Mother more educated	51.8	48.2	38.2	61.8	49.7	50.3	77.6	22.4	100	0	77.6	22.4
Mother less educated	8.7	91.3	17.4	82.6	43.6	56.4	26.8	73.2	0	100	67.8	32.2
Not stunted in infancy	44.6	55.4	31.5	68.5	45.1	54.9	69.0	31.0	75.4	24.6	100	0
Stunted in infancy	26.3	73.7	35.8	64.2	56.9	43.1	48.2	51.8	65.0	35.0	0	100

Note. Proportion of children in each of the row categories who are in each of the column categories. Proportions add up to 100 for each of the advantage/disadvantage column pairs. For example, row 1 shows that among the children in the "richer" category, half are in the "more able" category and half are in the "less able" category. See table 4 note for details on how each category is measured.

TABLE A3

INDIVIDUAL AND SCHOOL SAMPLE SIZES IN OVERLAP GROUP IN EACH OF THE ADVANTAGE-DISADVANTAGE CATEGORIES

	Vietnam		Peru	
	Children (1)	Schools (2)	Children (3)	Schools (4)
Richer	420	48	175	36
	(474)		(219)	
Poorer	510		127	
	(655)		(328)	
More able	346	61	144	47
	(357)		(178)	
Less able	660		223	
	(772)		(369)	
Male	568	77	212	63
	(578)		(263)	
Female	540		232	
	(551)		(284)	
Ethnic majority	. . .	. . .	128	25
			(198)	
Ethnic minority	. . .		90	
			(349)	
More educated mother	816	61	235	48
	(921)		(309)	
Less educated mother	185		145	
	(208)		(238)	
Not malnourished in infancy	775	63	264	54
	(900)		(410)	
Malnourished in infancy	217		120	
	(229)		(137)	

Note. Overlap group at the child level (cols. 1, 3) includes children who are in schools that have both advantaged and disadvantaged children for each definition of advantage. Overlap group at the school level (cols. 2, 4) includes all the schools in the sample that have both advantaged and disadvantaged children for each definition of advantage. Total number of children in each of the advantage and disadvantage groups is in parentheses. See table 4 note for details on how each category is measured.

TABLE A4

TESTS FOR EQUALITY OF NONSCHOOL VARIABLE COEFFICIENTS IN EQUATION (3) ACROSS ADVANTAGED AND DISADVANTAGED CHILDREN

	Vietnamese		Spanish	
Advantage Indicator	Math Score	Vietnamese Score	Math Score	Spanish Score
Being richer (top 2 wealth quintiles)	1.22	1.49*	1.09	.55
	(.21)	(.053)	(.25)	(.97)
Being more able (top 2 CDA quintiles)	.87	1.28	1.08	1.26
	(.65)	(.16)	(.36)	(.18)
Being male	.74	.87	1.40	1.32
	(.83)	(.65)	(.092)	(.13)
Being an ethnic majority	. . .	. . .	1.56*	1.98***
			(.06)	(.01)
Having a mother with at least completed primary schooling	1.30	1.36	1.49*	1.34
	(.15)	(.11)	(.07)	(.13)
Having not been malnourished in infancy	1.55**	1.28	3.91***	3.22***
	(.04)	(.15)	(.00)	(.00)

Note. *F*-test statistics and *p*-values in parentheses for tests of joint significance of interactions of nonschool variables with each of the advantage indicators used in the main analysis. See table 3 for complete set of nonschool variables and details of how they are measured. See table 4 note for details on how each advantage category is measured. CDA = Cognitive Development Assessment.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

TABLE A5

DIFFERENCES IN MAIN CHARACTERISTICS BETWEEN SUBSAMPLE OF CHILDREN INCLUDED IN THE SCHOOL SURVEYS (ANALYSIS SAMPLE IN THIS ARTICLE) AND WHOLE YOUNG LIVES SAMPLE OF YOUNGER COHORT CHILDREN

	Vietnam					Peru				
	Analysis Subsample		Whole Sample in Round 3		Difference	Analysis Subsample		Whole Sample in Round 3		Difference
	Mean	SD	Mean	SD	(SE)	Mean	SD	Mean	SD	(SE)
From household survey (time-invariant variables):										
Male (dummy)	.51	.50	.52	.50	−.00 (.02)	.48	.50	.50	.50	−.02 (.02)
Ethnic minority (dummy)	.08	.27	.14	.35	−.06*** (.01)	.36	.48	.31	.46	.05** (.02)
Father: years of schooling	7.69	3.53	7.41	3.88	.29** (.14)	9.23	3.77	9.15	3.95	.09 (.19)
Mother: years of schooling	7.17	3.30	6.82	3.81	.35*** (.13)	7.84	3.11	7.81	4.54	.02 (.21)
Measured in infancy (round 1):										
Birth size (maternal assessment; 1 = very large to 5 = very small)	3.08	.66	3.07	.68	.02 (.03)	3.11	.96	3.12	.99	−.00 (.05)
Health better than other children (dummy, maternal assessment)	.26	.44	.21	.41	.02 (.02)	.41	.49	.40	.49	.02 (.02)
Age in months	12.82	2.61	11.64	3.17	1.18*** (.11)	11.93	3.60	12.04	3.56	−.11 (.17)
Childcare: looked after by others /in crèche (dummy)	.42	.49	.38	.49	.04** (.02)	.21	.41	.24	.43	−.04* (.02)
Maternal mental health	4.40	4.02	4.36	3.91	.04 (.15)	5.78	4.26	5.68	4.29	.10 (.21)
Wealth index	.47	.20	.44	.22	.02*** (.01)	.43	.19	.43	.19	.01 (.01)

TABLE A5 (*Continued*)

	Vietnam					Peru				
	Analysis Subsample		Whole Sample in Round 3		Difference (SE)	Analysis Subsample		Whole Sample in Round 3		Difference (SE)
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Measured at age 5 (round 2):										
Height-for-age *z*-score	−1.30	.97	−1.34	1.04	.05 (.04)	−1.53	1.02	−1.54	1.12	.01 (.05)
PPVT Score (IRT transformed; whole sample mean = 300, SD = 50)	305.33	44.33	300.50	48.09	4.84*** (1.75)	301.34	44.84	300.10	48.95	1.24 (2.24)
CDA Score (IRT transformed; whole sample mean = 300, SD = 50)	306.80	46.56	300.17	49.42	6.63*** (1.79)	299.35	46.39	300.00	49.99	−.65 (2.28)
Time spent in preschool (hours per day)	5.79	2.45	5.05	3.07	.74*** (.10)	3.44	1.83	3.60	1.87	−.16 (.09)
Only child (dummy)	.22	.42	.24	.43	−.02 (.02)	.19	.40	.19	.39	.00 (.02)
Household size	4.57	1.39	4.67	1.51	−.09* (.05)	5.52	2.23	5.51	2.08	.01 (.11)
Log per capita real consumption (in local currency)	5.77	.54	5.76	.61	.01 (.02)	5.02	.67	5.01	.66	.01 (.03)
Area of land owned (hectares)	.42	1.07	.48	1.03	−.06 (.04)	1.09	3.34	2.54	19.69	−1.45*** (.47)
Number of observations	1,129		1,965			547		1,962		

Note. Asterisks indicate significance level of *t*-test of difference in means (analysis subsample − whole sample). Ethnicity minority status: Peru—mother's first language is not Spanish (i.e., is Quechua, Aymara, other); Vietnam—child reported by mother as not belonging to the Kinh group (i.e., is H'mong, Dao, Tay, other). Maternal mental health is measured using the Self-Reporting Questionnaire (SRQ20), a screening tool for risk of common mental disorders (rCMD) developed by the World Health Organization for developing countries (Beusenberg and Orley 1994). It is a scale consisting of 20 statements with yes/no answer options. Score reflects the number of yes responses out of possible 20; a higher score reflects higher risk of rCMD. Wealth index is a simple average of three individual indexes ranging from 0 to 1, including housing quality (rooms per person, floor and roof quality), consumer durables (mean of ownership of radio, bicycle, TV, car/truck, motorbike/scooter, landline telephone, bed/table), and access to services (electricity; piped water into dwelling/yard; access to own pit latrine/flush toilet; cooking fuel is electricity, gas, or kerosene). The Peabody Picture Vocabulary Test (PPVT) measures receptive vocabulary. The Cognitive Development Assessment (CDA) measures children's perceptions of quantity—testing understanding of concepts such as few, most, half, many, equal, and pair. IRT = Item Response Theory.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in Chicago Public High Schools." *Journal of Labor Economics* 25, no. 1:95–135.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton, NJ: Princeton University Press.

Alderman, Harold, and Hoyt Bleakley. 2013. "Child Health and Educational Outcomes." In *Education Policy in Developing Countries*, ed. P. Glewwe. Chicago: University of Chicago Press.

Altonji, Joseph G., and Richard Mansfield. 2011. "The Contribution of Family, School and Community Characteristics to Inequality in Education and Labor Market Outcomes." Photocopy, Department of Economics, Yale University.

Araujo, Maria C., Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *Quarterly Journal of Economics* 131, no. 3:1415–53.

Banerjee, Abhijit, and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Behrman, Jere. 2010. "Investments in Education: Inputs and Incentives." In *Handbook of Development Economics*, vol. 5, ed. D. Rodrik and M. Rosenzweig. Amsterdam: North-Holland.

Beusenberg, M., and John Orley. 1994. "A Users Guide to the Self Reporting Questionnaire." Division of Mental Health, World Health Organization, Geneva.

Botelho, Fernando, Ricardo Madeira, and Marcos Rangel. 2015. "Racial Differences in Performance Assessments in Grading: Evidence from Brazil." *American Economic Journal: Applied Economics* 7, no. 4:137–52.

Chay, Kenneth, Patrick McEwan, and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review* 95, no. 4:1237–58.

Crouch, Luis, Martin Gustafsson, and Pablo Lavado. 2009. "Measuring Educational Inequality in South Africa and Perú." In *Inequality in Education: Comparative and International Perspectives*, ed. D. B. Holsinger and W. J. Jacob, 461–84. Dordrecht: Springer.

Cueto, Santiago, Juan Leon, Gabriela Guerrero, and Ismael Munoz. 2009. "Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives." Technical Note no. 15, Young Lives, Oxford.

Currie, Janet, and Duncan Thomas. 2001. "Early Test Scores, School Quality and SES: Long-Run Effects on Wage and Employment Outcomes." *Research in Labor Economics* 20:103–32.

Das, Jishnu, and Tristan Zajonc. 2010. "India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement." *Journal of Development Economics* 92, no. 2:175–87.

Dee, Thomas. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95, no. 2:158–65.

Duncan, Greg, and Richard Murnane. 2011. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage.

Dunn, Lloyd, and Leota Dunn. 1997. *Examiner's Manual for the PPVT-III: Form IIIA and IIIB*. Circle Pines, MN: AGS.

Dunn, Lloyd, Eligio Padilla, Delia Lugo, and Leota Dunn. 1986. *Manual del Examinador para el Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana* [Peabody Picture Vocabulary Test: Hispanic-American adaptation]. Circle Pines, MN: AGS.

Escobal, Javier, Jaime Saavedra, and Renos Vakis. 2012. *¿Está el piso parejo para los niños en el Perú? Medición y comprensión de la evolución de las oportunidades*. Lima: Banco Mundial, GRADE.

Filmer, Deon, and Lant Pritchett. 1999. "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries." *Population and Development Review* 25, no. 1:85–120.

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, no. 2: 447–64.

———. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2, no. 3:210–40.

Glewwe, Paul, Eric Hanushek, Sarah Humpage, and Renato Ravina. 2013. "School Resources and Educational Outcomes in Developing Countries." In *Education Policy in Developing Countries*, ed. P. Glewwe. Chicago: University of Chicago Press.

Glewwe, Paul, and Hanan Jacoby. 2004. "Economic Growth and the Demand for Education: Is There a Wealth Effect?" *Journal of Development Economics* 74, no. 1:33–51.

Glewwe, Paul, and Michael Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In *Handbook of the Economics of Education*, ed. E. Hanushek and F. Welch. Amsterdam: North-Holland.

Glewwe, Paul, and Edward Miguel. 2008. "The Impact of Child Health and Nutrition on Education in Less Developed Countries." In *Handbook of Development Economics*, vol. 4, ed. T. P. Schultz and J. Strauss. Amsterdam: Elsevier.

Greene, William. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science* 320, no. 5880:1164–65.

Hamre, Bridget, Bridget Hatfield, Robert Pianta, and Faiza Jamil. 2014. "Evidence for General and Domain-Specific Elements of Teacher-Child Interactions: Associations with Preschool Children's Development." *Child Development* 85, no. 3: 1257–74.

Hanna, Rema, and Leigh Linden. 2012. "Discrimination in Grading." *American Economic Journal: Economic Policy* 4, no. 4:146–68.

Hoff, Karla, and Priyanka Pandey. 2006. "Discrimination, Social Identity, and Durable Inequalities." *American Economic Review* 96, no. 2:206–11.

Holsinger, Donald. 2005. "Inequality in the Public Provision of Education: Why It Matters." *Comparative Education Review* 49, no. 3:297–310.

Jacoby, Hanan G. 1994. "Borrowing Constraints and Progress through School: Evidence from Peru." *Review of Economics and Statistics* 76, no. 1:151–60.

Krueger, Alan, and Mikael Lindahl. 2001. "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39, no. 4:1101–36.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2011. "The Contribution of School Quality and Teacher Qualifications to Student Performance." *Journal of Human Resources* 46, no. 1:123–53.

McEwan, Patrick. 2004. "The Indigenous Test Score Gap in Bolivia and Chile." *Economic Development and Cultural Change* 53, no. 1:157–90.

McEwan, Patrick, and Marisol Trowbridge. 2007. "The Achievement of Indigenous Students in Guatemalan Primary Schools." *International Journal of Educational Development* 27:61–76.

Murray, Helen. 2012. "Is School Education Breaking the Cycle of Poverty for Children? Factors Shaping Education Inequalities in Ethiopia, India, Peru and Vietnam." Policy Paper 6, Young Lives, Oxford.

OECD (Organization for Economic Cooperation and Development). 2013. *PISA 2012 Results: What Students Know and Can Do*. Vol. 1, *Student Performance in Mathematics, Reading and Science*. Paris: OECD.

Rivkin, Steven, Eric Hanushek, and John Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73, no. 2:417–58.

Rockoff, Jonah. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94, no. 2:247–52.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay and Student Achievement." *Quarterly Journal of Economics* 125, no. 1:175–214.

Todd, Petra, and Kenneth Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1, no. 1:91–136.

UNESCO. 2011. *World Data on Education*. 7th ed., 2010/11. Paris: UNESCO/IBE. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Viet_Nam.pdf.

World Bank. 2004. *Vietnam Reading and Mathematics Assessment Study*. Vol. 1. Washington, DC: World Bank.

———. 2011. *Viet Nam: High Quality Education for All by 2020*. Vol. 1, *Overview/Policy Report*. Washington, DC: World Bank.

———. 2012. *World Development Indicators*. Washington, DC: World Bank. http://data.worldbank.org/data-catalog/world-development-indicators.

———. 2013. *World Development Indicators: Size of the Economy*. Washington, DC: World Bank. http://wdi.worldbank.org/table/1.1.