The Pennsylvania State University

The Graduate School

The Department of Human Development and Family Studies

ETHNIC GROUP DISPARITIES IN ACADEMIC ACHIEVEMENT ACROSS FOUR LOW-INCOME COUNTRIES

A Thesis in

Human Development and Family Studies

and Demography

by

Jessica Heckert

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

May 2011

The thesis of Jessica Heckert was reviewed and approved* by the following:

Rukmalie Jayakody
Associate Professor of Human Development and Family Studies and Demography
Thesis Advisor

Leif Jensen
Professor of Rural Sociology and Demography

Douglas M. Teti
Professor of Human Development and Family Studies and Psychology
Graduate Professor-in-Charge, Human Development and Family Studies

*Signatures are on file in the Graduate School.

ABSTRACT

Many countries experienced dramatic economic growth in recent decades, but within-country growth was unequally distributed, and disparities persist by race, ethnicity, and social class. Although education should promote equity across social groups, this idealization falls short when educational resources are unequally distributed. Instead, educational disparities perpetuate within-country inequality. I examine disparities in academic skill development by social group in Ethiopia, India, Peru, and Vietnam using data from the Young Lives Study, a longitudinal study of childhood poverty. I examine characteristics that predict academic achievement on three domains: math, literacy, and total completed grades, within each country. Then, I decompose the differences within each country to highlight the particular factors contributing to observed disparities. I discuss the findings in terms of policy implications to reduce educational disparities.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1. Introduction and Literature Review

Introduction

Dramatic economic growth in recent decades brought about declining poverty rates and living standards improvements (Ranis, Stewart, Ramirez, 2000). Despite these improvements, growth was unequally distributed, and large differences in social indicators persist within poorer countries (Cornia, 2003). Education is one way that inequalities manifest during childhood. To better understand within-country inequality this thesis examines academic skill acquisition during middle childhood and explores how social group membership perpetuates academic skill disparities.

Acquiring basic academic skills is a primary developmental task of middle childhood, and an important predictor of future productivity (Psacharopoulos & Patrinos, 2004). Ideally, education should improve equity, but instead, hierarchies exist whereby particular groups are undervalued and others maintain powerful influences. Current educational practices perpetuate these hierarchies and ensure continued inequality (Bechmann & Hannum, 2001). This phenomenon encourages the distribution of poverty to persist across generations.

Universal primary education became a central concern following the *Declaration of the Rights of the Child* in 1959, and many countries realized that national economic well-being necessitated childhood investments (Burnett, 2008). However, the dramatic improvements made in both primary and secondary education in the 1960s and 70s dwindled beginning in the 80s (Weisbrot, Baker, & Rosnick, 2007). In particular, substantial inequalities within countries persist, and aggregate enrollment indicators mask low performance and under-enrollment among groups with lower educational achievement (Ben-Arieh, Kaufman, Andrews, Goerge, Lee, & Aber, 2001; Burnett 2008). For example, in the current sample over 95 percent of children even in the most marginalized groups enrolled in school at some point. However, persistent school enrollment and performance indicators reveal a more nuanced story where disparities persist within each country.

Given differential investment in education, all enrolled children fail to experience equitable educational gains. These disparities are particularly prominent along divisions such as race, ethnicity, and caste. Therefore in this thesis I examine evidence from four settings with unique histories of social group inequalities in different stages of economic growth, including Ethiopia, Peru, Vietnam, and India's Andhra Pradesh state. A common feature of these four settings is that children from marginalized social groups encounter unequal educational opportunities and resources. Specifically, I examine aspects of educational inequality that are potentially modifiable or suggest direct policy implications, including family economic conditions, family characteristics, geographic distribution, language, and parental educational attitudes. I then ask to what extent do variations in these characteristics and differential returns among them contribute to disparities in the total number of completed grades, math skills, and reading skills.

I begin the literature review by describing marginalization and academic achievement within each of the four selected settings. I then proceed by examining previous work that examines potential causes of educational disparities within these contexts.

Country Contexts

Describing social group divisions in the four target countries provides a basis for understanding how social group membership generates educational achievement disparities. The following section explains national social groups divisions and provides evidence of educational disparities within each of the countries.

Peru

Peru's racial composition results from 500 years of *mestizaje*, mixing of Spanish colonizers, Indigenous inhabitants, and African slaves. Estimates[1] state the population is approximately 45 percent Indigenous, 37 percent *mestizo* (mixed), 15 percent white, and 3 percent other (primarily Afro-Peruvian, Chinese, and Japanese).

Research examining educational inequality within Peru emphasizes the disadvantaged status of Indigenous students. Indigenous status throughout Latin America predicts late school entry and grade repetition (Patrinos & Psacharopoulps, 1996), and although differences in completed schooling years between Indigenous and non-Indigenous Peruvians shrank between 1994 and 2004, a 2.3 year disparity persists (Hall & Patrinos, 2005). For those who do remain in school, Indigenous children achieve lower academic benchmarks; achievement test scores in both Bolivia and Chile demonstrate that they fall behind their same-grade non-Indigenous peers (McEwan, 2004).

Vietnam

Ethnicity divides social groups in Vietnam. The Kinh comprise 86 percent of the population, inhabit the fertile lowlands and river deltas, and garner better access to resources and infrastructure. The 53 ethnic minority groups include those inhabiting the region for centuries, such as the Tay, and more recent geopolitical migrants, such as some Hmong. Most inhabit isolated rural areas and speak minority languages.[2] Vietnam's economic growth following 1986

[1] This information is not reported in national census data.
[2] An exception is the Hao, or ethnic Chinese, who assimilated with the urban Kinh and flourished economically (Baulch, et al., 2004).

renovation policies marked a dramatic turnaround, but ethnic minorities benefitted much less and maintain substantially higher poverty rates (van de Walle & Gunewardena, 2001).

Ethnic minority status in Vietnam reveals substantial educational disparities. Ethnic minorities' net primary school enrollment rates are a sixth below the national average, and lower secondary school rates are half the national average (Baulch, Chuyen, Haughton, & Haugton, 2004). While only three percent of majority households have no literate members, 12 percent of ethnic minority households are so disadvantaged. Additionally, ethnic minorities' concentration in remote rural areas further impedes their access to schools; while majority households live an average 0.2 kilometers away from the closest lower secondary school and 6.1 kilometers from an upper secondary schools, minority households live 2.4 and 10.3 kilometers away (van de Walle & Guneweardena, 2001).

Ethiopia

Recent Ethiopian history is beset with ethnic inequality and conflict, such as the 1991 coup d'état (Mengisteab, 2001). Contemporary Ethiopia contains over 80 documented ethnic groups, but I limit this discussion to the Amhara and the Oromo. The Oromo are the largest ethnic group, comprise 32 percent of the population, but experience political and social marginalization. The next largest, the Amhara, are 30 percent of the population and have politically, socially, and linguistically dominated Ethiopia since colonizing their southern and western neighbors, including the Oromo, in the late 19^{th} century (Tronvoll, 2000).

Previous research documents ethnic inequality within Ethiopia in non-educational domains, such as political representation and health (Mengisteab, 2001). Historically, the Oromo homelands received fewer educational allocations, and education occurred in Amharic, the nationally dominant language (Tronvoll, 2000). Ethnic divisions within Ethiopia and ethnic disparities in health and economic indicators among other countries in the region, (Brockerhoff & Hewitt, 2000; Charasse-Pouélé & Fournier, 2006) suggest that educational disparities by ethnicity within Ethiopia likely exist.

India, Andhra Pradesh state

The 1949 Indian Constitution outlawed the caste system, but its vestiges still permeate Indian society, whereby, historically, birth determined membership, dictated employment, and allocated resources. Scheduled Tribes (ST) and Scheduled Castes (SC) are the most marginalized groups, followed by Other Backwards Classes (OBC) and members of higher classes, referred to as other. According to the 2001 Census, 16, 8, and 52 percent of the population are members of SCs, STs, and OBCs respectively.[3]

Despite engaging multiple strategies throughout the 1990s, educational reform for lower castes was largely insufficient, and remarkable disparities remain in educational attainment rates (Chanana, 1993; Rao, Cheng, and Narain, 2003). Members of lower castes report fewer completed years of education, even in more egalitarian states (Deshpande, 2000). Among higher castes, 58 percent are literate, but among lower castes and tribal groups, only 37 and 30 percent are literate (Dreze & Loh, 1995).

Possible Mechanisms

Multiple intertwined processes contribute to educational disparities among marginalized groups. Marginalized groups are generally poorer, which is strongly associated with less parental education and larger family size (McEwan, 2004). Poorer children also receive less nourishment (Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, Strupp, 2007) and contribute more

[3] Marginalization within India occurs on multiple domains, for example, by religion. However, given conditions in Andhra Pradesh state and data limitations, I limit the examination of disparities to caste.

time to family production strategies (Hall & Patrinos, 2005). Both of these contribute to cognitive development and schooling. Furthermore there is concern that minority families expect fewer returns to educational investments or undervalue its usefulness (van de Walle & Gunewardena, 2001). Additionally, geographic concentration limits access to resources and power, and traditional languages are undervalued and underused in current educational systems (McEwan, 2004). The following section elaborates on the processes occurring within these domains and provides examples from different national settings.

One explanation for educational disparities is simply that marginalized social groups are poorer, and poor families cannot afford direct and indirect educational costs. Poverty rates are higher among indigenous Peruvians, Vietnamese ethnic minorities, lower castes in India, and marginalized ethnic groups in Ethiopia (Trivelli, 2005; Baulch, Chuyen, Haughton, & Haugton, 2004), and direct schooling costs, including books, materials, uniforms, transportation, and meal costs strain families' limited resources. Strong positive associations between household income and child educational attainment persist, even in Vietnam, where the socialist government has emphasized educational equality via progressive school fees; the prohibitive cost of uniforms, books, school supplies, and transportation undermine the potential benefits of reduced school fees (Behrman & Knowles, 1999).

Limited resource accesses directly impairs educational outcomes by inhibiting cognitive development due to undernourishment in early childhood. Extensive studies on the cognitive consequences of poor nutrition across multiple middle and low-income countries find that early undernourishment, prevalent among poor families, leads to inadequate brain development, poorer cognitive skills, fewer years of education, decreased knowledge accumulation per year of school, and lower earnings in adulthood (Grantham-McGregor et al., 2007).

When studying poverty, it is challenging to untangle the multiple correlates such as large family size and low parental education. Larger families spread resources among more children. The fertility rate among ethnic minorities in Vietnam is 25 percent higher than among the Kinh (Baulch, Chuyen, Haughton & Haughton, 2002), and higher fertility persists among indigenous Latin Americans despite regional fertility decreases (Rodríguez Wong, 2009). Yet, family size consequences are likely limited. Only Vietnamese children in families with four or more children achieved less education, and the effects were limited to secondary and tertiary education (Anh, Knodel, Lam, & Friedman, 1998).

In addition to limited economic resources, poor families often have limited parental education. Higher grade repetition occurs among children in families with lower parental education, and teachers may discriminate against students with uneducated parents (Patrinos & Psacharopoulps, 1996). Less educated parents experience difficulty negotiating the educational system and may find school does not meet intended family goals. For example, female education may have limited value if a daughter is groomed primarily for marriage (Chanana, 1993). Moreover, group membership is associated with lower educational returns and labor market discrimination, suggesting that parents may correctly perceive education as less valuable (Trivelli, 2005; Zoninsein, 2001; van de Walle & Gunewardena, 2001).

Beyond direct educational costs, poor families suffer from opportunity costs. When school attendance replaces child labor, paid or unpaid, families receive fewer direct benefits from the child. Indigenous children in Latin America work at higher rates, and working children learn less while in school (Hall & Patrinos, 2005; Psacharopoulos, 1997). However, evidence from Ethiopia indicates that children will work regardless of school enrollment, suggesting these activities do not directly replace school but compliment it (Rose & Al-Samarrai, 2001). Yet, household and labor responsibilities expend energy and compete with school tasks, limiting students' ability to engage in school and homework.

Spatial economic variation further disadvantages minority groups that are concentrated in areas with poor infrastructure and limited production potential. For example, Vietnamese ethnic minorities, indigenous Peruvians, and India's STs are more heavily concentrated in difficult to access mountainous areas, and limited infrastructure characterizes the Oromo homelands in Ethiopia. Within urban areas, marginalized populations are also commonly concentrated in slums or other areas with limited infrastructure and resources, characteristic among lower castes in urban India and rural to urban Peruvian migrants.

Rural areas have traditionally lagged behind in primary education access. Parents incur higher costs and rural areas attract less well trained teachers (Ilon & Moock, 1991; McEwan, 2004). Even amidst high primary education enrollment rates, rural schools receive fewer resources, compromising educational quality (Trivelli, 2005). Furthermore, national policy makers often overlook indigenous regions, and even as their economic resources increase, national resources are not directed to areas with high indigenous concentrations (Hall & Patrinos, 2005).

Language further challenges those who speak minority languages. Children's preferred languages may have lower status and dominant languages may be the primary means of student-teacher communication. Schools historically discouraged, even punished, local language use, and although instructional language preference and bilingual education are increasing, they are still relatively rare and linguistic barriers persist (McEwan, 2004). Finding well trained teachers who speak local languages is challenging and supporting materials in local languages are underdeveloped (Aikman & Pridmore, 2001; Wagaw, 1999). Furthermore language preferences may limit parental navigation of the educational system.

Finally, studying additional languages increases the learning burden at school. For example, Ethiopian children in non-Amharic speaking regions receive language instruction in the regional language, later adding Amharic, the *de facto* national language, and then English, the secondary education instructional language (Wagaw, 1999).

In summary, previous studies reveal that marginalized social groups across diverse contexts complete fewer grades in school and learn less during the years that they are enrolled. Poorer performance among marginalized groups is associated with the unequal distribution of family economic resources, family characteristics such as parental education and family size, residing in areas with fewer schooling resources, speaking non-dominant languages, and valuing education less. In this thesis I examine how variations in the aforementioned characteristics contribute to academic disparities among young people in Ethiopia, India, Peru, and Vietnam. I examine these differences in terms of the number of completed grades and math ability at age 12 and literacy at age eight. In the following sections, I examine how much these characteristics differ between groups within each country, the extent to which these factors contribute to each outcome, and whether academic disparities are associated with variations in these characteristics or differing returns to them.

Chapter 2. Method

Data

I employ data from the Young Lives Study, a unique longitudinal study of childhood poverty conducted in Ethiopia, India-Andhra Pradesh state, Peru, and Vietnam (Huttly, Jones, & Boyden, 2009). Target children and their families were selected and initially interviewed in 2002 at age eight and again in 2006 at age 12. I draw on responses from both rounds. Children and caregivers provided information about household resources, family member characteristics, attitudes and perceptions, and the child's schooling, health, and activities. Child participants also completed a battery of items measuring basic academic skills.

Sampling Approach & Corrections

Within each country, the sampling procedure selected respondents using a multistage stratified design. The process selected 20 sampling clusters per country, and incorporated both random and purposive procedures to oversample poor households and represent geographical and social diversity within the country. The relatively small sample, goal of sample diversity, and logistical feasibility limited the ability to fully randomize cluster selection, eliminating more remote and volatile districts. Once the sampling design selected districts, it randomly selected sampling clusters from within the districts. Within sampling clusters, approximately 50 respondent households were randomly selected from households with children between 7.5 and 8.5 years old.

District selection procedures varied by country, and in the cases of Ethiopia and Andhra Pradesh, India, I limit the sample to particular subgroups for reasons I describe below. I conducted all analyses using STATA 11.0 SVY procedures, which correct for complex sample design (Stata Corp, 2009). However purposive sample selection procedures are statistically uncorrectable. Due to variations among countries, I make different adjustments for each country. Below I explain country-specific conditions. Few cases contained incomplete data (< 5%), and I therefore employ list-wise deletion.

In the case of Peru, districts were stratified by poverty status, and higher poverty districts had a greater probability of inclusion. Within each stratum, each district's probability of inclusion was equivalent to its population size. Ten possible samples were randomly generated, and the survey designers selected one they felt was feasible and adequately represented the country.

In the case of Vietnam, four regions were selected for inclusion. Within each region districts were selected in part by randomization and in part by consultation with local officials.

In the case of Ethiopia sampling site selection was not randomized and determined by balancing logistical feasibility and national diversity. Therefore, I only correct for geographically clustered responses and do not account for complex sample designs. As a result of the sampling procedure, the sample includes more-advantaged rural areas and poorer urban areas, limiting representativeness (Outes-Leon & Sanchez, 2008). I limit the Ethiopian sample to Oromo and Amhara respondents. The Amhara and Oromo reflect the two largest ethnic groups within the country and are both well represented in the data

In the case of India, the sample was regionally stratified. Then districts within each region were chosen through a combination of randomization and purposive selection. Additionally, beginning in the late 1990s, before the index children began school, the Andhra Pradesh state initiated a large scale, multi-systemic approach to improve human development indicators for Scheduled Castes (SC) and Scheduled Tribes (ST) (Government of Andhra

Pradesh, 2003). Specifically, the sample districts and the intervention districts overlap considerably, and a major component included sending SC and ST children to residential schools (World Bank OPCS, 2005; Kumra, 2008). As a result, children from the most vulnerable circumstances excelled on both health and educational indicators. Both the current sample and Andhra Pradesh State reports document these improvements, whereby students in the targeted residential schools earn high statewide exam scores (Government of Andhra Pradesh, 2003). Though these improvements are promising, they complicate data analysis. The current study aims to understand inequalities, and SC and ST children were not heavily targeted outside these areas. Therefore, I exclude from the analyses the four sampling clusters where a large portion of the children attended residential boarding schools. I revisit this topic in the discussion.

Measures

Academic achievement

Math. Children completed a ten-item arithmetic inventory at age 12. I first exclude items with differential item functioning by gender or test administration language (Cueto, Leon, Guerrero, & Muñoz, 2009). I then estimate a standardized composite score using Item Response Theory (IRT) procedures and settle on a three parameter model using PARAM-3PL Version 0.89 (Rudner, 2007). The three parameter model estimates each item’s difficulty, estimates the degree to which it discriminates among respondents of differing abilities, and adjusts for guessing on multiple choice items. Scores for each country are estimated separately.

The scores I calculate differ from the IRT scores provided by the survey designers, because I include respondents with missing items. An advantage of IRT is that it includes estimates for respondents with incomplete response series without assuming that unanswered items are incorrect. IRT procedures produce standard errors of estimates that vary by the estimated score, and these values were unacceptable only for respondents answering less than half the items. These few individuals were eliminated from final analyses.

Total grade. Child respondents reported current school information. Total grade reflects their current grade for those still in school or the most recent grade for those not currently enrolled in school.

Literacy. To assess literacy, interviewers asked children to read and write two separate sentences in the child's selected language. I use age eight literacy scores, except in the case of Ethiopia whereby I use age 12 literacy scores. Interviewers scored responses according to whether the child read letters, words, complete sentence, or nothing and whether the child wrote easily, with difficulty, or not at all. The combined literacy score is scaled from 0 to1. Cronbach's alpha reveals adequate reliability ($\alpha_{Et,Am} = 0.68$, $\alpha_{Et,Or} = 0.71$, $\alpha_{In,S} = 0.66$, $\alpha_{Ind,O} = 0.72$, $\alpha_{Pe,In} = 0.78$, $\alpha_{Pe, Eur} = 0.65$, $\alpha_{Vi,EM} = 0.93$, $\alpha_{Vi,Ki} = 0.70$). Additional analyses of literacy scores at other ages are not possible due to poor reliability from ceiling effects at age 12 in every country but Ethiopia and low school enrollment at age eight in Ethiopia and hence no reading assessment.

Social group category

The adult respondent reported ethnicity, race, or caste information for the index children and their parents. I use responses from the second round of data collection, capitalizing on their improved specificity. In a later section I compare disparities between two discrete groups, and therefore I divide respondents in each country into two groups.

For Ethiopia I include only those who identify as Amhara or Oromo. In the Indian sample I divide the respondents into those who identify as SC or ST and all others. In the case of

Vietnam, I aggregate all ethnic minorities groups and compare them to the Kinh, the majority ethnic group. The limited sample size does not allow for further disaggregation.

Identifying Peruvian social groups with survey data is not straight-forward when Peruvians of vastly different origins self-identify as *mestizo* (Ñopo, Saavedra, & Torero, 2007; Trivelli, 2005). I divide respondents into two groups: 'more-Indigenous' and 'more-European.' The more-Indigenous group contains all children who first spoke an Indigenous language, identify as Indigenous, have at least one parent whose mother tongue is an Indigenous language, or have at least one parent who identifies as Indigenous. A trivial number of Afro-Peruvian descendents (less than one percent) are in the sample, and I classify them with the more-Indigenous group for analytic purposes due to their historically marginalized status. All other children are classified as less-Indigenous. I justify this approach based on previous research. First, when language alone identifies Indigenous households, it underestimates Indigenousness, particularly now that growing migration to urban centers is increasing Spanish-speaking self-identification (Trivelli, 2005), and many Indigenous parents in urban areas encourage Spanish usage among their children to avoid potential discrimination (Anderson, 2007). Additionally, a comparison of self-identification and phenotypic characteristics in labor market discrimination finds discrimination based on physical features even for those who do not self-identify as Indigenous (Ñopo, Saavedra, & Torero, 2007).

Independent Variables

Wealth index. Information on ownership of consumer durables, housing quality, and access to utilities produced a household wealth index at each time point. It is equivalent to the wealth index calculated in the Demographic and Health Surveys. I conduct analyses using the wealth index from the same time point when the dependent variable was observed. The exception is models of age eight literacy in Vietnam, where I use the wealth index from age 12. I do this because of a large amount of missing wealth index data within particular clusters in the first round of data collection, suggesting a data collection error. Using a wealth index from a later time point is justified, both because of the high correlation between wealth indexes at ages 8 and 12 for available data points ($r > .8$), and because, unlike period measures of household income, wealth indexes are relatively stable over time.

Number of Children. An adult respondent reported the total number of still living children born to the child's biological mother at the time the child was eight years old. I classified families as having 1) 1 or 2 children, 2) 3 children, or 3) 4 or more children.

Maternal education. The adult respondent reported maternal education years during the age 12 interview. I classified responses into categories according to the educational systems of each country and then collapsed categories according to their distribution. Categories for India and Peru included 1) no education or incomplete primary school, 2) completed primary school, and 3) completed secondary school. Categories for Ethiopia include 1) no school experience, 2) some school experience, and 3) completed primary education. Categories for Vietnam include 1) no education or incomplete primary school, 2) completed primary school, and 3) completed lower secondary school. I included adult literacy program participants with those not completing primary education, because few reported they could read.

Standardized height for age. Extremely short stature, even in adulthood, reflects early childhood undernourishment. Children were measured at age eight and child height and exact age in months, were transformed into height for age z-scores based on updated standards from the World Health Organization's Multicentre Growth Reference Study (Borghi, de Onis, et al., 2006).

Hours working. Children reported the daily hours they contributed to the family livelihood strategy, including working for money and non-paid contributions such as farming activities, household chores, and caring for others. I combined these contributions into a single variable so that it is similar for boys and girls, and because these activities are similar in their competition for school related time.

Language. A child's primary spoken language is an important component of school experience, particularly when children do not speak the language of instruction. It is also highly confounded with social group identification. However, there is language diversity among more-Indigenous Peruvian respondents. Therefore I include a variable indicating if the child's first language was Spanish.

Parental schooling attitudes. I include variables describing family attitudes towards school and education. The surveys were not consistent in content and meaning across the four settings, and I include different variables for each country.

First, I include a measure of parental school efficacy in Ethiopia, India, and Peru. Parent respondents reported if they believed there was something they could do to help their child do better in school. Next, I include whether parents believed that school was essential in Vietnam. Parents were asked if formal schooling was or would have been essential for them. Finally, I include a variable indicating whether parents thought the child would achieve a high level of education in Ethiopia, India, and Vietnam.

Region and Urbanicity. Both the region of residence and whether the community was rural or urban were identified in the survey using national census information. Urbanicity is defined dichotomously as either rural or urban. Peru is divided into the Coast, Sierra, and Jungle regions, and the sample represents all three regions. The Vietnamese sample represents four of the country's eight geographic regions: the Central Coast, the Mekong River Delta, the Northern Uplands, and the Red River Delta. The Indian state of Andhra Pradesh is divided into the Telangana, Rayalaseema, and Coast regions, and the sample includes districts from all three regions. The Ethiopian sample used in this analysis is limited to four geographic regions: Addis Ababa, which is the capital, the state of Amhara, the state of Oromia, and the Southern Nations, Nationalities, and People's Region.

Descriptive Statistics

I began by calculating the means and proportions for all variables. I calculate them separately for the two groups within each country. Point estimates do not vary when accounting for clustering and sample selection. However, standard errors do vary, and confidence intervals expand when accounting for clustering and sample selection, and Taylor Series approximations of the standard errors are reported.

Multivariate Models

In order to examine conditions associated with academic achievement in a multivariate framework, I fit regression models predicting academic achievement within each group. In the case of each country I predict math ability at age 12, total years of school at age 12, and literacy at age eight, except in the case of Ethiopia, where literacy is assessed at age 12.

For each dependent variable, in the first model I use only an indicator for social group to estimate the effect of group membership without controlling for additional characteristics. In the second model I estimate a fixed effects model by including dummy indicators for N-1 clusters to control for unobservable characteristics unique to the clusters. The third model adds child gender. The fourth model includes household wealth. The fifth model includes indicators for family size and maternal education, factors that are strongly associated with poverty. The sixth model adds standardized height for age and working hours, factors considered direct consequences of poverty. In the case of Peru, an additional model adds an indicator for Spanish as a first language. In the final model I include the variables that measure parental educational attitudes.

Models predicting literacy skills at age eight do not include variables for time spent working or parental school attitudes, because these factors were measured at age 12. Regional and urbanicity variables are not included in the multivariate models due to overlap with the cluster variables.

Oaxaca-Blinder Decomposition

A drawback of using linear regression with a variable indicating group membership, as done in the multivariate models above, is that it assumes the coefficients operate similarly for both groups. Linear regression models assume that two groups have unequal dependent variable means because they differ in the quantity of the independent variables. However, previous findings suggest ethnic minority groups often respond differently to poverty alleviation strategies and may experience differential returns to educational investments. Hence, linear regression assumptions may not be appropriate.

To account for the possibility of differing effects, I decompose means and coefficients using the Oaxaca-Blinder decomposition (Blinder, 1978; Oaxaca, 1978). This method was initially developed to decompose wage differentials by gender and has been applied to other disparities, including test score gaps between Indigenous and non-Indigenous students in Bolivia and Chile (McEwan, 2004), Maori and European descended students in New Zealand (Lock & Gibson, 2008), and Black and White students in the U.S. (Myers, Kim, Mandala, 2004). This technique uses regression estimates for each subgroup to divide the observed gap into an ‘explained’ portion, attributable to differences in characteristics, and an ‘unexplained’ portion, attributable to differing effects of the independent variables. This technique has two advantages. First, it examines how social groups vary within each country both in the quantity of their characteristics and the differing effects of these characteristics. Furthermore this technique identifies the portion of the gap attributable to the differing quantity and differing effects of each included variable.

In this analysis I specify a pooled model, a version of the Oaxaca-Blinder decomposition proposed by Neumark (1988), which pools regression estimates over the two models as a source of comparison. The pooled regression decomposition model is expressed as follows:

$$\bar{y}^A - \bar{y}^B = \Delta\bar{X}\hat{\beta}^{Pooled} + \bar{X}^A(\hat{\beta}^A - \hat{\beta}^{Pooled}) + \bar{X}^B(\hat{\beta}^{Pooled} - \hat{\beta}^B),$$

whereby $\bar{y}^A$ and $\bar{y}^B$ are the mean values of the dependent variables for the two subgroups, $\Delta\bar{X}$ is the absolute value of the difference of the variables in vector $\bar{X}$ for the two subgroups, $\bar{X}^A$ and $\bar{X}^B$ are vectors of the means of the predictor variables for the two subgroups, $\hat{\beta}^A$ and $\hat{\beta}^B$ are estimated slope coefficients for the two subgroups, and $\hat{\beta}^{Pooled}$ is the combined regression estimate for each predictor variable. According to this model, the explained component, or the portion of the difference due to differing quantities of the endowments, is the sum of the last two terms. The unexplained portion of the difference, or the portion due to differing returns to the characteristics, is the first term on the right hand side of the equation. A critique of the pooled model compared to other versions of the Oaxaca-Blinder decomposition is that it incorrectly attributes too large a portion of the disparity to explained components. However, this bias is avoided by including a group membership variable in the pooled version of the models (Jann, 2008).

I decomposed the differences in means and coefficients using the Oaxaca command (Jann, 2008), again adjusting for the survey design using SVY in STATA 11.0 (Stata Corp, 2009). In the pooled model, results are divided into explained and unexplained components for each variable or group of variables. In the following section I explain their interpretation.

Interpretation of Decomposition Components

As mentioned above, the decomposition is divided into the explained and unexplained components. The explained component is the portion of the gap due to differing quantities of the independent variables. For example if two groups experience similar returns to maternal education, but one group reports substantially lower levels of maternal education, the differing levels of maternal education will contribute to the explained portion of the achievement gap.

The unexplained component is the portion of the gap attributable to the differing effects of the variables included in the model. If two groups have similar levels of maternal education but the effect of high maternal education is larger in one group, this difference in returns would contribute to the unexplained component. For example, a highly educated mother from one group may gain access to resources more easily than a highly educated mother from another group.

Often the unexplained component is considered discrimination, but it can also be due to omitted variables. Furthermore, discrimination can be conceptualized as either current discrimination or historically occurring discrimination towards a group whereby the two groups' behaviors adapted in different contexts, creating differing returns for similar characteristics.

Confusion often occurs when a portion of the explained or unexplained component is negative. A negative component suggests that the particular variable or group of variables works to the advantage of the more marginalized group. A negative explained component indicates that the more marginalized group has more favorable characteristics in this regard. A negative unexplained component suggests that certain aspects of discrimination advantage the marginalized group.

Chapter 3. Results

Descriptive Findings

I begin by describing the unadjusted differences on the three outcome variables: total grades, math score, and literacy, within each country (see Tables 1 to 4). In all four countries, the more marginalized groups perform lower on all three indicators. The largest gaps are present in Vietnam where for example Kinh children are more than two-thirds of a school year ahead of ethnic minority children and surpass them on literacy and math skills as well. In Peru, the more-European children perform better than more-Indigenous children on all three indicators, but although they are only a third of a grade ahead, notable differences in skill levels persist. In India children from SCs and STs perform below other children on math and literacy skills, whereas only examining total grades reveals an insignificant difference, although one that favors non-SC and ST children. The Ethiopian results are less clear. Amhara children perform substantially better than Oromo children on the literacy test. However, once accounting for the sampling design, there is no significant difference between the two groups on math scores and the total number of grades completed in school.[4]

Findings reveal that children from marginalized groups in India, Peru, and Vietnam live in households that report less wealth. Furthermore, their mothers were more likely to have never attended school or not completed primary school, whereas the maternal counterparts in non-marginalized groups more often completed primary or secondary school. However, wealth and maternal education differences in Ethiopia were not significant and not in the expected direction. Marginalized children in Peru and Vietnam are more often growing up in families with larger

[4] Differences for total completed grades and math scores are highly significant before accounting for sampling clusters. However, the high degree of similarity among respondents in each cluster leads to insignificant differences.

numbers of children, possibly stretching parental resources, whereby family size in Ethiopia and India only varied trivially between the two groups.

Marginalized children in India, Peru, and Vietnam were smaller than their counterparts, reflecting poorer early childhood nutrition. They also dedicated more time to family livelihood strategies, such as working for pay and household tasks. However, in Ethiopia, the reverse was true, and Oromo children are somewhat taller than Amhara children.

In general parents from the two groups differed very little in their expectations about their children's education. The exception is Vietnam where notably more Kinh parents report that they believe education is essential for success. In every case but Ethiopia, children from marginalized groups more heavily concentrated in rural areas, and the regional distributions reveal that groups are concentrated in particular areas within each country.

Multivariate Results

I now turn to the multivariate results, where I first examine the outcomes from Peru (Tables 5 to 7). In the initial models for each of the three outcomes, indigenous status strongly predicts lower academic achievement. However, after accounting for household wealth, family size, and maternal education, the effect of indigenous status weakens considerably. Greater household wealth predicts both literacy and the total number of grades completed in school. However, wealth is not associated with math ability. Being from a larger family predicts lower math ability, but it is only weakly related to total grades and literacy ability. Higher levels of maternal education, particularly secondary school completion are related to both higher math and literacy abilities, after controlling for additional factors. However, maternal education is not associated with total grades. Gender in not significantly associated with either math or literacy ability, but being male is associated with completing fewer grades in school. Standardized height is positively associated with completing more years of school, but it is not related to math or literacy ability. Children who work more have lower math skills, but they have not completed fewer grades. Speaking Spanish first, rather than an indigenous language, predicts better outcomes, even when controlling for other factors. Parental school efficacy was weakly associated with higher math ability, but not associated with the total number of completed grades in school.

Moving to Vietnam (see Tables 8 to 10), in the case of both total grades completed and math ability being an ethnic minority child remains a highly significant predictor of lower achievement in the initial models but reduces significance as the models become more complex. In predicting literacy, ethnic minority status significantly predicts lower literacy skills in all the models. Household wealth is strongly associated with higher school achievement in all models, as are maternal completion of secondary education and smaller family size. More stunted children also have significantly lower levels of academic achievement on all three outcomes. The amount of time children spend working however has only a weak negative association with higher achievement. Parental belief that the child will go far in school is positively associated with higher achievement, though parental confidence and child ability are likely mutually reinforcing phenomena. Parental belief that education is essential is not associated with achievement.

Turning to India (see Tables 11 to 13), SC/ST membership significantly predicts lower achievement on math and literacy skills but is not significantly associated with completing fewer grades. Being male is associated only with higher math scores and does not predict other outcomes. More household wealth is associated with both higher math and literacy ability, but it is weakly and insignificantly associated with the total number of completed grades. Being in a family with four or more children is associated with lower math performance, but family size is not related to the other indicators. Maternal completion of secondary education predicts higher math and literacy achievement but again is not related to completing more grades in school. Standardized height however is strongly and positively associated with completing more grades in school, whereas it is not associated with other outcomes. Children who work more have poorer achievement measured by both completed grades and math ability. Both parental beliefs that the child will achieve a high level of education and beliefs that they have the ability to help the child do better in school are associated with the total number of completed grades and math ability.

I now turn to the results from Ethiopia (see Tables 14 through 16). Being a member of the historically more marginalized Oromo group only predicts lower academic achievement in the case of literacy, before the fixed effects are added to the model. In all other models, this coefficient is not significant. This finding may be partially due to the high geographic concentration of individuals by ethnic group, particularly in rural areas and a high degree of similarity on the academic achievement outcomes between individuals within the same cluster. The intraclass correlations for academic achievement outcomes by cluster are large (ranging from 0.15 to 0.32). These relationships suggest that although individual ethnicity may not be important, residing in a primarily Oromo area may still be an important factor.

I continue the analysis by examining other expected predictors of educational differences at the individual level. Being male is associated with higher achievement only in the case of math ability; it is not significant in any of the other models. Household wealth strongly predicts higher achievement in both the total grades and math ability, but it is not associated with literacy ability. Family size is not significantly associated with better outcomes, and the coefficients, though insignificant, are in the opposite direction from expected, which may suggest that larger family size is associated with better outcomes. The unexpected direction may be due to a spurious association with the more consistent presence of two parents or, because the family size variable measures the number of living children, other factors that predict child survival. Additionally, the positive effect of larger families may be related to positive consequences of siblings who divide family production burdens or assist one another with school related tasks; an analysis that examines the order and gender characteristics of the sibling structure could potentially shed light on these processes. Results also suggest that even small amounts of maternal education improve school achievement factors. Standardized height, a proxy for better early childhood nutrition, is positively associated with the outcomes, even when controlling for other factors. This effect is particularly large in the case of total completed grades. Working more is only associated with lower achievement in the case of total grades and not measured skills. Turning to parental attitudes, parents' belief that their children will achieve a high level of education is positively associated with total grades and literacy skills. This finding may reflect accurate perceptions of the child's ability. However, parental beliefs that they can help their children do better in school is not associated with higher achievement.

In summary, belonging to a marginalized social group predicts lower educational achievement in Peru, Vietnam, and India. However, these differences are reduced dramatically after controlling for the other factors in the models. In the full models, factors such more household wealth, smaller family size, higher levels of maternal education, better early childhood nutrition, spending less time in household production, and speaking the majority language first contribute to higher educational achievement. The direction of these factors is largely consistent with previous findings.

Regression Decomposition Results

I continue by describing the decomposition results to further develop an understanding of the observed academic achievement disparities. The previous section emphasized factors associated with academic achievement assuming that the processes within each country operated similarly for the two groups. The decomposition results examine how differing quantities and effects of the characteristics are associated with the differences. Consistent with the method's intended use, I limit the decompositions to findings where the initial results reveal a significant difference between the two groups: all three indicators for Peru and Vietnam and math and literacy skill indicators for India. I report these findings as proportions of the overall difference (Table 17).

Across all models, findings reveal that significant differences are attributable to the total explained component. This finding indicates that disparities are strongly associated with differing amounts of the independent variables included in the models. The portions of the differences attributable to the unexplained components are smaller, but in several cases are still significant, suggesting that differing processes may also contribute to the observed gap.

The most notable contributions to the explained component are the lower amounts of household wealth and maternal education. Differences in wealth alone explain approximately one third of the gap between SCs /STs and other Indian children on both math and literacy outcomes. Differing effects of wealth, indicated by the large unexplained wealth component suggest that a large portion of the difference is attributable to differing effects of wealth whereby

SCs/ STs have fewer returns to their wealth. Wealth differences also explain nearly two thirds of the total grades gap in Peru and a quarter of it in Vietnam. These findings suggest that economic resources may be a key factor in maintaining school enrollment for more-Indigenous Peruvian and ethnic minority Vietnamese children.

Less education among ethnic minority mothers in Vietnam significantly explain between a tenth and a quarter of the gap between ethnic minority and Kinh children's outcomes, and less education among more-Indigenous Peruvian mothers accounts for approximately a quarter of the difference in Peru. There are also several negative unexplained coefficients for maternal education, notably in the case of Peruvian literacy scores. Unexplained negative components indicate that the maternal education coefficients are stronger for the more-Indigenous group and hence the returns to maternal education in this case are actually higher.

Family size differences are not significantly associated with the observed gaps after controlling for other factors. However, larger family size among ethnic minorities in Vietnam may make a small contribution to differences in math and reading skills.

Standardized height differences, a measure for nutritional quality in early childhood contributes significantly to the observed differences, suggesting that nutritional disparities are an important factor in educational differences. In the case of Peru there is an interesting contrast when examining the portion of the achievement gaps attributable to height. The component explained by differences in height is much larger for the total number of grades completed than for actual measures of academic skill. The much stronger association with total number of grades completed than actual cognitive abilities, measured by academic skills, may suggest that taller children may simply enroll in school before their smaller peers, and its contribution to cognitive development may be overestimated. This merits examination in future work.

Differences in the time spent working between the two groups are significant in the case of math scores in Peru and marginally significant in the case of math scores in Vietnam. Increased involvement with family production strategies may be particularly detrimental to the learning process at this age even if children maintain school enrollment. Language differences are associated with the achievement gap within Peru such that more-Indigenous children who first learned a language other than Spanish are further disadvantaged in the current school system.

Overall, differences in family educational attitudes contribute a relatively small if any amount to achievement disparities. This contradicts assumptions that members of lower status groups place less value on education or expect fewer returns for investing in their children's education.

The explained effects of gender are nearly zero in all cases, which is expected given the similar distribution of gender between the two populations within each country. In several cases there is a large but insignificant negative portion of the gap attributable to the unexplained effects of being male. This finding hints that there are gender differences that disadvantage girls from disadvantaged social groups more than their male peers. This is not a conclusive finding but it merits further investigation.

In summary, among Peruvian children, educational disparities are explained largely by measurable differences between more-Indigenous and more-European households. The strength of these factors differs across indicators, but household wealth, maternal education, early childhood nutrition, and the amount of time spent working contribute to educational disadvantages among more-Indigenous children. A similar story is told among Indian children where differences in math and reading scores are attributable primarily to measurable differences between SC/ST households and other households, particularly differences in household wealth. Vietnam reveals a different story. Differences in wealth, maternal education, and early childhood nutrition are important; however, poorer educational achievement is also strongly associated with fewer returns to these endowments.

Chapter 4. Discussion

At the turn of the 21st century the United Nations (UN) established the Millennium Development Goals (MDGs), specific objectives for improving social and economic conditions in the poorest countries by 2015 (UN, 2000). While the MDGs target gender equity in primary education, comparable goals advancing ethnic minorities and other marginalized groups are absent (Kabeer, 2006).

Achieving equity for marginalized social groups has typically lagged behind gender equity in UN discourse. For example, the UN General Assembly adopted the *Declaration of the Elimination of Discrimination against Women* in 1967, whereas the *Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities* and the *Declaration on the Rights of Indigenous Peoples* were not adopted until 1992 and 2007 respectively.

As initially presented, the primary purpose of this thesis is to examine within-country disparities by social group. As a whole, the findings provide strong evidence that there is a need to develop large global efforts that will reduce educational disparities between social groups. The findings reveal a clear trend that educational disparities persist by social groups, particularly within India, Peru, and Vietnam. There is also limited evidence suggesting that ethnic disparities exist between the Amhara and Oromo in Ethiopia, although it is not possible to differentiate the consequences of individual ethnicity and the community level characteristics where these ethnic groups are concentrated. However, it is also possible that ethnicity is not driving inequality in Ethiopia.

Additionally, our findings also reveal information about the processes contributing to the observed disparities. Overall, differences in family level endowments, such as wealth and education are most strongly associated with academic achievement disparities. Furthermore in certain circumstances, family endowments are operating through early childhood nutritional deficiencies and the amount of time children dedicate to family livelihood strategies. There is no evidence suggesting that family educational attitudes contribute to achievement disparities.

Implications for Improving Educational Outcomes

The current findings suggest that targeting marginalized social groups is necessary to reduce within-country educational disparities. However, they reveal much less regarding how attempts to reduce disparities should proceed. Education's importance for predicting lifetime productivity will only increase in future decades as technological inputs increase and the global economy becomes more knowledge based (Psacharopoulos & Patrinos, 2004). If educational disparities are further neglected during earlier parts of the life course, economic and social disparities will likely persist.

Because a large portion of the achievement gap is due to differences in family endowments, targeting higher poverty families may effectively reduce educational disparities. Previously, Knodel and Jones (1996) demonstrated that educational disparities by relative income status within countries exceed the educational disparities found between boys and girls in these countries. They argued that universal primary education goals should shift their emphasis from gender and target high poverty children. However, gender equity remains the principal focus of primary education goals.

Opposition to targeting educational disparities by poverty status is largely logistical. Unarguably, it is a difficult task, especially in countries where accurate income documentation is rare. Additionally, as explained in the introduction, ethnic minorities experience fewer gains from anti-poverty policies. Therefore it is also important that these policies are specifically designed to benefit minority families.

An alternative solution to reduce educational disparities is to target policies and services to marginalized groups. This strategy seems like the most effective means for maximizing the effects of limited resources. A number of strategies, such as conditional cash transfers and school quality improvement are effective and could be modified to target marginalized groups.

However, there are drawbacks to targeting marginalized groups in this manner. Namely, there is concern that it would increase jealousy and conflict within already strained group dynamics. Knowing that one’s equally poor neighbor from a different ethnic group was receiving a subsidy would likely create animosity and conflict before it could alleviate disparities. For example, community members in post-ethnic conflict Nepal felt that targeting cash transfers only to marginalized group members would increase community divisiveness (Köhler, Calì, & Stirbu, 2009).

Geographically targeted interventions also have potential. Marginalized groups are often geographically concentrated. Targeting regions with a high proportion of minority group members has the potential to benefit marginalized groups and avoid the divisiveness that would likely occur when targeting individuals. Regional targeting appears to have been highly effective in Andhra Pradesh, India. Previously, I discussed the need to exclude several clusters from the analyses due to being targeted by multisystemic poverty alleviation strategies that included putting children in boarding schools. The efforts emphasized improved economic opportunities, access to health facilities, and particularly improving educational quality (Government of Andhra Pradesh, 2003; World Bank OPCS, 2005). Consequently, children from Scheduled Tribes and

Schedules Castes in these areas exceeded state-wide performance averages (Government of Andhra Pradesh, 2003).

However, a risk of geographically targeted efforts is that marginalized groups may not respond to efforts that do not specifically target their needs. There are two perspectives for addressing ethnic inequalities. One assumes that the majority model will improve the outcomes of the minority group if it is applied to the minority group, and the other deems it necessary to work within a unique framework, driven by the needs of the disadvantaged group (van de Walle & Gunewardena, 2001).

Furthermore, when assuming a majority model within geographically targeted poverty reduction strategies, residents from the majority ethnic group may be better equipped to capitalize on newly available resources. This was the case in Vietnam where government policies emphasized poverty alleviation in the rural mountain regions, home to ethnic minority groups. However, the Kinh in these areas benefitted more than the ethnic minorities, because the poverty alleviation model assumed ethnic minority income earning and service use strategies paralleled the Kinh's (van de Walle & Gunewardena, 2001). Therefore, efforts should consider the specific needs and behaviors of the targeted group.

Implications for Studying Educational Disparities

A consistent finding in these analyses is that underlying disparities are masked by convergence in the completed number of grades. The analyses predict both completed grades and ability. The factors that predict achievement and differences vary in these different domains. This is not surprising given that schools attended by children from marginalized social groups often receive fewer investments. Additionally, teachers may have lower expectations of certain children. Using skill indicators allows for a more nuanced understanding of academic achievement. Incorporating extensive skill indicators into large household surveys may not be feasible, but studies that aim to understand educational processes should use skill indicators

Conclusion

This thesis makes several unique contributions. It addresses within-country educational achievement disparities from an internationally comparative perspective, across four distinctly different settings. This design grants the opportunity to draw broader conclusions about the nature of achievement disparities. Additionally, it uses math and literacy abilities in addition to the child's grade in school. Given that the schools resources are unequally distributed, it is important to use indicators that evaluate whether learning actually occurs.

However, this thesis also has several limitations. One concern is testing bias; this is the possibility that one group earns higher scores due to test features and not intellectual ability. Despite extensive efforts in the survey design and later item elimination due to differential item functioning, the possibility cannot be discounted. A testing bias that favors the majority group would overestimate the observed skill differences. A further weakness is that it is not possible to measure change in the same indicators over time. Examining growth in particular academic domains would provide a clearer understanding of how academic skill disparities develop throughout childhood and should be incorporated in future research. Finally, an additional limitation is the sample size, which is further handicapped by the sampling design.

In conclusion, I examine disparities in academic skill development by social group in Ethiopia, India, Peru, and Vietnam using data from the Young Lives Study, a longitudinal study of childhood poverty. I examine characteristics that predict academic achievement on three domains: math, literacy, and total completed grades, within each country. Then, I decompose the differences within each country to highlight the particular factors contributing to observed disparities. I conclude that educational disparities persist by social group, and I discuss several possible strategies for alleviating these disparities.

References

Aikman, S., & Pridmore, P. (2001). Multigrade schooling in remote areas of Vietnam. *International Journal of Educational Development*, *21*(6), 521-536.

Anderson, J. (2007). Urban poverty reborn. *Journal of Developing Societies*, *23*(1-2), 221 -241.

Anh, T. S., Knodel, J., Lam, D., & Friedman, J. (1998). Family Size and Children's Education in Vietnam. *Demography*, *35*(1), 57-70.

Baulch, B., Chuyen, T. T. K., Haughton, D., & Haughton, J. (2004). Ethnic minority development in Vietnam: A socioeconomic perspective. In P. Glewwe, N. Agrawal, & D. Dollar (Eds.), *Economic growth, poverty, and household welfare in Vietnam*. The International Bank for Reconstruction and Development, The World Bank, Washington, DC.

Behrman, J. R., & Knowles, J. C. (1999). Household income and child schooling in Vietnam. *The World Bank Economic Review*, *13*(2), 211-256.

Ben-Arieh, A., Kaufman, N. H., Andrews, A. B., George, R., Lee, B. J., & Aber, J. L. (2001). *Measuring and Monitoring Children's Well-Being*. Social Indicators Research Series. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, *8*(4), 436-455.

Borghi, E., Onis, M. D., Garza, C., Broeck, J. V. D., Frongillo, E. A., Grummer-Strawn, L., Buuren, S. V., et al. (2006). Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine*, *25*(2), 247-265.

Brockerhoff, M., & Hewitt, P. (2000). Inequality of child mortality among ethnic groups in Sub-Saharan Africa. *Bulletin of the World Health Organization*, *70*, 30-41.

Buchmann, C., & Hannum, E. (2001). Education and stratification in developing countries: A review of theories and research. *Annual Review of Sociology*, *27*, 77-102.

Burnett, N. (2008). Education for All: An imperative for reducing poverty. *New York Academy Sciences Annals*, *1136*, 269-275.

Chanana, K. (1993). Accessing higher education: The dilemma of schooling women, minorities, Scheduled Castes and Scheduled Tribes in contemporary India. *Higher Education*, *26*(1), 69-92.

Charasse-Pouélé, C., & Fournier, M. (2006). Health disparities between racial groups in South Africa: a decomposition analysis. *Social Science & Medicine*, *62*(11), 2897-2914.

Cornia, G. A. (2003). The Impact of liberalisation and globalisation on within-country income Inequality. *CESifo Economic Studies*, *49*(4), 581-616.

Cueto, S., Leon, J., Guerrero, G., & Muñoz Ismael. (2009). Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives. Young Lives Technical Report No. 15. Retrieved from http://www.younglives.org.uk/pdf/ publication-section-pdfs/technical-notes-pdfs/YL-TN15-Cueto.pdf

Deshpande, A. (2000). Does caste still define disparity? A look at inequality in Kerala, India. *The American Economic Review*, *90*(2), 322-325.

Dreze, J., & Loh, J. (1995). Literacy in India and China. *Economic and Political Weekly*, *30*(45), 2868-2878.

Government of Andhra Pradesh. (2003). Strategy Paper on Welfare of Scheduled Caste and Scheduled Tribe Population. Retrieved from http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN015043.pdf

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60-70.

Hall, G., & Patrinos, H. A. (2005). Indigenous peoples, poverty, and human development in Latin America: 1994-2004. The World Bank. Retrieved September 21, 2009, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2005/07/20/000112742_20050720171830/Rendered/PDF/330200ENGLISH01Poverty1Exec1Summ1en.pdf.

Huttly, S., Jones, N. & Boyden, J., (2009) “Young Lives: an International Study of Childhood Poverty: Round 1, 2002 & Round 2, 2006 [computer file].” Colchester, Essex: UK Data Archive [distributor], August 2009. SN: 5307.

Ilon, L., & Moock, P. (2001). School attributes, household characteristics, and demand for schooling: A case study of rural Peru. *International Review of Education*, *37*(4), 429-451.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Kabeer, N. (2006, March). Social exclusion and the MDGs: The challenge of durable inequalities' in the Asian context. Institute of Development Studies and Overseas Development Institute. Retrieved September 21, 2009, from http://www.eldis.org/vfile/upload/1/document/0708/DOC21178.pdf.

Knodel, J., & Jones, G. (1996). Post-Cairo population policy: Does promoting girls' education miss the mark? *Population and Development Review*, 22(4), 683-702.

Köhler, G., Calì, M., & Stirbu, M. (2009). Rethinking poverty and social exclusion: Responses in post-conflict Nepal: Child-sensitive social protection. *Children, Youth, and Environments*, 19(2), 229-249.

Kumra, I. (2008). An asessment of the Young Lives sampling approach in India. Young Lives Technical Report No. 2. Retrieved from http://www.younglives.org.uk/files/technical-notes/an-assessment-of-the-young-lives-sampling-approach-in-andhra-pradesh-india

Lock, K., & Gibson, J. (2008). Explaining Maori under-achievement in standardized reading tests: the role of social and individual characteristics. *Kotuitui: NZ Journal of Social Sciences Online*, *3*, 1-13.

McEwan, P. J. (2004). The indigenous test score gap in Bolivia and Chile. *Economic Development and Cultural Change*, *53*(1), 157-190.

McNab, C. (1990). Language policy and language practice: Implementing multilingual literacy education in Ethiopia. *African Studies Review*, *33*(3), 65-82.

Mengisteab, K. (2001). Ethiopia's ethnic-based federalism: 10 years after. *African Issues*, *29*(1/2), 20-25.

Myers, S. L., Kim, H., & Mandala, C. (2004). The effect of school poverty on racial gaps in test scores: The case of the Minnesota Basic Standards Tests. *The Journal of Negro Education*, *73*(1), 81-98.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *The Journal of Human Resources*, *23*(3), 279-295.

Ñopo, H., Saavedra, J., & Torero, M. (2007). Ethnicity and earnings in a mixed-race labor market. *Economic Development and Cultural Change*, *55*(4), 709-734.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693-709.

Outes-Leon, I., & Sanchez, A. (2008). An assessment of the Young Lives sampling approach in Ethiopia. Young Lives Technical Report No. 1. Retrieved from http://www.younglives.org.uk/files/technical-notes/an-assessment-of-the-young-lives-sampling-approach-in-ethiopia

Patrinos, H. A., & Psacharopoulos, G. (1996). Socioeconomic and ethnic determinants of age-grade distortion in Bolivian and Guatemalan primary schools. *International Journal of Educational Development*, *16*(1), 3-14.

Psacharopoulos, G. (1997). Child labor versus educational attainment Some evidence from Latin America. *Journal of Population Economics*, *10*(4), 377-386.

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, *12*(2), 111-134.

Ramaiah, A. (1992). Identifying Other Backward Classes. *Economic and Political Weekly*, *27*(23), 1203-1207.

Ranis, G., Stewart, F., & Ramirez, A. (2000). Economic growth and human development. *World Development*, *28*(2), 197-219.

Rao, N., Cheng, K., & Narain, K. (2003). Primary schooling in China and India: Understanding how socio-contextual factors moderate the role of the state. *International Review of Education*, *49*(1/2), 153-176.

Rose, P., & Al Samarrai, S. (2001). Household constraints on schooling by gender: Empirical evidence from Ethiopia. *Comparative Education Review*, *45*(1), 36-63.

Swinkels, R., & Turk, C. (2006). Explaining ethnic minority poverty in Vietnam: a summary of recent trends and current challenges. Background paper for the Committee for Ethnic Minorities/Ministry of Planning and Investment. The World Bank, Hanoi.

Tronvoll, K. (2000). Ethiopia: A new start? Minority Rights Group International. Retrieved from http://www.addisvoice.com/resources/Ethiopia-new-start.pdf

United Nations. (2000). United Nations Millennium Declaration. Resolution of the United Nations General Assembly, A/RES/55/2.

Wagaw, T. G. (1999). Conflict of ethnic identity and the language of education policy in contemporary Ethiopia. *Northeast African Studies*, *6*(3), 75-88.

van de Walle, D., & Gunewardena, D. (2001). Sources of ethnic inequality in Viet Nam. *Journal of Development Economics*, *65*(1), 177-207.

Weisbrot, M., Baker, D., & Rosnick, D. (2006). The scorecard on development: 25 years of diminished progress. *International Journal of Health Services*, *36*(2), 211-234.

World Bank, Operations Policy and Country Services (OPCS). (2005). Status of Projects in Execution - FY05 SOPE, Region: South Asia, Country: India. Retrieved from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/09/15/000333037_20090915024840/Rendered/PDF/498220AR0India0050Box341974B01PUBLIC1.pdf

Zoninsein, J. (2001). The economic case for combating racial and ethnic exclusion in Latin American and Caribbean countries. Washington, DC: Inter-American Development

Bank. Retrieved September 21, 2009, from http://www5.iadb.org/sds/doc/soc-ZoninsteinJonasi.pdf.

APPENDIX: TABLES

Table 1. Peru: Descriptive Statistics

	More-European n = 388		More-Indigenous n = 272		
Dependent Variables	Mean	Lin. SE	Mean	Lin. SE	
Math score- age 12 (-3 to 3)	0.67	0.12	0.07	0.20	*
Total grades- age 12	6.07	0.07	5.76	0.14	*
Literacy- age 8[1] (0 to 1)	0.84	0.02	0.71	0.04	**
Independent Variables					
Male	0.53	0.02	0.54	0.03	
Wealth Index- age 12	0.60	0.02	0.46	0.05	**
Wealth Index- age 8*	0.56	0.03	0.45	0.05	*
Children in family = 1 or 2	0.45	0.04	0.29	0.05	*
Children in family = 3	0.25	0.02	0.21	0.02	
Children in family = 4 or more	0.30	0.03	0.50	0.06	*
Maternal Ed- Incomplete primary ed.	0.15	0.02	0.55	0.08	***
Complete primary ed.	0.40	0.04	0.28	0.04	*
Complete secondary ed.	0.44	0.05	0.18	0.05	***
Standardized height	-1.17	0.06	-1.61	0.12	*
Works	4.08	0.19	5.53	0.25	***
Spanish First	0.00	0.00	0.71	0.10	*
School Parental School Efficacy	0.74	0.03	0.69	0.04	
Region					
Rural	0.16	0.07	0.38	0.12	
Coast	0.53	0.12	0.23	0.07	*
Sierra	0.30	0.11	0.64	0.13	*
Jungle	0.17	0.10	0.13	0.12	

[1]Variables specific to round 1, More-European n = 378, More Indigenous n = 256

† p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 2. Vietnam: Descriptive Statistics

	Kinh		Ethnic Minority		
	n = 821		n = 103		
Dependent Variables	Mean	Lin. SE	Mean	Lin. SE	
Math score- age 12 (-3 to 3)	1.30	0.09	-0.16	0.31	***
Total grades- age 12	6.67	0.03	5.93	0.24	**
Literacy- age 8 (0 to 1)	0.93	0.01	0.52	0.16	*
Independent Variables					
Male	0.50	0.02	0.51	0.07	
Wealth Index- age 12	0.49	0.01	0.27	0.07	**
Children in family = 1 or 2	0.61	0.02	0.19	0.06	***
Children in family = 3	0.23	0.02	0.33	0.04	*
Children in family = 4 or more	0.16	0.01	0.48	0.09	***
Maternal Ed- Incomplete primary ed.	0.18	0.02	0.82	0.12	***
Complete primary ed.	0.34	0.03	0.14	0.10	*
Complete low secondary ed.	0.48	0.04	0.05	0.03	***
Standardized height	-1.29	0.05	-2.25	0.06	***
Time working	2.02	0.06	3.58	0.38	***
Parent believes school is essential	0.92	0.08	0.69	0.10	*
Parent believes child can go far in school	0.76	0.03	0.80	0.06	
Region					
Rural	0.77	0.01	1.00	0.00	***
Northern Uplands	0.11	0.03	0.82	0.16	**
Red River Delta	0.22	0.01	0.00	0.00	***
Central Coast	0.44	0.02	0.18	0.16	**
Mekong River Delta	0.22	0.01	0.00	0.00	***

† $p \leq .10$, *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

Table 3. Andhra Pradesh, India: Descriptive Statistics

	Other n = 579		Scheduled n = 155		
Dependent Variables	Mean	Lin. SE	Mean	Lin. SE	
Math score- age 12 (-3 to 3)	0.71	0.17	0.19	0.17	*
Total grades- age 12	6.64	0.08	6.45	0.12	
Literacy- age 8* (0 to 1)	0.71	0.03	0.58	0.04	***
Independent Variables					
Male	0.50	0.02	0.52	0.04	***
Wealth Index- age 12	0.39	0.04	0.27	0.02	***
Wealth Index- age 8*	0.39	0.04	0.27	0.03	***
Children in family = 1 or 2	0.47	0.04	0.36	0.05	†
Children in family = 3	0.35	0.03	0.39	0.04	
Children in family = 4 or more	0.18	0.03	0.25	0.04	
Maternal Ed- Incomplete primary ed.	0.58	0.07	0.82	0.05	***
Complete primary ed.	0.25	0.03	0.08	0.03	***
Complete secondary ed.	0.14	0.05	0.04	0.02	*
Standardized height	-1.47	0.08	-1.69	0.10	†
Works	1.60	0.17	1.90	0.20	
Parent believes child will go far in school	0.88	0.02	0.84	0.03	
Parental school efficacy	0.78	0.03	0.68	0.07	
Region					
Rural	0.66	0.12	0.81	0.10	*
Coast	0.35	0.07	0.19	0.07	*
Rayalaseema	0.29	0.06	0.38	0.08	
Telangana	0.36	0.06	0.43	0.09	

[1]Variables specific to model predicting literacy at age 8; Other n = 585, SCs/STs n = 168

† p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 4. Ethiopia: Descriptive Statistics

	Amhara		Oromo		
	n = 249		n = 196		
Dependent Variables	Mean	Lin. S.E	Mean	Lin. S.E	
Math score- age 12 (-3 to 3)	-0.06	0.22	-0.49	0.25	
Total grades- age 12	4.66	0.28	4.18	0.43	
Literacy- age 12* (0 to 1)	0.85	0.03	0.72	0.04	**
Independent Variables					
Male	0.49	0.03	0.55	0.03	
Wealth Index- age 12	0.16	0.03	0.18	0.03	
Children in family = 1 or 2	0.26	0.06	0.13	0.05	
Children in family = 3	0.17	0.03	0.18	0.04	
Children in family = 4 or more	0.56	0.05	0.68	0.06	
Maternal Ed- No schooling	0.44	0.10	0.33	0.06	
Incomplete primary ed.	0.38	0.06	0.52	0.05	†
Complete primary ed.	0.17	0.06	0.14	0.05	
Standardized height	-1.54	0.12	-1.19	0.16	†
Works	4.10	0.46	4.54	0.37	
Parental school efficacy	0.55	0.05	0.52	0.07	
Parent believes child will go far in school	0.96	0.01	0.91	0.03	
Region					
Rural	0.55	0.18	0.58	0.21	
Addis Ababa	0.19	0.12	0.19	0.15	
Amhara State	0.68	0.15	0.01	0.01	***
Oromia State	0.09	0.06	0.77	0.16	***
SNNP	0.04	0.04	0.04	0.04	

† $p \leq .10$, *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

Table 5. Peru: Regression models predicting grades in school at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7			Model 8		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds language			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
More Indigenous	-0.32	0.14	*	-0.26	0.08	**	-0.26	0.08	**	-0.13	0.07	†	-0.06	0.08		-0.05	0.08		-0.04	0.08		-0.04	0.08	
Gender (male = 1)							-0.16	0.08	†	-0.17	0.08	*	-0.17	0.08	*	-0.16	0.08	†	-0.17	0.08	*	-0.17	0.08	*
Wealth Index										2.04	0.29	**	1.77	0.26	**	1.54	0.24	**	1.48	0.25	***	1.48	0.25	***
Children in family (ref = 1 or 2)																								
3 children													-0.21	0.12	†	-0.21	0.11	*	-0.22	0.12	†	-0.21	0.11	†
4 or more children													-0.20	0.11	†	-0.17	0.11		-0.16	0.11		-0.16	0.10	
Maternal Ed (ref = no primary)																								
Complete primary ed.													0.11	0.09		0.12	0.09		0.07	0.10		0.08	0.10	
Complete secondary ed.													0.18	0.13		0.16	0.13		0.12	0.13		0.13	0.13	
Standardized height																0.15	0.05	**	0.15	0.05	**	0.15	0.05	**
Time working																-0.03	0.02		-0.03	0.02	*	-0.03	0.02	
Spanish spoken first																			0.43	0.17	*	0.43	0.17	*
Parental school efficacy																						-0.03	0.10	
R^2 adjusted	0.02			0.14			0.15			0.23			0.24			0.26			0.26			0.26		
Sample size	660			660			660			660			660			660			660			660		

Models 2-8 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 6. Peru: Regression models predicting math ability at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7			Model 8		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds language			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
More Indigenous	-0.60	0.22	*	-0.38	0.13	**	-0.38	0.13	**	-0.32	0.12	*	-0.22	0.12	†	-0.20	0.12	†	-0.19	0.11	†	-0.18	0.11	†
Gender (male = 1)							0.05	0.07		0.05	0.07		0.05	0.07		0.04	0.07		0.03	0.07		0.01	0.07	
Wealth Index										0.92	0.34	*	0.45	0.43		0.36	0.44		0.29	0.41		0.29	0.40	
Children in family (ref = 1 or 2)																								
3 children													-0.36	0.14	*	-0.37	0.14	*	-0.37	0.14	*	-0.39	0.15	*
4 or more children													-0.26	0.10	*	-0.27	0.09	**	-0.27	0.09	**	-0.29	0.09	**
Maternal Ed (ref = no primary)																								
Complete primary ed.													0.12	0.14		0.11	0.14		0.07	0.15		0.06	0.15	
Complete secondary ed.													0.33	0.13	*	0.30	0.13	*	0.26	0.13	†	0.26	0.12	*
Standardized height																-0.01	0.05		-0.01	0.05		0.00	0.05	
Time working																-0.06	0.02	*	-0.06	0.02	*	-0.06	0.02	*
Spanish spoken first																			0.42	0.20	*	0.43	0.20	*
Parental school efficacy																						0.20	0.11	†
R^2 adjusted	0.05			0.21			0.21			0.22			0.23			0.24			0.24			0.25		
Sample size	660			660			660			660			660			660			660			600		

Models 2-8 include dummy variables for N-1 clusters; † $p \le .10$, *$p \le .05$, **$p \le .01$, ***$p \le .001$

Table 7. Peru: Regression models predicting literacy at age 8

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds language		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
More Indigenous	-0.13	0.03	**	-0.07	0.02	*	-0.07	0.02	*	-0.04	0.03		-0.01	0.03		-0.01	0.03		0	0.03	
Gender (male = 1)							-0.02	0.03		-0.02	0.02		-0.02	0.02		-0.02	0.03		-0.02	0.02	
Wealth Index										0.32	0.06	**	0.21	0.07	**	0.2	0.08	*	0.18	0.07	*
Children in family (ref = 1 or 2)																					
3 children													-0.04	0.03		-0.04	0.03		-0.04	0.03	
4 or more children													-0.05	0.02	*	-0.04	0.02	†	-0.04	0.02	†
Maternal Ed (ref = no primary)																					
Complete primary ed.													0.09	0.04	*	0.09	0.04	*	0.08	0.04	*
Complete secondary ed.													0.12	0.04	**	0.11	0.03	**	0.11	0.04	*
Standardized height																0.02	0.01		0.02	0.01	
Spanish spoken first																			0.14	0.07	†
R^2 adjusted	0.05			0.18			0.18			0.21			0.23			0.24			0.25		
Sample size	634			634			634			634			634			634			634		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, p ≤ .001

Table 8. Vietnam: Regression models predicting total grades in school at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Ethnic Minority	-0.74	0.24	***	-0.53	0.21	*	-0.53	0.21	*	-0.42	0.19	*	-0.33	0.18	†	-0.24	0.16		-0.25	0.16	
Gender (male = 1)							0.04	0.06		0.05	0.05		0.06	0.06		0.06	0.06		0.07	0.06	
Wealth Index										1.14	0.25	*	1.01	0.26	***	0.85	0.24	**	0.81	0.04	**
Children in family (ref = 1 or 2)																					
3 children													0.11	0.04	*	0.14	0.04	**	0.14	0.04	**
4 or more children													-0.09	0.08		-0.05	0.08		-0.06	0.08	
Maternal Ed (ref = no primary)																					
Complete primary ed.													0.07	0.06		0.08	0.06		0.08	0.06	
Complete low secondary ed.													0.16	0.05	**	0.15	0.05	**	0.14	0.05	*
Standardized height																0.14	0.04	**	0.14	0.04	**
Time working																-0.02	0.02		-0.02	0.03	
Parent believes school is essential																			-0.02	0.07	
Parent believes child can go far in school																			0.13	0.06	†
R^2 adjusted	0.10			0.20			0.20			0.23			0.24			0.26			0.27		
Sample size	938			938			938			938			938			938			938		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 9. Vietnam: Regression models predicting math ability at age 12

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	Ethnic indicator		Cluster fixed effects		Adds gender		Adds wealth		Adds children & edu		Adds std. ht. and work		Adds parent sch factors	
	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE
Ethnic Minority	-1.46	0.31 ***	-1.18	0.27 ***	-1.18	0.27 ***	-0.97	0.29 **	-0.55	0.24 *	-0.36	0.25	-0.4	0.24
Gender (male = 1)					-0.12	0.09	-0.09	0.08	-0.12	0.09	-0.14	0.1	-0.11	0.1
Wealth Index							2.17	0.4 ***	1.47	0.4 **	1.16	0.38 **	0.95	0.38 *
Children in family (ref = 1 or 2)														
3 children									-0.06	0.12	-0.02	0.12	-0.02	0.11
4 or more children									-0.42	0.16 *	-0.36	0.16 *	-0.37	0.17 *
Maternal Ed (ref = no primary)														
Complete primary ed.									0.38	0.13 *	0.41	0.13 **	0.39	0.12 **
Complete low secondary ed.									0.76	0.14 ***	0.75	0.14 ***	0.68	0.13 ***
Standardized height											0.19	0.05 **	0.19	0.05 **
Time working											-0.09	0.04 *	-0.07	0.04 †
Parent believes school is essential													-0.09	0.17
Parent believes child can go far in school													0.59	0.11 ***
R^2 adjusted	0.08		0.16		0.16		0.18		0.21		0.23		0.25	
Sample size	938		938		938		938		938		938		938	

Models 2-7 include dummy variables for N-1 clusters; † $p \leq .10$, *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

Table 10. Vietnam: Regression models predicting literacy at age 8

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Ethnic indicator		Cluster fixed effects		Adds gender		Adds wealth		Adds children & edu		Adds std. ht. and work	
	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE
Ethnic Minority	-0.41	0.16 *	-0.23	0.05 ***	-0.23	0.05 ***	-0.2	0.05 **	-0.15	0.05 **	-0.13	0.05 *
Gender (male = 1)					-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01
Wealth Index							0.26	0.07 **	0.18	0.07 *	0.15	0.07 *
Children in family (ref = 1 or 2)												
3 children									0.01	0.01	0.02	0.01
4 or more children									-0.07	0.03 *	-0.06	0.03 *
Maternal Ed (ref = no primary)												
Complete primary ed.									0.03	0.02	0.04	0.02
Complete low secondary ed.									0.08	0.03 **	0.08	0.02 **
Standardized height											0.04	0.01 **
R^2 adjusted	0.27		0.41		0.41		0.43		0.45		0.47	
Sample size	938		938		938		938		938		938	

Models 2-6 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 11. Andhra Pradesh, India: Regression models predicting total grades in school at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Scheduled Caste or Tribe	-0.19	0.12		-0.05	0.09		-0.05	0.09		-0.05	0.08		0.07	0.07		0.08	0.08		0.09	0.09	
Gender (male = 1)							0.01	0.09		0.00	0.09		-0.05	0.08		-0.15	0.07		-0.11	0.07	
Wealth Index										1.26	0.29	***	1.08	0.28	***	0.54	0.23	*	0.45	0.24	†
Children in family (ref = 1 or 2)																					
3 children													-0.08	0.07		-0.05	0.07		-0.04	0.07	
4 or more children													-0.3	0.17	†	-0.25	0.15	†	-0.22	0.14	
Maternal Ed (ref = no primary)																					
Complete primary ed.													0.13	0.1		0.03	0.11		0.01	0.11	
Complete secondary ed.													0.12	0.11		-0.01	0.1		-0.03	0.1	
Standardized height																0.18	0.03	***	0.18	0.04	***
Time working																-0.19	0.03	***	-0.14	0.03	***
Parent believes child will go far in school																			0.49	0.18	*
Parental school efficacy																			0.20	0.08	*
R^2 adjusted	0.01			0.16			0.16			0.18			0.19			0.31			0.33		
Sample size	734			734			734			734			734			734			734		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 12. Andhra Pradesh, India: Regression models predicting math ability at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Scheduled Caste or Tribe	-0.52	0.21	*	-0.32	0.19	†	-0.34	0.19	†	-0.13	0.16		-0.07	0.16		-0.08	0.15		-0.06	0.16	
Gender (male = 1)							0.41	0.14	**	0.38	0.13	**	0.38	0.14	**	0.25	0.14	†	0.34	0.14	*
Wealth Index										2.86	0.48	***	2.30	0.49	***	1.88	0.52	**	1.72	0.49	**
Children in family (ref = 1 or 2)																					
3 children													-0.07	0.11		-0.05	0.10		0.00	0.09	
4 or more children													-0.48	0.21	*	-0.44	0.19	*	-0.38	0.17	*
Maternal Ed (ref = no primary)																					
Complete primary ed.													0.28	0.15	†	0.21	0.15		0.19	0.15	
Complete secondary ed.													0.56	0.27	*	0.52	0.25	*	0.51	0.24	*
Standardized height																0.02	0.08		0.02	0.08	
Time working																-0.18	0.03	***	-0.08	0.04	*
Parent believes child will go far in school																			0.52	0.17	**
Parental school efficacy																			0.76	0.11	**
R^2 adjusted	0.02			0.13			0.15			0.19			0.22			0.26			0.29		
Sample size	734			734			734			734			734			734			734		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 13. Andhra Pradesh, India: Regression models predicting literacy at age 8

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Scheduled Caste or Tribe	-0.13	0.03	***	-0.10	0.04	*	-0.10	0.04	*	-0.07	0.04	†	-0.05	0.04		-0.06	0.04	
Gender (male = 1)							0.04	0.02	†	0.04	0.02	†	0.04	0.02	†	0.04	0.02	†
Wealth Index										0.44	0.11	**	0.33	0.11	**	0.32	0.11	**
Children in family (ref = 1 or 2)																		
3 children													-0.01	0.03		-0.01	0.03	
4 or more children													-0.06	0.04		-0.06	0.04	
Maternal Ed (ref = no primary)																		
Complete primary ed.													0.06	0.03	†	0.05	0.03	†
Complete secondary ed.													0.14	0.06	*	0.13	0.06	*
Standardized height																0.02	0.01	
R^2 adjusted	0.03			0.13			0.13			0.17			0.18			0.19		
Sample size	753			753			753			753			753			753		

Models 2-6 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 14. Ethiopia: Regression models predicting total grades at age 12

	Model 1		Model 2		Model 3		Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator		Cluster fixed effects		Adds gender		Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE	b	Lin SE	b	Lin SE	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Ethnic Group (1 = Oromo)	-0.47	0.45	0.06	0.24	0.06	0.25	0.11	0.23		0.14	0.21		0.17	0.27		0.21	0.28	
Gender (male = 1)					-0.01	0.14	0.01	0.13		0.00	0.13		-0.08	0.14		-0.06	0.15	
Wealth Index							2.97	0.49	***	2.46	0.63	***	2.17	0.41	***	2.30	0.43	**
Children in family (ref = 1 or 2)																		
3 children										0.14	0.24		0.21	0.21		0.22	0.21	
4 or more children										0.14	0.22		0.12	0.20		0.10	0.20	
Maternal Ed- (ref = no schooling)																		
Incomplete primary ed.										0.35	0.15	*	0.37	0.12	**	0.38	0.12	**
Complete primary ed.										0.62	0.25	*	0.53	0.29	†	0.51	0.28	†
Standardized height													0.44	0.05	***	0.44	0.05	**
Time working													-0.14	0.03	***	-0.13	0.03	**
Parent believes child will go far in school																0.27	0.09	**
Parental school efficacy																0.31	0.22	
R^2 adjusted	0.02		0.31		0.31		0.33			0.34			0.46			0.47		
Sample size	445		445		445		445			445			445			445		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 15. Ethiopia: Regression models predicting math ability at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Ethnic Group (1 = Oromo)	-0.32	0.29		-0.15	0.19		-0.19	0.19		-0.15	0.18		-0.08	0.18		-0.07	0.19		-0.06	0.18	
Gender (male = 1)							0.30	0.13	*	0.32	0.13	*	0.29	0.13	*	0.25	0.13	†	0.25	0.14	†
Wealth Index										2.93	0.93	**	2.50	0.97	*	2.38	0.93	*	2.39	0.93	*
Children in family (ref = 1 or 2)																					
3 children													-0.09	0.21		-0.07	0.21		-0.07	0.21	
4 or more children													0.19	0.17		0.18	0.16		0.18	0.17	
Maternal Ed- (ref = no schooling)																					
Incomplete primary ed.													0.38	0.07	***	0.38	0.09	***	0.39	0.09	***
Complete primary ed.													0.63	0.16	***	0.60	0.16	***	0.61	0.15	***
Standardized height																0.11	0.05	*	0.11	0.05	*
Time working																-0.06	0.05		-0.05	0.05	
Parent believes child will go far in school																			0.12	0.13	
Parental school efficacy																			-0.06	0.19	
R^2 adjusted	0.01			0.21			0.22			0.24			0.27			0.28			0.28		
Sample size	445			445			445			445			445			445			445		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 16. Ethiopia: Regression models predicting literacy at age 12

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6			Model 7		
	Ethnic indicator			Cluster fixed effects			Adds gender			Adds wealth			Adds children & edu			Adds std. ht. and work			Adds parent sch factors		
	b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE		b	Lin SE	
Ethnic Group (1 = Oromo)	-0.13	0.04	**	-0.01	0.03		-0.01	0.03		-0.01	0.03		0.00	0.03		0.00	0.04		0.01	0.04	
Gender (male = 1)							0.01	0.03		0.01	0.03		0.02	0.02		0.01	0.03		0.02	0.03	
Wealth Index										0.28	0.18		0.31	0.20		0.29	0.20		0.32	0.19	
Children in family (ref = 1 or 2)																					
3 children													0.03	0.03		0.03	0.03		0.03	0.02	
4 or more children													0.00	0.02		0.00	0.02		-0.01	0.02	
Maternal Ed- (ref = no schooling)																					
Incomplete primary ed.													0.07	0.03	*	0.07	0.02	*	0.07	0.02	**
Complete primary ed.													0.02	0.04		0.02	0.05		0.01	0.04	
Standardized height																0.02	0.01	†	0.02	0.01	†
Time working																-0.01	0.01		-0.01	0.01	
Parent believes child will go far in school																			0.05	0.02	*
Parental school efficacy																			0.10	0.07	
R^2 adjusted	0.06			0.18			0.18			0.19			0.20			0.21			0.23		
Sample size	445			445			445			445			445			445			445		

Models 2-7 include dummy variables for N-1 clusters; † p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001

Table 17. Decomposition of Achievement Differences

			Total		Wealth Index		Family Size		Maternal		Standardized		Hours Working		Language		Parental School		Gender		Cluster Fixed		Intercept	
India	Math Score	Explained	88.1**	(28.2)	38.5*	(14.5)	4.9	(4.4)	15.9 †	(8.8)	0.9	(3.3)	4.4	(3.7)			18.0	(10.8)	-1.3	(3.3)	6.9	(19.9)		
	Gap = .52	Unexplained	11.9	(30.0)	11.6	(64.8)	-69.7†	(34.0)	-15.4	(12.7)	17.4	(34.1)	-29.8	(28.4)			-22.2	(87.8)	-20.5	(28.2)	133.6	(81.0)	7.0	(1.3)
	Literacy Score	Explained	54.1**	(13.8)	28.0*	(11.4)	2.4	(2.7)	17.5 †	(9.3)	2.9	(2.9)							-0.2	(1.5)	3.4	(9.3)		
	Gap = .13	Unexplained	45.9	(27.5)	144.0**	(46.5)	-21.0	(27.2)	4.2	(7.1)	43.9	(40.8)							6.3	(23.6)	218.1**	(69.0)	-349.6	(73.4)
Peru	Total Grades	Explained	87.3*	(41.5)	63.5*	(26.1)	7.1	(6.8)	13.7	(14.7)	20.3*	(9.5)	14.9	(11.1)	39.2†	(21.1)	-0.7	(1.8)	0.7	(2.0)	-71.3*	(33.8)		
		Unexplained	12.7	(19.3)	1.8	(101.3)	37.5	(36.8)	6.3	(43.7)	-44.2	(31.8)	37.8	(70.6)	-125.1*	(56.6)	31.9	(38.9)	0.6	(35.0)	49.1	(37.6)	17.1	(128.3
	Math Score	Explained	70.3*	(34.0)	6.5	(9.4)	6.7	(4.1)	12.9	(8.5)	-0.2	(4.0)	13.9*	(6.4)	20.5	(12.3)	1.8	(1.7)	0.0	(0.0)	8.3	(26.5)		
		Unexplained	29.7*	(13.8)	56.2	(64.8)	0.8	(19.3)	-11.2	(22.9)	26.0	(21.2)	3.8	(37.6)	-59.5†	(29.3)	38.3	(27.6)	-30.4†	(17.3)	43.1	(26.6)	-37.4	(93.9)
	Literacy Score	Explained	96.0**	(26.0)	12.6	(8.2)	2.6	(3.4)	25.6*	(11.2)	6.7	(5.2)			27.7	(18.1)			0.2	(0.1)	20.6	(22.8)		
		Unexplained	4.0	(15.1)	-15.3	(56.6)	-17.0	(20.0)	-65.1*	(20.2)	-10.0	(28.9)			-75.6	(59.0)			-7.0	(12.4)	-12.7	(16.8)	206.7	(117.8
Vietnam	Total Grades	Explained	65.6*	(26.8)	24.4*	(10.2)	1.0	(4.0)	10.2*	(4.7)	17.7**	(5.7)	3.7	(4.8)			-1.0	(2.6)	-0.3	(0.1)	9.9	(25.8)		
	Gap = .74	Unexplained	34.4	(16.1)	-25.8	(75.2)	-5.2	(26.2)	-7.3	(9.6)	72.6	(52.7)	17.3	(24.6)			-43.7	(31.8)	-34.9	(19.6)	65.1	(27.7)	-3.6	(123.9
	Math Score	Explained	72.6***	(17.6)	14.6†	(7.3)	8.6†	(4.6)	25.9***	(6.2)	12.4**	(3.6)	7.9†	(4.4)			-2.2	(4.3)	0.2	(0.1)	5.2	(12.7)		
	Gap = 1.46	Unexplained	27.4*	(12.6)	31.3	(18.9)	10.3	(12.3)	-6.6	(7.5)	-10.5	(18.8)	3.5	(13.6)			15.5	(13.8)	8.3	(8.3)	4.3	(7.7)	-28.6	(50.1)
	Literacy Score	Explained	67.4†	(36.6)	8.1†	(4.6)	4.7†	(2.6)	10.0*	(3.9)	9.6**	(2.1)							0.1	(0.0)	35.0	(32.9)		
	Gap = .41	Unexplained	32.6***	(8.4)	-50.0†	(25.6)	35.6*	(13.9)	-5.4	(5.4)	0.6	(12.3)							1.2	(3.4)	18.3	(15.4)	32.3	(25.1)

† p ≤ .10, *p ≤ .05, **p ≤ .01, ***p ≤ .001; Standard errors are shown in parentheses