ESSAYS ON HUMAN CAPITAL INVESTMENTS IN DEVELOPING COUNTRIES

BY

EMILIE TERESA BAGBY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Werner Baer, Chair
Assistant Professor Richard Akresh
Associate Professor Elizabeth Powers
Associate Professor Darren Lubotsky

ABSTRACT

This dissertation encompasses three chapters that explore determinants of parental investments in their children's health and education in developing countries. Below are the individual abstracts for each chapter.

Chapter 1: Child Ability, Parental Investments and Child Nutrition in Ecuador

This paper investigates the role of family composition and child cognitive ability in explaining how resource-constrained households make nutritional investment decisions in their children. Parents have private information about their children's abilities and health that is typically not available to researchers. I use a unique panel household dataset from Ecuador's Bono de Desarrollo Humano that contains a measure of child cognitive ability and allows me to estimate its affect on resource allocation. I address reverse causality due to the effects of investments on ability and I use within household fixed effects to look at children to look at the intra-household investment decision. Findings point to the existence of sibling rivalry due to resource constraints; children with more siblings, and children in poor households, are less likely to eat high-quality food. Children with higher abilities are less likely to share a nutritional supplement with another family member, suggesting that parents must decide how to invest their limited resources, and child ability informs that decision. Within households of more than one child, children with higher abilities are more likely to eat higher quality foods than their siblings, even after controlling for child body size.

Chapter 2: Child Ability and Household Human Capital Investment Decisions in Burkina Faso

Using data we collected in rural Burkina Faso, we examine how children's cognitive abilities influence resource constrained households' decisions to invest in their education. We use a direct measure of child ability for all primary school-aged children, regardless of current school enrollment. We explicitly incorporate direct measures of the ability of each child's siblings (both absolute and relative measures) to show how sibling rivalry exerts an impact on the parent's decision of whether

and how much to invest in their child's education. We find children with one standard deviation higher own ability are 16 percent more likely to be currently enrolled, while having a higher ability sibling lowers current enrollment by 16 percent and having two higher ability siblings lowers enrollment by 30 percent. Results are robust to addressing the potential reverse causality of schooling influencing child ability measures and using alternative cognitive tests to measure ability.

Chapter 3: Risk and Protective Factors for School Dropout in Mexico and Chile

Fourteen percent of Chilean youth and 30 percent of Mexican youth have dropped out prior to completing secondary school. Of these youth, 90 to 97 percent are considered "at risk," meaning that they engage in or are at risk of engaging in risky behaviors that are detrimental to their own development and to the well-being of their societies. This paper uses youth surveys from Chile and Mexico to demonstrate that early school dropout is strongly correlated with a range of risky behaviors as well as typically unobservable risk and protective factors. We test which of a large set of potential factors are correlated with dropping out of school early and other risky behaviors. These factors range from relationships with parents and institutions to household behaviors (abuse, discipline techniques) to social exclusion. We use stepwise regressions to sort out which variables best explain the observed variance in risky behaviors. We also use a non-parametric methodology to characterize different sub-groups of youth according to the amount of risk in their lives. We find that while higher socioeconomic status emerges as key explanatory factors for school dropout and six additional risky behaviors for boys and girls in both countries, it is not the only one. A good relationship with parents and peers, strong connection with local governmental institutions and schools, urban residence, younger age, and spirituality also emerge as being strongly correlated with school dropout and different risky behaviors. Similarly, young people that leave school early also engage in other risky behaviors. The variety of factors associated with leaving school early suggests that while poverty is important, it is not the only risk factor. This points to a wider range of policy entry points than currently used, including targeting parents and the relationship with schools.

ACKNOWLEDGEMENTS

I owe a great deal of thanks to a number of people who helped me in making this research possible. I wish to thank my advisor, Werner Baer, for his continued support throughout graduate school as well as the many insightful suggestions he provided related to my project. I am also incredibly grateful to my professor and co-author, Richard Akresh, for his guidance, professional support, and tireless energy in helping me with all aspects of this project. His efforts have made me a better researcher, including when I engage in fieldwork, analysis, and writing. I also wish to thank the other members of my thesis committee, Elizabeth Powers and Darren Lubotsky, for their time and effort in providing me professional advice and comments related to my research. Further, I wish to thank my co-author of Chapter 3, Wendy Cunningham, for her professional support and guidance throughout this journey. I also wish to thank my co-authors, Damien de Walque and Harounan Kazianga, for their support in all aspects of writing Chapter 2, including their support and flexibility in supporting me during the field work process. I am thankful to the University of Illinois at Urbana-Champaign and the World Bank, both of which provide me with the financial means to complete various aspects of this project. I also would like to thank the government of Ecuador for providing me with the data to write Chapter 1. There are a number of other individuals who provided important feedback to strengthen my research, including Dan Bernhardt, Vidisha Vachharajani, and Jason Dunick.

Finally, I wish to thank my parents, Elaine and Larry Bagby, my sisters, Allison and Madelyn, and my best friend, Wladimir Benalcazar. Without their continued love and support during this journey, none of this work would have been possible.

TABLE OF CONTENTS

INTRODUCTION

Parental decisions about whether and how much to invest in their children's human capital depend on many factors, and these decisions have long-lasting impacts on each child's future wellbeing. A large literature attempts to understand the source of inequalities for children's educational investments within a household building on seminal work by Becker and Tomes (1976) that delineates the tradeoff between the quantity of children and their 'quality.' For poor households seeking to maximize the returns to their human capital investments, schooling and health decisions will depend on parent perceptions about the returns to human capital for a given child and that child's ability. In making investment decisions, parents will have information about a child's ability and that information will often not be available to researchers. This partly explains why much of the empirical research on the determinants of household investments in children's schooling focuses on easy to observe demographic characteristics of the child such as gender, birth order, and family composition. Recent research attempts to overcome this difficulty by using datasets with typically unavailable information about child characteristics such as cognitive ability, child comportment, and household environment, to better understand factors that influence investment decisions.

This work contributes to this body of literature, with three chapters that use typically unavailable information about child characteristics to better understand health and education investments in children. In the first chapter, a measure of child cognitive ability is used to understand nutrition investments and within household differences in these investments in Ecuador. The second chapter uses alternative measures of child cognitive ability to better understand education investments in Burkina Faso. In the third chapter, the relationship between different risk and protective factors and school dropout of young people in Mexico and Chile is explored.

Chapter 1, entitled "Child Ability, Parental Investments and Child Nutrition in Ecuador" investigates the role of family composition and child cognitive ability in explaining how resource-constrained households make nutritional investment decisions in their children. I use a unique panel household dataset from Ecuador's Bono de Desarrollo Humano which contains a measure of child cognitive ability, the Spanish version of the Peabody Picture Vocabulary Test, and allows me to estimate its affect on resource allocation. Reverse causality of the effects of investments on ability is addressed through the use of panel data. Findings point to sibling rivalry due to resource constraints; children with more siblings, and children in poor households, are less likely to eat high quality food. Children with higher abilities are less likely to share a nutritional supplement with another family member, suggesting that parents must decide how to invest their limited resources, and child ability informs that decision. Within households with more than one child, children with higher abilities are more likely to eat higher quality foods than their siblings, even after controlling for child body size.

Chapter 2, entitled "Child Ability and Household Human Capital Investment Decisions in Burkina Faso" and co-authored with Richard Akresh, Damien de Walque, and Harounan Kazianga, looks at schooling decisions in Burkina Faso. Using data we collected in rural Burkina Faso, this paper examines how children's cognitive abilities influence resource constrained households' decisions to invest in their education. We directly measure child ability for all primary-school-aged children, regardless of current school enrollment, using the Raven's Colored Progressive Matrices and the Wechsler Intelligence Scales. We explicitly incorporate direct measures of the ability of each child's siblings (both absolute and relative measures) to show how sibling rivalry exerts an impact on the parent's decision of whether and how much to invest in their child's education. We find children with one standard deviation higher own ability are 16 percent more likely to be currently enrolled, while having a higher ability sibling lowers the likelihood of current enrollment by 16 percent and having two higher ability siblings lowers it by 30 percent. Results are robust to addressing the potential reverse causality of schooling influencing child ability measures and using alternative cognitive tests to measure ability.

Chapter 3, entitled "Risk and Protective Factors for School Dropout in Mexico and Chile" and co-authored with Wendy Cunningham, aims to characterize the heterogeneity in the youth population and identify risk and protective factors related to school dropout. Using non-parametric methodologies and specialized youth surveys from Mexico and Chile, different sub-groups of youth are characterized according to the amount of risk in their lives that might contribute to school dropout and other poor outcomes. Then factors and behaviors that are correlated with school dropout, and yet typically not observed in large household datasets, are identified. These factors range from relationships with parents and institutions to household behaviors (abuse, discipline techniques) to social exclusion.

Overall this dissertation aims to improve our understanding of why young people in developing countries may not receive desired levels of education and health investments. Improving child nutrition, health, education and human capital development are recognized as important goals as indicated by the Millennium Development Goals. Understanding if and why parents in resource constrained households may sacrifice one child's educational or nutritional investments for another, or if children self-select out of school, thus contributing to overall inequalities in human capital, can inform policy makers' decisions when attempting to improve human capital levels.

CHAPTER 1
CHILD ABILITY, PARENTAL INVESTMENTS, AND CHILD NUTRITION IN ECUADOR

1. Introduction

Parents face a well-documented tradeoff between the quantity of children they might have and the level of investments in human capital that they make in each child, as posited by Becker and Lewis (1973). Children in the same household might compete with one another to receive human capital investments since parents must divide resources amongst all children in the household and may perceive the payoff to investments in one child to be greater than in another. This inequality in investments may, in turn, lead to inequalities in their success later in life. This project studies these intra-family dynamics by investigating whether or not child cognitive ability can help to explain differences in nutritional investments in young children in Ecuador, a country with high levels of malnutrition.

Using panel data from Ecuador's cash transfer program, the Bono de Desarrollo Humano (BDH), this paper models various measures of children's nutritional investment as a function of children's cognitive ability, controlling for current health status, family composition, and other factors. I use two measures of nutritional investment or lack thereof: the first is an indicator that a child was given a nutritional supplement, but had to share it with a sibling or adult in the family. This variable indicates whether or not parents redistribute nutritional investments away from the intended recipient to other members within the household. Parents in the survey are also asked to report whether their children ate certain types of food during the past week. I classify these foods as either "nutritious" or "non-nutritious," and my second measure of nutritional investment is a measure of how often the child eats more nutritious foods.

This paper builds upon a body of literature that explores the determinants of human capital investments. Much of the empirical research is limited to the use of easy-to-observe demographic characteristics such as gender, birth order, birth spacing, and sibling sex-mix (Garg and Morduch 1998). Recent studies use different direct measures of child ability including cognitive tests (Ayalew 2005 and Akresh, Bagby, de Walque and Kazianga 2010), IQ scores (Kim 2005), achievement tests (Glick and Sahn 2010) and birth weight (Datar et al. 2010, Rosenzweig and Zhang 2009, Loughran et al. 2004) to understand how parents use information about their children when making investment decisions.

This study is unique in identifying the role that child ability plays in explaining child nutrition investments. First, I use family composition in the estimates of nutritional investments to see if there is a tradeoff between the quantity of children and the quality of their nutritional investments, while controlling for child body size and ability. This approach takes current family composition and child characteristics as given at the time the parents make the investment decision. I then use within-household fixed effects to identify how a child's cognitive ability relative to the ability of his or her siblings plays a role in decisions about nutrition to control for any unobserved characteristics that are common to all children in a household, but vary from one household to another. Additional specifications explore the precise role that sibling ability plays in a child's nutrition.

The key challenge in using measures of child ability to explain parental investment decisions is that a child's ability is endogenous and reflects past parental investments and other factors that are unobservable to the researcher. This paper addresses this in several ways. First, I use two measures of body size, child height-for-age and weight-for-height, from the first wave of the panel to control for the stock of previous investments. Height-for-age is considered to be a good measure of long term investments in a child, while weight-for-height is a more contemporaneous measure of a child's nutrition. Second, since current nutrition can affect a child's ability to perform well on a cognitive ability test, I use child cognitive ability measured in the first wave of the panel to explain nutrition investments in the second wave.

Another challenge is that children themselves may influence the investment decision through their own preferences and behaviors. They may make their own choices as to what foods to eat based on their tastes and preferences rather than relying on what their parents tell them to consume. Also, more able children might be more willing to ask for or take the foods they want rather than waiting for someone to tell them what to eat. I cannot address either of these potential explanations for different nutritional consumption among siblings using this dataset. A third explanation has to do with child labor activities; children engaging in demanding physical labor have been shown to consume more calories than their siblings engaging in less demanding activities. However, I argue that since the age of the children at the time of the nutritional input is between 4 to 7 years in this study, the children are at similar points in their lives and therefore less likely to be treated differently because of age related activities. Young children are less likely than older youth to be engaging in labor requiring high energy, which limits this as a possible explanation for why one sibling might consume more than another. In addition, since these children are not spaced far apart, they are more likely to be competing based upon what parents observe about their children rather than with feedback from sources outside of the household (such as through the use of grades in school).

The results from this study indicate that higher ability children receive greater nutritional investments than their lower ability siblings. Children with a one standard deviation higher ability consume 5 percent more nutritious foods. Results are robust to the use of different measures of nutritious food consumption. I also find that young children in poor households with more children have lower levels of consumption of nutritious foods compared to children in non-poor households with similar numbers of children. Children with one additional sibling are 16 percentage points more likely to share nutritional supplements and three percentage points less likely to consume high levels of nutritious foods. These findings point to inequalities within the household as one source of inequality in nutrition amongst children in Ecuador.

The results in this paper are consistent with, and build upon, those in previous work by Akresh, Bagby, de Walque and Kazianga (2010), who examine a similar question with regards to school enrollment in Burkina Faso. In that paper we focus on the role of ability on the household decision to invest in the education of each child between the ages of 5-15, finding that higher ability children receive more education investments than their lower ability siblings. A potential concern with that paper is that cognitive ability is measured at the time of and just prior to the education investment, and therefore might be capturing all investments in the life of the child up until age 5 or older. In this paper I focus on a younger age group, ages 3-6 at the time of the ability measure, and look at the relationship between ability and parent investments in nutrition. There are still previous investments from birth to age 3, but this window is smaller, and controlling for height-for-age and weight-for-height captures some of the previous investments.

Ecuador is a country with a well documented level of failure to provide adequate nutrition to all children, with 23 percent of the children under 5 being stunted.[1] Understanding how households make decisions to invest in each of their children can inform policies aimed at eliminating malnutrition as well as those focused on improving overall human capital. Targeted nutritional interventions or non-targeted poverty reduction strategies can have different affects on the nutrition of children in the same household. Under Becker's (1974) unitary model of the household, a total increase in household resources is equivalent to transfers to individual members in the household. If the cost of targeting to individuals is high, or the benefit to the household is lower from in-kind nutrition transfers to the household than from cash transfers or poverty alleviation programs, then poverty reduction strategies could have a larger impact on child nutrition. On the other hand, targeted transfers to individuals may be more likely to stick to the individual than to be reallocated away, or act as "flypaper", as discussed in Jacoby (2002) and Afridi (2010). Similarly, the presence of

[1] See World Bank (2007) for a description of the stunting in Ecuador

nutrition programs that target individuals may have a "labeling" effect as discussed by Kooreman (2000), or a change the way parents make decisions because of the new information they have about the nutritional status of their children.

The remainder of the paper is organized as follows. Section 2 discusses the conceptual framework about the household nutrition investment decision. Section 3 describes the survey data used in the analysis and explains the construction of the different nutrition measures. Section 4 describes the empirical identification strategy and section 5 presents the main results as well as robustness tests. Section 6 concludes.

2. Conceptual Framework

Becker and Lewis' (1973) tradeoff between the quantity of children and the quality of each child in the household has been documented around the world. In addition, inequalities in human capital outcomes for children in the same family have been observed. Parental preferences and resource constraints are one explanation for these inequalities. If parents have limited resources, they may have to make sacrifices when choosing how to distribute these resources within the household. These decisions are made by parents based on preferences over their children's futures, knowledge about the labor and marriage markets and information about their children.

A large body of literature attempts to explain why children in the same household might have differing human capital outcomes or investments. Many developing country empirical analyses use characteristics that are easily observable and typically available in large household surveys, such as gender, birth order, sibling sex mix and other measures of family composition.[2] If parents are forced to choose amongst their children in distributing limited resources, then siblings become rivals and family composition is important in determining investments. Garg and Morduch (1998) find that sibling rivalry plays a strong role in child health outcomes including height-for-age, weight-for-

[2] See Strauss and Thomas (1995), Strauss and Beegle (2000) and Glewwe and Kremer (2006) for reviews of the literature.

height, stunting, wasting and being underweight in Ghana. They also find that children with only sisters are better off than their counterparts with only brothers, and children with older siblings are better off than children with younger siblings. Their explanation of credit constraints and differences in relative returns to education is consistent with findings by Dahan and Gaviria (2003) who propose a model of inequalities in human capital investments within households even if children have no differences in abilities.

Additional recent studies that look at different human capital outcomes find results consistent with sibling rivalry. For instance, Edmonds (2007) in Nepal, Ota and Moffatt (2007) in India, and Dammert (2010) in Nicaragua and Guatemala find evidence consistent with sibling rivalry in education investments and in child labor. Most of these developing country studies also find birth order and gender effects; children with older siblings are better off than their older siblings, and children with sisters are better of than children with brothers. For instance, Parish and Willis (1993) find that children receive different investments based on gender as a result of cultural traditions in Taiwan. On the other hand, studies in the United States suggest different patterns. Butcher and Case (1994) find evidence of sibling rivalry; however they also find that children with only brothers receive more education than children with only sisters. Such studies do not address the relationship between child ability and human capital investments.

One set of models of the parents' decision making process from the human capital literature takes into account the fact that parents have more information about their children's potential to succeed in the labor or marriage markets than is available to the typical researcher. These models assume that parents have preferences over their children's future success and use the information they have about a child's ability when making investment decisions.[3] Becker and Tomes (1976) and

[3] These models assume that parents come to a consensus when making these decisions as opposed to other models which take into account bargaining power in the household. See Behrman (1997) for a complete explanation of the consensus parental preferences models

Behrman, Pollack and Taubman (1982) use child ability in the model to attempt to identify parental preferences in making the investment decision. They examine the role of endowments without directly observing child abilities or endowments (additional related papers include Rosenzweig and Schultz, 1982 and Behrman, Rosenzweig and Taubman, 2004). Results from these studies suggest that parents slightly reinforce differences in their children's human capital by providing more investments to the more able children.

Direct measures of a child's endowments, or innate ability, are virtually impossible to come by. A growing body of literature uses the first observable measure of endowment, a child's birth weight, while another strand uses different direct measures of a child's ability such as height or cognitive ability tests to look at both health and education investments and outcomes.

Empirical analyses that use measures of cognitive ability or birth weight to explain educational outcomes find results consistent with those studies that do not directly measure endowments. Behrman and Rosenzweig (2004) use birth weight to explain later success in earnings and completed education. A growing literature that directly measures a child's cognitive ability to explain the education investment decision includes papers by Ayalew (2005) in Ethiopia, Kim (2005) in Wisconsin, and Glick and Sahn (2010) in West Africa. Akresh, Bagby, de Walque and Kazianga (2010) expand on this literature by measuring child ability for all children ages 5-15 in the household, regardless of school enrollment status in Burkina Faso. By measuring the cognitive abilities of all children in the household, we show that sibling ability rivalry impacts the parents' decisions about whether and how much to invest in each child's education.

The literature that evaluates the determinants of health investments is mixed. Some analyses in developing countries have found that parents tend to compensate for differences in health endowments while others find that parents reinforce for differences in child endowments when making health investments early in life. As with the educational investments literature, some studies that look at health investments (or outcomes) use birth weight as a measure of child endowment. Rosenzweig and Zhang (2009) use information about birth weight in twins in China, finding evidence of the quantity-quality tradeoff and of preferences towards stronger children. Loughran, Kilburn and Datar (2007) find that low birth weight children fare better in smaller families than in larger ones. Datar, Kilburn and Loughran (2010) and Datar, Ghosh and Sood (2007) find that parents tend to reinforce for differences in birth weight when making investments in a child's health early in life. Ayalew (2005), on the other hand, uses predicted values from a health production function as a measure of exogenous child endowment, finding that children with lower health endowments receive more health inputs. While Pitt, Rosenzweig and Hassan (1990) find that households are inequality averse, they also find that boys receive more calories because they are more likely to engage in activities requiring better health status to be more productive. These studies do not consider the role that cognitive ability plays separately from the child's health.

Studies that explore the relationship between nutrition and cognitive development in developing countries find that malnutrition, resulting from a low level and poor mix of nutritional investments, can explain poor cognitive development. For instance, Glewwe and King (2001) find that malnutrition is an important determinant of cognitive development. Aldermen et al (2006) find large effects of the impact of malnutrition prior to starting school on later human capital formation. To the author's knowledge, there are no papers to date which directly explore the role of child cognitive ability in explaining parental investments in a child's nutrition.

3. Data

The data used in this analysis were collected to evaluate Ecuador's cash transfer program, the Bono de Desarrollo Humano (BDH). The data have a panel structure; the baseline was collected in 2003-2004 and the follow-up in 2005. The data were collected to evaluate the impact of the health intervention which focuses on health inputs and outcomes of children under age 8 in the second wave. Households residing in 6 different provinces, Pichincha, Loja, and Azuay in the highlands and Los Ríos, El Oro and Esmeraldas on the coast, were randomly assigned to treatment. Only households with at least one child under age 6 and no child over age 15 during the first wave were eligible.

The sample used in this analysis is comprised of 3,485 children between the ages of 4-7 years in the second wave (or 3-6 years in the first wave) as shown in the summary statistics in Table 1.1. Of these children, 769 are children with siblings in the same age range and these children are in 389 households. Most children in this sample have one sibling in the same age group (4-7), of which half are female. 69 percent of the sample is located in rural areas. On average, both father's and mother's education is just under 8 years.[4] 72 percent of the children have their father present in the home.

Families were assigned to treatment after having qualified for the intervention through the use of the Selben index, which is a system of selecting beneficiaries of social programs. This index is comprised of 27 different variables including demographic and education characteristics of the household and household assets.[5] 41 percent of the sample was assigned to the treatment group and is classified as "poor" in this analysis. Beneficiaries received $15 per family per month, which is approximately 15 percent of the average monthly expenditure of eligible families.

The survey uses the Spanish version of the Peabody Picture Vocabulary Test (TVIP) to evaluate the cognitive development of young children and is used as the measure of child ability in this analysis. The TVIP or its English counterpart, the PPVT, are tests of language comprehension and are internationally used as approximations of IQ. Because of the age component to the test, this study normalizes the score to have a mean zero and standard deviation of one for each age group.

[4] Additional information about the mother is available, including the mother's TVIP score and her height. I do not report these here since the main result of child ability does not change in the regressions with the inclusion of mother's ability or height.

[5] See Ponce and Bedi (2008) for a detailed explanation of the Selben

In addition to cognitive ability, these data have anthropometric measures for all children in the household under age 8. Height-for-age and weight-for-height z-scores were normed using CDC growth charts. On average, the height-for-age z-score is -1 and weight-for-height z-score is -0.2 for this sample, reflecting Ecuador's high level of malnutrition.

This analysis uses two different measures of investment in nutrition. The first measure is an indicator for whether or not the children who received a government nutrition supplement, called Nutrinnfa, shared their supplement with someone else in the household, either a sibling or adult. This supplement is a powdered food that is reinforced with micronutrients.[6] 528 children in the sample received the nutrition supplement. Parents redistributed this nutritional investment away from the intended recipient to other members within the household in 63 percent of the cases.

The second measure of nutrition investments aggregates different food consumption variables into a single variable which represents nutritional investments in nutritious foods. Respondents were asked of each young child whether or not the child ate a particular type of food in the past week and if so, how many times. The food categories include: liver; cow internal organs; noodles, bread and other wheat products; spinach or chard; citric fruits; other non-citrus fruits; carrot or similar vegetable; soda or ice-cream; cookies or pastries; fried snacks including fries, fried bananas, and pork skin; and candies, chocolate, lollipops, or gum. The values of liver, cow internal organs, noodles or bread or other foods derived from wheat flour, spinach or chard, and carrots are summed together to create a nutritious food variable. Wheat flour in Ecuador is fortified with iron and other nutrients, and these other nutritious foods are mechanisms of receiving iron or vitamin A amongst other micronutrients.[7] Deficiencies from iron and Vitamin A are well documented in Ecuador and iron is of particular importance in the high altitudes many of the children are living. While fruits are also nutritious foods, they are very common and very cheap in Ecuador and are

[6] World Bank (2007) has a discussion of the supplement and how it fits within Ecuador's strategy to improve nutrition.
[7] World Bank (2007) discusses Ecuador's fortification of wheat flour.

therefore not included. Children consumed the total of the individual nutritious food items a total of nine times in the previous week with a standard deviation of five times. Four percent of the children ages 4-7 with siblings in the same age group consumed more than one standard deviation above the household average of the sum of nutritious foods.

This measure of nutritious food consumption is somewhat imprecise; it simply asks whether the child ate a certain type of food in the past week or not, and if so, the number of times. There is no information as to how much of each food the child ate at each sitting. In addition, studies have observed that food recall is imprecise.[8] However, this analysis is concerned primarily with the differences in sibling nutrition, and since recall for the nutrition of children in the same family should have similar measurement error, the use of these variables is reasonable. In addition, the impact evaluation of this intervention, by Paxson and Schady (2007), finds that while the intervention had an impact on health and education outcomes, there is no impact on food consumption using these variables. Therefore, it is reasonable to use the nutrition variables from the second wave of the survey, after the intervention, as outcome variables.

To check the robustness of the results, alternative forms of high nutritious food consumption, groupings of the nutritious foods, and a measure of non-nutritious foods are used. First, an indicator for whether or not the child received more than one standard deviation above the average family consumption of the nutritious food variable, since within-household differences in nutritional investments are of interest. Second, consumption of liver and spinach only is therefore summed as an alternative measure of nutritious food investments. Liver and spinach are both foods that provide iron, a vitamin of importance for children living in high altitudes. Finally, non-nutritious foods are then considered, to see if similar decisions are being made as in nutritious foods.

[8] See Baranowski, Sprague, Baranowski and Harrison (1991)

The measure of non-nutritious foods sums the number of times the child consumed soda or ice-cream, cookies or pastries, fried snacks, and candies.

Figure 1.1a shows that children in non-poor households consume more nutritious food than children in poor households and that children with more siblings consume less on average than children with fewer siblings in poor households. On the other hand, as shown in Figure 1.1b, children in both poor and non-poor households consume similar levels of non-nutritious foods, regardless of the number of siblings. This suggests that children in poor households are more likely to be rivals for parental investments in their human capital than children in non-poor households. Panel B of Table 1.1 shows that 30-36 percent of the overall sample variation in weight-for-height and height-for-age z-scores arises from within household variation, while three percent of the overall variation in consumption of nutritious foods arises from variation within households.

4. Empirical Specification

The empirical strategy is an expansion of the typical parental investment or sibling rivalry models, which do not include a measure of child ability. The role of family composition in the investment decision is identified using an ordinary least squares model as follows:

$$I_{ih2} = \beta_0 A_{ih1} + \omega_0 S_{ih1} + \omega_1 F_{ih1} + \alpha_0 X_{ih1} + \alpha_1 Z_{h1} + \eta_{ih} \quad (1)$$

where I_{ih2} is the nutrition investment for child *i* in household *h* in period *2*, A_{ih1} is a measure of observed child ability in period *1*, S_{ih1} is a count of the total number of siblings the child has in period *1*, F_{ih1} is a count of the total number of female siblings the child has in period *1*, X_{ih1} is a vector of individual characteristics from period *1* such as age, gender, height-for-age and weight-for-height z-scores that might influence parental investments, Z_{h1} is a vector of observed household characteristics in period *1*, and η_{ih} is composed of the unobserved characteristics of child *i* in household *h* that might influence the decision, all unobserved characteristics about the household that are constant across siblings, and a random idiosyncratic error term.

Several measures of the nutrition investment are considered in this specification, including whether or not a child shared a nutrition supplement with another household member, consumption of nutritious foods, and consumption of non-nutritious foods. The interpretation of ω_0 is the change in I_{ih2} associated with an additional male sibling. The interpretation of ω_1 is the change in I_{ih2} associated with converting a sibling from a male to a female while holding the number of siblings constant. $\omega_0 + \omega_1$ is the change in I_{ih2} associated with the addition of a female sibling. To address for the reverse causality of nutrition affecting ability, as well as for individual child heterogeneity, this study uses a value added approach. The ability measure from the first wave of the survey, A_{ih1}, is used to look at investments in the second wave, I_{ih2}. In addition, the height-for-age z-score and the weight-for-height z-score from the first wave of the survey are included in X_{ih1} to control for previous investments in the children. X_{ih1} also includes child gender, birth order and age dummies.

Estimates from equation (1) identify the role of family composition in explaining nutrition investments. These estimates, however, do not look at the decision within households. A household fixed effects model is therefore used to identify the role of child ability in the decision within the household. This model controls for observed and unobserved household level characteristics that are constant across siblings. This compares a child's own ability to the average ability for siblings in the household as follows:

$$(2) \qquad I_{ih2} = \beta_0 A_{ih1} + \alpha_0 X_{ih1} + \lambda_{h1} + \varepsilon_{ih}$$

where I_{ik2}, A_{ih1}, and X_{ih1} are as defined above, λ_{h1} is the household fixed effect that captures all characteristics about the household that are constant across siblings, and ε_{ih} is a random, idiosyncratic error term. Observed household characteristics, Z_{h1} and number of siblings, S_{ih1}, have no variation across children within the household and are therefore dropped. The number of sisters, F_{ih1}, is also dropped because it is constant within a given household for children of the same gender. The coefficient β_0 can be interpreted as the impact on nutritious food consumption of child *i*'s ability relative to his or her siblings' average abilities. This specification is used to consider consuming nutritious and non-nutritious foods. Sharing a nutrition supplement is not considered in this specification for sample size issues. The sample is restricted to children between the ages of 4-7 who have at least one sibling in the same age range. While this is a data limitation, since only children in this age range have ability measures, the result is to compare only children that are in similar situations in life as previously discussed.

The household fixed effects estimation compares own ability to average sibling ability; however it is possible that alternative forms of the relationship of a child's ability with his or her siblings' abilities might be more representative of reality. For instance, it could be that parents consider only the highest ability amongst their children when making investment decisions rather than average overall ability. I therefore estimate a variation of equation (1) to include explicit measures of sibling ability as follows:

$$I_{ih2} = \beta_0 A_{ih1} + \beta_1 g(A_{-ih1}) + \omega_0 S_{ih1} + \omega_1 F_{ih1} + \alpha_0 X_{ih1} + \alpha_1 Z_{h1} + \eta_{ih} \quad (3)$$

where all terms are as defined in equation (1) except for the new term $g(A_{-ih1})$ which is the direct measure of the abilities of other children *(-i)* in household *h* during period *1*. The coefficients β_0 and β_1 respectively give an estimate of the impact of child *i*'s own ability and his sibling's ability on child *i*'s investments. The different measures of sibling ability which are used include the average and the highest ability. Similar controls for sibling height-for-age and weight-for-height are included.

5. Results

5.1 Sibling Rivalry and Child Ability

The results presented in Table 1.2 provide evidence that parents decide to reallocate resources within the household and that child ability plays a role in the decision. One purpose of this dataset is to evaluate the effects of a government intervention to provide nutritional supplements to young children. Poor households were assigned to treatment groups that receive different supplements for each young child in the household depending on each child's age. A probit analysis is used to predict the probability that a child who received the government nutrition supplement shares their supplement with someone else in the household, either a sibling or adult. If a child is sharing the supplement that is intended for him or her, then parents are deciding that someone else in the household deserves the supplement in addition to that child and the child should share it and not consume their total allotted amount. Regressions control for individual and family characteristics from first wave of the panel with the nutrition supplement received afterwards. Regressions in columns 1 to 4 include controls for family wealth level and living in a rural area.[9]

Results point to sibling rivalry since children with more siblings are more likely to share the supplement than children with fewer siblings. An additional sibling increases the probability of sharing the supplement by 15 to 20 percentage points. Column 1 presents the results for the typical sibling rivalry analysis, which does not include child ability as an explanatory term, showing evidence of sibling rivalry. The point estimate decreases in magnitude with the inclusion of additional controls in columns 2 to 6 such ability, child body size, and other family characteristics, however it remains significant and large at 15 percentage points. Columns 1 and 3 suggest that gender may be important as well. Boys are more likely to share their supplement with another member of the household, however once controlling for additional household characteristics, the coefficient is no longer significant in column 5.[10] Column 2 presents the results with controls for child height-for-age and weight-for-height, neither of which is significant. Column 3 shows that children with higher cognitive abilities are 12 percentage points less likely to share their nutritional supplement . Column

[9] Regressions include child gender and age dummies and province fixed effects. Correlation among the error terms of children in a given parish experiencing the same nutrition environment might bias the standard errors downward, so in all regressions the standard errors are cluster by parish.
[10] Additional specifications that interact gender with ability were run and are not presented here. They show that while ability is protective, boys of higher ability are more likely to share the supplement. Additional specifications where ability was interacted with birth order were run, with no significant effect of the interaction term.

4 includes controls for height-for-age and weight-for-height and column 5 includes additional family characteristics including parent's educations, father's presence in the home, and being indigenous. Children with a one standard deviation higher ability are 12 to 15 percentage points less likely to share their supplement with another member of the household. Also, height-for-age and weight-for-height z-scores are not significant, suggesting that cognitive ability may play a larger role in this decision than body size. Concerns about the multi-colinearity of the height-for-age and weight-for-height z-scores might suggest that only one of these measures be included as a control. However, results are robust to controlling for only height-for-age or weight-for-height z-scores, or both.[11]

Ordinary least squares estimates in equation (1) describe the role of family composition and child ability across households in Tables 1.3 and 1.4. The dependant variable in Table .3 is the nutritious food variable, and in Table 1.4 is an indicator for whether or not the child consumed more than one standard deviation above the household mean of nutritious food.[12] Results were separated for poor and non-poor households since wealthier households consume more and have less inequality in food consumption as shown in Figure 1.1a. Cognitive ability, body size, and family and child characteristics as measured in the baseline are used to explain nutrition in the follow-up. Regressions include controls for family wealth level, living in a rural area and receiving the cash transfer intervention.[13] Regressions in columns 4 and 8 for both tables also include parent's education, and father's presence in the home and the sample of children in all households is used.

Results suggest the presence of sibling rivalry in the household; controlling for household wealth, children with more siblings are less likely to consume as much nutritious foods as children with fewer siblings (columns 1-2 Table 1.3, columns 1-8 Table 1.4). Birth order effects consistent

---

[11] Results are consistent with and without controls for being a treatment household or not and for controlling for mothers TVIP score and mother's height. In the interest of space, only the results when including both height-for-age and weight-for-height are presented.
[12] Results for the binary variable are consistent when run using a logit model rather than ordinary-least-squares.
[13] Regressions include child gender and age dummies and province fixed effects. Correlation among the error terms of children in a given parish experiencing the same nutrition environment might bias the standard errors downward, so in all regressions the standard errors are cluster by parish.

with the developing country literature are also present with the nutritious food variable in Table 1.3; younger children are more likely to consume more nutritious foods than their older siblings. Child ability is also positively correlated with food consumption in the specifications that do not include controls for parent education. Having a one standard deviation higher ability is correlated with consuming up to five percent more nutritious foods. Once additional controls for parent's education levels, for being indigenous, and for the father being present in the home, are included child ability is no longer significant. This might suggest that children in wealthier households or with more educated parents have higher abilities and therefore these household controls are absorbing the effect of child ability. In Table 1.4, when considering high levels of food consumption relative to one's siblings, the coefficient on ability remains significant and positive for children in poor households in all specifications. Having a one standard deviation higher ability is correlated with 0.02 percentage point higher probability of consuming one standard deviation above the household mean. In non-poor households, the coefficient for ability is zero but insignificant. Results are robust to controlling for height-for-age, weight-for-height, or both.[14] With these controls, body size appears to be important as well with as strong correlation between nutrition and body size. The point estimates for children in poor households are similar to those in non-poor households in Table 1.3, but are very different in Table 1.4. To look more closely at the relationship between resource constraints and ability, the sample was pooled and ability was interacted with household poverty indicator. In those regressions, the point estimates for Table 1.3 do not vary, ability is positively correlated with nutritious food consumption, but once including controls for parent education the coefficient is not significant. Also, the interaction of ability with poverty level was not significant.[15]

[14] Again, in the interest of space, only the results when including both height-for-age and weight-for-height are presented. If included individually results are consistent with those presented.
[15] These results will be provided upon request.

The findings suggest that children in larger households with more siblings consume less nutritious foods and are more likely to be in competition with each other in poor households than in non-poor households. Also, these findings suggest that while higher ability children are consuming more nutritious foods across households, this could simply reflect that more able children are in more able households that can afford to purchase more healthy foods. The results in Tables 1.3 and 1.4 do not inform us about what is happening within-households.

In order to understand the intra-household relationship between ability and food consumption, it is important to look at the determinants of food consumption within, rather than across, households. Table 1.5 presents the within-household fixed effects estimates from equation (2) for the two alternative measures of nutritious food consumption used in Tables 1.3 and 1.4. The sample is limited to households with at least 2 children between the ages of 4-7 during the second wave of the survey to compare children that are in similar places in life, and is split by household wealth.[16] Results provide evidence that within the household, child ability is correlated with food consumption for poor households but not in non-poor households.[17] Poor children with a one standard deviation higher ability consume nutritious foods 0.55 more times in a week, which corresponds to consuming 6 percent more nutritious food. In other terms, children with a one standard deviation higher ability have a 0.04 percentage point higher likelihood of consuming more than one standard deviation above the household mean. Cognitive ability matters more than body size, regardless of whether height-for-age or weight-for-height are included individually or together.[18]

[16] This is due to data restrictions, since only children under age 8 in the follow-up have ability measures and other characteristics.
[17] These findings are even more interesting when considering that the OLS regressions (tables 3 & 4) on this sample of household with at least 2 children. OLS regressions across households in Tables 3 and 4 do not show a relationship between ability and food consumption on this sample.
[18] Results when pooling the sample and interacting child ability with the household povery indicator are insignificant. The point estimates for poor and non-poor households are different, and leaving the sample split by household wealth shows the different coefficients.

Table 1.6 presents the estimates from equation (3) which explicitly measures sibling ability to look closely at sibling ability rivalry.[19] Results are consistent with the within-family fixed effects estimates from Table 1.5. A child with siblings of higher ability is less likely to consume nutritious foods. A child whose siblings have a one standard deviation higher average ability consumes nutritious foods 0.6 fewer times in a week and has a 0.02 percentage point lower likelihood of consuming one standard deviation above the household mean. The point estimate for the average and the highest sibling ability values is similar because most children have only one sibling with whom they are being compared in this sample. This data limitation also explains why I am unable to use the varied specifications used in Akresh et al. (2010). In that paper, we had large households with many children between the ages of 5-15 and explicitly measured sibling ability using three different forms: highest sibling ability, a dummy for having any sibling with a higher ability and dummies for the number of siblings of higher ability. This sample does not allow such varied specifications. I do, however, include explicit measures of sibling height-for-age and weight-for-height in Table 1.6. Their coefficients are largely insignificant, as expected given the results in Table 1.5.

5.2 Robustness Checks

The subsequent specification considers the relationship between child ability and consumption of non-nutritious foods within households. If parents are concerned about their children's nutritious development when making choices of how to invest their resources, then eating non-nutritious foods should not be a significant factor. The non-nutritious food variable is the summation of the number of times the child consumed each of 4 different categories of food in the previous week. These categories are soda or ice-cream, cookies or pastries, fried snacks, and candy; all are foods that do not provide high levels of micronutrients necessary for healthy child

[19] The sample is split by poor and non-poor households with at least 2 children.

development. Results from the within household fixed-effects specification are presented in Table 1.7. Contrary to its relationship with nutritious foods, child ability does not play an explanatory role in the consumption of non-nutritious foods in poor and non-poor households. Regardless of the form that the non-nutritious food variable takes, the coefficient on child ability is non-zero and insignificant.

As a second robustness check, a different measure of nutritious food consumption is presented in Table 1.8. The measure is an indicator for whether or not the child consumed more than one standard deviation above the household mean consumption of liver and spinach. Both are foods high in iron, which is particularly important to those living in high altitudes as are many of Ecuador's children. Results are consistent with those from previous tables. Having a sibling with a one standard deviation higher ability corresponds to a 0.02 percentage point lower probability of eating more than one standard deviation of the household mean consumption of liver or spinach. The sample of poor families that consume liver or spinach is small since both are relatively expensive foods, which can explain why the point estimates are small across the various specifications and insignificant in the within-household fixed effect specification. Results are present for children in poor households and not in non-poor households.

6. Conclusion

In this paper I provide evidence that sibling rivalry affects nutritional investments in Ecuadorian children using multiple measures of nutritional consumption. Children receiving supplements from a government intervention are more likely to share their supplement with other members of a household when they have more siblings. However, children with higher abilities are less likely to share their supplement than children with low abilities. Children with a one standard deviation higher ability are 15 percentage points less likely to share a government-provided nutritional supplement, which corresponds to a 24 percent lower likelihood of sharing the supplement. Child cognitive ability is also significant in explaining within household differences in the consumption of nutritious foods. Children with a one standard deviation higher ability consume 6 percent more nutritious food than their lower ability siblings. Results hold regardless of whether or not controls for previous investments and individual child heterogeneity, by using height-for-age or weight-for-height or both, are included. Taken together, these findings suggest that parents use information about child ability in making the decision to (re)allocate resources within the household. Results are present for poor households and are not present for non-poor households, suggesting that resource constraints are one explanation for why children may experience sibling ability rivalry.

Improving child nutrition, health and human capital development are recognized as important goals as indicated by the Millennium Development Goals. If parents in resource constrained households sacrifice one child's nutritional investments for another, thus contributing to inequalities in human capital, policy makers must take this into account when attempting to reduce malnutrition. Direct nutritional interventions targeting a single child may be thwarted by within-household decisions that divert resources to 'stronger' or more able sibling rivals. However, non-targeted overall reduction in poverty and reduced credit constraints might actually help the weaker siblings particularly if the overall benefit to the household is larger from the poverty reduction strategy than from a transfer targeted at individuals. While studies do find "flypaper" effects of nutrition programs, where the transfer sticks to the targeted individual, who benefits from the more nutritious supplement that was distributed, they do find more reallocation in poorer households. Any project to evaluate the impact of nutritional interventions should include measures of cognitive ability in addition to child health, body size and work effort to understand the intra-household investment implications.

FIGURES AND TABLES FOR CHAPTER 1

Table 1.1. Summary Statistics

Panel A: Sample of households with at least 1 child in the 4-7 age range during Second Wave	Poor Households Only			Non-Poor Households Only		
	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev
Own Ability (TVIP age adjusted z-score)	2650	-0.23	0.83	835	0.79	1.11
Height-for-age Z-Score	2393	-1.07	1.07	802	-0.96	0.96
Weight-for-height Z-Score	2535	-0.22	1.05	783	-0.07	1.03
Number of Siblings 14 and Younger	2650	1.22	1.15	835	0 .83	0.89
Number of Female Siblings 14 and Younger	2650	0.59	079	835	0.41	065
Number of Siblings Ages 4-7 during Second Wave	2650	1.19	0.92	835	0.91	092
Male	2648	0.50	0.50	833	0.52	0.50
Age in Years	2650	5.42	0.84	835	5.04	0.87
Treatment or Control Household	2650	0.54	0.50	835	0.01	0.09
Quintile grouping from Selben Index	2650	12	0	835	34	0
Log Selben (Wealth Index)	2650	13.89	0.59	835	14.11	0.75
Indigenous Indicator	2650	0.06	0.24	835	0.04	0.19
Rural Indicator	2650	0.59	0.49	835	1	0
Father's years of education	2409	7.11	3.20	796	10.42	3.72
Mother's years of education	2647	7.13	3.11	835	10.42	3.69
Father is present in the home	2649	0.69	0.46	835	0.80	0.40
Sharing of the Nutrition Supplement (Nutrinnfa)	465	0.63	0.48	63	0.65	0.48
Nutritious Food	2649	8.48	4.93	835	11.99	4.93
Greater than 1st dev above mean HH Nutritious Food	1992	0.07	0.25	451	0.09	0.29
Non-nutritious Food	2649	7.61	5.96	835	7.95	6.0
Greater than 1st dev above mean HH Non-nutritious Food	1992	0.13	0.33	451	0.15	0.36
Panel B: Percent of overall variation arising from variation within households with at least 2 children ages 4-7 during wave 2						
Height-for-age Z-Score	30%					
Weight-for-height Z-Score	36%					
Nutritious Food	3%					
Non-nutritious Food	4%					

Notes: : All summary statistics are based on information for the 3485 children age 4 to 7 in Wave 2 unless data is missing for the variable. Child ability is measured using the Spanish language version of the Peabody Picture Vocabulary Test (TVIP) and normed by age (z-score). 528 children received the government nutrition intervention, a supplement called Nutrinnfa. The nutritious food variable is the summation of the number of times the child consumed liver, cow internal organs, noodles or bread or other wheat flour products, spinach/chard, and carrots or similar vegetables in the previous week. The non-nutritious food variable is the summation of the number of times the child consumed soda or ice-cream, cookies or pastries, fried snacks, and candies in the previous week. The Selben is an index created by the government to identify potential beneficiaries of social programs. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.2. Probability that a Child Receiving Nutrition Supplement (Nutrinnfa) Shares it with Siblings or Other Family Members, As a Function of Child Ability, All Children Receiving the Supplement

	(1)	(2)	(3)	(4)	(5)
Own Ability (TVIP age adjusted z-score)			-0.126*	-0.103	-0.152**
			[0.067]	[0.070]	[0.076]
Height-for-age Z-Score		-0.015		-0.009	-0.027
		[0.064]		[0.064]	[0.071]
Weight-for-height Z-Score		-0.017		-0.014	-0.035
		[0.057]		[0.056]	[0.059]
Number of Siblings	0.200***	0.183**	0.195***	0.179**	0.163**
	[0.070]	[0.073]	[0.068]	[0.072]	[0.075]
Number of Sisters	-0.118	-0.108	-0.119	-0.105	-0.123
	[0.099]	[0.101]	[0.098]	[0.101]	[0.116]
Male	0.201*	0.177	0.207*	0.184	0.210
	[0.110]	[0.117]	[0.112]	[0.118]	[0.130]
Birth order	-0.048	-0.030	-0.035	-0.019	-0.003
	[0.179]	[0.189]	[0.182]	[0.192]	[0.194]
Father's years of education					0.016
					[0.025]
Mother's years of education					0.015
					[0.022]
Father is present in the home					-0.088
					[0.153]
Indigenous Identifier					0.337
					[0.318]
Rural	0.475**	0.391*	0.471**	0.389*	0.459*
	[0.219]	[0.224]	[0.223]	[0.226]	[0.239]
Log Selben (Wealth Index)	0.144	0.142	0.139	0.134	0.144
	[0.110]	[0.115]	[0.110]	[0.115]	[0.116]
Age FE	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes
Observations	528	470	528	470	428

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All probit regressions run with age and province fixed effects. Independent variables are from wave 1 of the panel, while the dependent variable is from wave 2 of the panel. Regression sample includes 528 children that were eligible and received the Nutrinnfa supplement and had taken the TVIP test, of which 470 had anthropometric measures and 428 had anthropometric measures and additional information about the household. All children are in poor households and therefore eligible for the BDH intervention, or children are in households that were just above the cutoff for being poor. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.3. Relationship between Nutritious Food Investments and Child Ability, by Household Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Poor				Non-Poor			
Own Ability (TVIP age adjusted z-score)		0.436***	0.378***	0.076		0.333*	0.333*	0.215
		[0.119]	[0.128]	[0.132]		[0.169]	[0.170]	[0.155]
Height-for-age Z-Score			0.307***	0.231**			0.156	0.010
			[0.112]	[0.109]			[0.157]	[0.175]
Weight-for-height Z-Score			0.206*	0.219*			0.260	0.293
			[0.112]	[0.117]			[0.189]	[0.177]
Number of Siblings	-0.263**	-0.214*	-0.143	-0.068	0.041	0.091	0.212	0.278
	[0.122]	[0.123]	[0.141]	[0.149]	[0.235]	[0.224]	[0.262]	[0.283]
Number of Sisters	-0.018	-0.029	-0.045	-0.081	-0.295	-0.330	-0.535	-0.561
	[0.185]	[0.188]	[0.204]	[0.209]	[0.382]	[0.367]	[0.400]	[0.392]
Male	-0.120	-0.123	-0.105	-0.182	-0.011	-0.019	0.130	0.188
	[0.190]	[0.189]	[0.201]	[0.210]	[0.257]	[0.253]	[0.263]	[0.284]
Birth order	0.072	0.055	0.023	0.175	0.767*	0.777*	1.096**	0.941**
	[0.265]	[0.266]	[0.280]	[0.277]	[0.437]	[0.436]	[0.430]	[0.472]
Father's years of education				0.141***				0.170***
				[0.039]				[0.047]
Mother's years of education				0.244***				-0.019
				[0.042]				[0.046]
Father is present in the home				0.111				0.195
				[0.270]				[0.339]
Indigenous Identifier				0.932*				0.010
				[0.542]				[0.769]
Household Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,647	2,647	2,352	2,143	833	833	766	731

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All OLS regressions run with age and province fixed effects. Independent variable from wave 1 of panel and dependent variable from wave 2 of the panel. Regression sample in columns 1-4 includes 2647 poor children that had taken the TVIP cognitive ability test, of which 2352 had both anthropometric measures and 2143 had additional household information available. Regression sample in columns 5-8 includes 833 non-poor children that had taken the TVIP cognitive ability test, of which 766 had both anthropometric measures and 731 had additional household information available. The nutritious food variable is the summation of the number of times the child consumed liver, cow internal organs, noodles or bread or other wheat flour products, spinach or chard, and carrots or similar vegetables in the previous week. Household Characteristics include a rural indicator, log wealth index and an indicator for whether or not the household received the cash transfer. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.4. Relationship between Investments above one Standard Deviation of Household Mean Nutritious Food and Child Ability, by Household Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Poor				Non-Poor		
Own Ability (TVIP age adjusted z-score)		0.018*	0.025**	0.022**		0.003	0.004	0.008
		[0.010]	[0.011]	[0.011]		[0.015]	[0.015]	[0.016]
Height-for-age Z-Score			-0.005	-0.006			0.020	0.026
			[0.006]	[0.006]			[0.020]	[0.018]
Weight-for-height Z-Score			-0.013**	-0.012*			-0.005	0.002
			[0.006]	[0.006]			[0.011]	[0.010]
Number of Siblings	-0.020***	-0.018**	-0.020***	-0.030***	-0.089***	-0.088***	-0.083***	-0.094***
	[0.007]	[0.007]	[0.008]	[0.009]	[0.020]	[0.020]	[0.023]	[0.024]
Number of Sisters	-0.007	-0.008	-0.008	-0.005	0.031	0.030	0.026	0.029
	[0.008]	[0.008]	[0.008]	[0.008]	[0.023]	[0.022]	[0.024]	[0.026]
Male	-0.004	-0.005	0.002	0.006	-0.037	-0.037	-0.030	-0.025
	[0.012]	[0.012]	[0.013]	[0.014]	[0.023]	[0.023]	[0.024]	[0.024]
Birth order	-0.047***	-0.048***	-0.050***	-0.036**	0.007	0.007	0.006	0.027
	[0.015]	[0.015]	[0.015]	[0.016]	[0.033]	[0.034]	[0.035]	[0.031]
Father's years of education				-0.001				-0.005
				[0.002]				[0.005]
Mother's years of education				0.002				-0.003
				[0.003]				[0.006]
Father is present in the home				-0.019				-0.207**
				[0.017]				[0.103]
Indigenous Identifier				0.015				-0.000
				[0.025]				[0.054]
Household Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,991	1,991	1,770	1,633	450	450	411	403

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All OLS regressions run with age and province fixed effects. Independent variable are from wave 1 of panel and dependent variables are from wave 2. Regression sample in columns 1-4 includes 1991 poor children that had taken the TVIP cognitive ability test, of which 1770 had both anthropometric measures and 1633 had additional household information available. Regression sample in columns 5-8 includes 450 non-poor children that had taken the TVIP cognitive ability test, of which 411 had both anthropometric measures and 403 had additional household information available. The nutritious food variable is an indicator for whether or not the child consumed more than 1 standard deviation above the nutritious food variable (the summation of the number of times the child consumed liver, cow internal organs, noodles/bread, spinach/chard, and carrots or similar in the previous week). Household Characteristics include a rural indicator, log wealth index and an indicator for whether or not the household received the cash transfer. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.5. Relationship between Investments and Child Ability Within the Household, by Household Wealth

	Nutritious Food				Greater than 1 standard deviation from household mean of Nutritious Food			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Poor		Non-Poor		Poor		Non-Poor	
Own Ability (TVIP age adjusted z-score)	0.466**	0.547**	-0.389	-0.232	0.038**	0.039*	-0.049	-0.025
	[0.182]	[0.217]	[0.458]	[0.532]	[0.019]	[0.022]	[0.040]	[0.044]
Height-for-age Z-Score		0.140		0.065		0.023*		0.035
		[0.141]		[0.296]		[0.013]		[0.031]
Weight-for-height Z-Score		0.101		0.301		-0.009		0.033
		[0.121]		[0.193]		[0.013]		[0.021]
Male	-0.019	-0.039	0.348	0.386	0.009	0.009	0.082	0.081*
	[0.188]	[0.237]	[0.407]	[0.373]	[0.017]	[0.020]	[0.052]	[0.047]
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	No	No	No	No	No	No	No	No
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	643	572	126	115	639	569	126	115
Number of Households	326	313	63	61	322	310	63	61

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All household fixed effects regressions run with age fixed effects. Independent variables are from wave 1 of the panel, while the dependent variable is from wave 2 of the panel. Regression sample limited to children in households with at least two young children between the ages of 4-7 in the follow-up survey and includes 769 children that had taken the TVIP cognitive ability test (643 poor children, 126 non-poor children), of which 687 had both anthropometric measures (572 poor children, 115 non poor children). The sample size in columns 5 to 8 is slightly smaller from a lack of variation in some households. The nutritious food variable is the summation of the number of times the child consumed each of 5 different categories of food in the previous week. These categories are liver, cow internal organs, noodles/bread, spinach/chard, and carrots or similar vegetables. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.6. Relationship between Investments, Child Ability and Sibling Ability, by Household Wealth

	Nutritious Food				Greater than 1 standard deviation from household mean of Nutritious Food			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Poor		Non-Poor		Poor		Non-Poor	
Average Sibling Ability (TVIP age adjusted z-score)	-0.627**		-0.100		-0.023**		0.039	
	[0.253]		[0.636]		[0.011]		[0.039]	
Highest Sibling Ability (TVIP age adjusted z-score)		-0.581**		-0.089		-0.023**		0.035
		[0.253]		[0.630]		[0.011]		[0.038]
Average Sibling Height-for-age Z-Score	0.012		0.026		-0.012		-0.034	
	[0.197]		[0.356]		[0.009]		[0.033]	
Average Sibling Weight-for-height Z-Score	0.419**		0.203		0.001		-0.029	
	[0.208]		[0.320]		[0.007]		[0.021]	
Highest Sibling Height-for-age Z-Score		-0.100		-0.005		-0.011		-0.045
		[0.187]		[0.340]		[0.007]		[0.032]
Highest Sibling Weight-for-height Z-Score		0.279		0.083		0.002		-0.031
		[0.194]		[0.275]		[0.007]		[0.019]
Own Ability (TVIP age adjusted z-score)	-0.064	-0.077	-0.159	-0.139	0.015	0.016	0.004	0.001
	[0.204]	[0.203]	[0.556]	[0.537]	[0.015]	[0.016]	[0.042]	[0.041]
Height-for-age Z-Score	0.311	0.358*	0.377	0.390	0.017**	0.017*	0.010	0.008
	[0.199]	[0.208]	[0.459]	[0.505]	[0.009]	[0.009]	[0.031]	[0.032]
Weight-for-height Z-Score	0.359**	0.406**	0.715*	0.736*	-0.002	-0.002	0.007	0.006
	[0.160]	[0.174]	[0.348]	[0.362]	[0.009]	[0.009]	[0.025]	[0.024]
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	512	512	109	109	510	510	109	109

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All OLS regressions run with age and province fixed effects. Independent variables are from wave 1 of the panel, while the dependent variable is from wave 2. Regression sample includes 655 children from households that had taken the TVIP cognitive ability test and had at least one sibling between the ages of 4-7 in the follow-up survey that also had taken the TVIP cognitive ability test, of which 540 were from poor households and 115 are from non-poor households. The sample size in columns 5 to 8 is slightly smaller from a lack of variation in some households. The nutritious food variable is the summation of the number of times the child consumed each of 5 different categories of food in the previous week. These categories are liver, cow internal organs, noodles or bread or other wheat flour products, spinach or chard, and carrots or similar vegetables. Individual Characteristics include number of siblings, number of sisters, gender and birthorder. Household Characteristics include father's years of education, mother's years of education, father's presence in the home, an indigenous identifier, a rural indicator, log wealth index and an indicator for whether or not the household received the cash transfer. Data source: Ecuador's BDH data from 2003-2005.

Table 1.7. Relationship between Non-nutritious Food Consumption and Child Ability Within the Household, by Household Wealth

	Non-nutritious Food		Greater than 1 standard deviation from household mean of Non-nutritious Food	
	(1)	(2)	(3)	(4)
	Poor	Non-Poor	Poor	Non-Poor
Own Ability (TVIP age adjusted z-score)	-0.069	0.078	0.015	0.033
	[0.260]	[0.330]	[0.025]	[0.024]
Height-for-age Z-Score	0.148	-0.387	0.009	-0.055
	[0.236]	[0.306]	[0.021]	[0.046]
Weight-for-height Z-Score	-0.172	-0.151	-0.017	-0.001
	[0.200]	[0.312]	[0.020]	[0.037]
Male	-0.397	0.432	-0.012	0.006
	[0.444]	[0.606]	[0.031]	[0.059]
Age FE	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes
Observations	572	115	569	115
Number of Households	313	61	310	61

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All household fixed effects regressions run with age fixed effects. Independent variables are from wave 1 of the panel, while the dependent variable is from wave 2 of the panel. Regression sample includes 687 children that had taken the TVIP cognitive ability test and had both anthropometric measures (572 poor children, 115 non-poor children). The sample size in column 3 is slightly smaller from a lack of variation in some poor households. The non-nutritious food variable is the summation of the number of times the child consumed foods from four different categories in the previous week. These categories are soda or ice cream, cookies or pastries, fried snacks, and candies. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Table 1.8. Relationship between an Alternative Measure of Nutritious Food Investments and Child Ability, by Household Wealth

	(1)	(2)	(3)	(4)	(5)	(6)
		Poor			Non-poor	
Average Sibling Ability (TVIP age adjusted z-score)		-0.016*			0.026	
		[0.009]			[0.030]	
Highest Sibling Ability (TVIP age adjusted z-score)			-0.016*			0.026
			[0.009]			[0.030]
Own Ability (TVIP age adjusted z-score)	0.027	0.012	0.012	-0.002	0.008	0.008
	[0.021]	[0.015]	[0.015]	[0.057]	[0.036]	[0.036]
Height-for-age Z-Score	0.016	0.011	0.011	-0.008	-0.006	-0.006
	[0.016]	[0.009]	[0.009]	[0.046]	[0.022]	[0.022]
Weight-for-height Z-Score	0.009	-0.003	-0.003	-0.007	0.020	0.020
	[0.010]	[0.005]	[0.005]	[0.035]	[0.018]	[0.018]
Number of Siblings		-0.010	-0.010		-0.012	-0.012
		[0.009]	[0.009]		[0.033]	[0.033]
Number of Sisters		0.007	0.007		-0.010	-0.010
		[0.010]	[0.010]		[0.039]	[0.039]
Male	-0.009	-0.003	-0.003	0.029	0.038	0.038
	[0.024]	[0.016]	[0.016]	[0.048]	[0.045]	[0.045]
Birth order		-0.006	-0.006		-0.110**	-0.110**
		[0.021]	[0.021]		[0.047]	[0.047]
Father's years of education		-0.000	-0.000		-0.001	-0.001
		[0.003]	[0.003]		[0.007]	[0.007]
Mother's years of education		-0.000	-0.000		0.002	0.002
		[0.002]	[0.002]		[0.009]	[0.009]
Father is present in the home		0.018	0.018		-0.082	-0.082
		[0.017]	[0.017]		[0.096]	[0.096]
Household Characteristics	No	Yes	Yes	No	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	No	Yes	Yes	No	Yes	Yes
Household Fixed Effects	Yes	No	No	Yes	No	No
Observations	569	537	537	115	115	115
Number of Households	310			61		

Notes: Robust standard errors in brackets, clustered at parish level. * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions run with age fixed effects. The regressions in columns 1 and 4 are run with household fixed effects, while the regressions in columns 2, 3, 5 and 6 are run with province fixed effects. Independent variable is from wave 1 of the panel, while the dependent variable is from wave 2 of the panel. Regression sample includes 684 children that had taken the TVIP cognitive ability test and had both anthropometric measures (569 poor children, 115 non-poor children), of which 652 had additional household information available (537 poor children, 115 non poor children). The nutritious food variable is an indicator of whether or not the child ate more than one standard deviation above the household mean of the summation of the number of times the child consumed each of 2 different categories of food in the previous week. These categories are liver and spinach or chard. Household Characteristics include an indigenous identifier, a rural indicator, log wealth index and an indicator for whether or not the household received the cash transfer. Data source: Ecuador's Bono de Desarrollo Humano (BDH) data from 2003-2005.

Figure 1.1a. Graph of the average number of times nutritious food was eaten in the past week by number of siblings and by household wealth, full sample

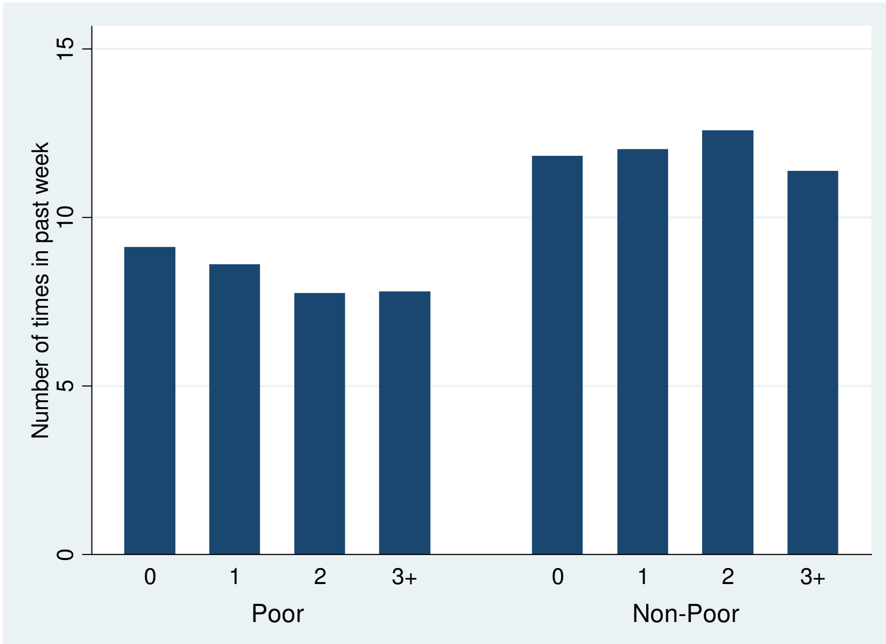
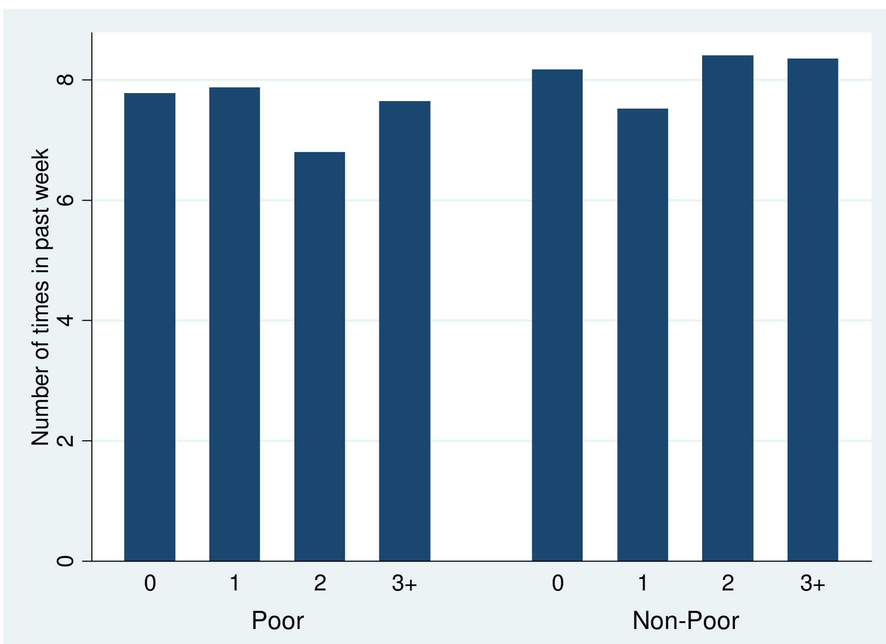


Figure 1.1b. Graph of the average number of times non-nutritious food was eaten in the past week by number of siblings and by household wealth, full sample



CHAPTER 2
CHILD ABILITY AND HOUSEHOLD HUMAN CAPITAL INVESTMENT DECISIONS IN BURKINA FASO[20]

1. Introduction

Parental decisions about whether and how much to invest in their children's human capital depend on many factors, and these decisions have long-lasting impacts on each child's future earnings, marital prospects, and overall welfare. A large literature attempts to understand the source of inequalities for children's educational investments within a household building on seminal work by Becker and Tomes (1976) that delineates the tradeoff between the quantity of children and their 'quality.' In making the schooling investment decision, parents will have information about a child's ability and that information will often not be available to researchers, which partly explains why much of the empirical research on the determinants of household investments in children's schooling focuses on easy to observe demographic characteristics of the child such as gender, birth order, and family composition (Parish and Willis 1993; Garg and Morduch 1998; Black, Devereux, and Salvanes 2005).[21] More recent papers attempt to use direct measurements of a child's ability such as IQ scores (Kim 2005) or cognitive tests (Ayalew 2005) to better understand which factors influence investment decisions.

In this paper, we build on the seminal empirical work by Rosenzweig and Schultz (1982) to examine the role that a child's cognitive ability plays in a resource constrained household's decision to invest in that child's education. For poor households seeking to maximize the returns to their human capital investments, schooling decisions will depend on parent perceptions about the returns to school for a given child and that child's ability. In a setting where few households ever enroll all

[20] This work is co-authored with Richard Akresh, Damien de Walque and Harounan Kazianga

[21] See Strauss and Thomas (1995) and Glewwe and Kremer (2006) for reviews of the literature. Related research explores the relationship between these demographic characteristics and the non-schooling outcomes of employment (Kessler 1991), risky behaviors (Aizer 2004), and child labor (Emerson and Souza 2008).

of their children in school, as is true in many developing countries, understanding the link between child ability and school enrollment and school continuation decisions is critical for developing policy prescriptions to improve educational outcomes.

We make four main contributions to the literature on explaining household school investment decisions. First, we employ direct measures of a child's ability for all children of primary school age (5 to 15), regardless of whether they are currently enrolled in school. This differs from existing papers that tend to have ability measures only for children that are currently enrolled in school (Glick and Sahn 2010). We use the Raven's Colored Progressive Matrices (CPM) and the Weschler Intelligence Scales (WISC) Digit Span as measures of a child's cognitive ability. Second, our paper is unique in explicitly incorporating direct measures of the ability of each child's siblings (both absolute and relative measures) and to show how sibling ability 'rivalry' exerts a strong impact on the parents' decision of which child to send to school. Third, the survey instrument asks parents to provide their perceptions about the likely chances of future economic success for each of their children, information that is not often gathered in surveys. We show that a similar pattern of sibling rivalry is observed using either these parent perceptions or the externally validated cognitive ability tests measuring child ability. Fourth, we address potential concerns about schooling influencing measures of child ability by exploiting the panel data structure and focusing on the relationship between the enrollment decision in the survey's second year and the ability measures observed in the survey's first year for the subset of young children who were not enrolled and not yet of typical school age in the first year.

We explore both the extensive margin of school enrollment during the 2007-2008 school year and grade progression measures, as well as the intensive margin of school related expenses. We find that a child with a one standard deviation higher ability test score has a 16 percent higher likelihood of being currently enrolled in school, while a child with a higher ability sibling is 16 percent less likely to be currently enrolled and having two higher ability siblings lowers a child's probability of enrollment by 30 percent. Household fixed effects regressions show that within a given household, a child with one standard deviation higher ability compared to the average ability of their siblings is 31 percent more likely to be enrolled. On the intensive margin, controlling for household fixed effects, we find that a child with one standard deviation higher ability receives 20 percent more discretionary school expenditures by the parents.

The remainder of the paper is organized as follows. Section 2 discusses the conceptual framework about sibling rivalry and the household schooling investment decision. Section 3 describes the survey data used in the analysis and explains the construction of the different child ability measures. Section 4 describes the empirical identification strategy and section 5 presents the main results as well as robustness tests. Section 6 concludes.

2. Sibling Rivalry Conceptual Framework

A number of studies examine the interaction of siblings to understand schooling outcomes and why girls often receive less education than their brothers. Butcher and Case (1994), using United States data that contains explicit information on an individual's completed education and the education of their brothers and sisters, find that women with only brothers receive significantly more education on average than women with any sisters.[22] Their finding differs from what is typically found in developing country studies. Parish and Willis (1993) examine how sibling sex composition influences girls' education in Taiwan. They emphasize that cultural traditions favoring male descent can cause parents to manipulate daughters for the benefit of their sons. Garg and Morduch (1998) emphasize sibling rivalry using data from Ghana. Child education decisions in credit constrained households are

[22] They highlight three potential explanations for why sibling sex composition might influence education decisions: sibling resource competition, sex-typing of tasks, and peer effects. Resource competition occurs if boys and girls have different relative prices for educational investments or investment returns. Sex-typing stems from parents sending messages to children describing appropriate behaviors and goals, while peer effects come from children developing traits that depend on how they interact with their siblings.

influenced by the number of children they have, resource dilution, and the sex composition of their children, sibling rivalry. Resource dilution occurs because more children imply fewer resources per child and credit constraints limit the family's ability to borrow against future returns. Sibling rivalry occurs because all children benefit from having fewer educated siblings with comparatively higher returns on investment.

Resource dilution and sibling rivalry in educational investments in poor countries is well documented. In addition to the Taiwan and Ghana studies, the list includes: Binder (1998) for Mexico, Morduch (2000) for Tanzania, Edmonds (2007) for Nepal, Ota and Moffatt (2007) for India, and Dammert (2010) for Guatemala and Nicaragua. While Garg and Morduch focus on credit constraints and differences in relative returns to education as the cause for sibling rivalry, Edmonds (2007) emphasizes that comparative advantage in home production can lead to similar implications when girls have comparative advantage and it is not possible to hire labor for home production. Both Edmonds (2007) and Dammert (2010) find evidence consistent with this sibling rivalry interpretation.

Such models of sibling rivalry neglect that parents have additional knowledge about their children's capabilities and use this information to make school investment decisions. A literature embedded in testing the one-period consensus parental preferences model of human capital investment of Becker and Tomes (1976) and Behrman, Pollack, and Taubman (1982) uses child endowments in modeling the investment decision.[23] Most studies that examine the investment decision process have to work around the fact that actual child ability or endowment is typically not observed (Rosenzweig and Schultz 1982; Behrman, Pollack and Taubman 1982; Behrman, Rosenzweig and Taubman 1994). Some recent studies are able to use direct measures of child ability. Kim (2005) uses an IQ test administered to Wisconsin high school juniors and finds that higher

[23] See Behrman (1997) for an overview of the consensus parental preferences models.

ability children receive more parent transfers. Glick and Sahn (2010) use achievement test scores from Senegalese children taken in Grade 2 to explain school outcomes seven years later, but this neglects the role of siblings that may not have been enrolled in school.

The most closely related paper to ours is by Ayalew (2005) who uses Raven's CPM test scores for school-age children in one village in Ethiopia to measure child ability and using a household fixed effects model finds that parents consider child ability when making school enrollment decisions. There are several key differences between our papers. First, we focus on absolute and relative direct measures of the ability of a child's siblings to generate inferences about the role of sibling rivalry in influencing schooling decisions. Second, we explore both alternative ability measures by using different cognitive tests and alternative outcomes such as school expenditures and grade progression, in addition to current enrollment, which is the focus of the Ayalew paper. Third, we exploit the panel data structure as a robustness check to address potential reverse causality concerns about schooling influencing measures of child ability.

3. Burkina Faso Social Protection Evaluation Survey

The panel survey was conducted in June 2008 (Year 1) and June 2009 (Year 2) in Nahouri province in southern Burkina Faso, located approximately 100 miles from the capital and bordering Ghana. Households were randomly selected from a village-level census conducted by our project team immediately prior to the Round 1 survey in the 75 rural villages of Nahouri province that each has a primary school. The survey is part of an ongoing project evaluating social protection strategies in Burkina Faso. Households in this region are predominantly subsistence farmers growing sorghum and groundnuts and have mean annual per capita expenditures of approximately $90.

Our analysis focuses on primary school-aged children ages 5 to 15 in households with multiple biological children. There are 4,635 children in this age range in 1,507 different households. As shown in Table 2.1, parental schooling is low, with only 13 percent of the children having a parent that ever attended school. Fifty-four percent of this children's sample is male and the average age is 9.4 years old. On average, these children have 3.8 siblings under age 15, including 1.8 sisters. They live in households with an average of 8.9 individuals, including a head of household, 1.5 wives, 4.8 biological children of the household head under age 15, 0.4 children under age 15 that are not the biological children of the head, and 1.2 other members that include grandparents, aunts, uncles, and other extended family members.

Parents were directly asked about the chance of future success they believe each of their children will have in formal employment, a reasonable measure of parental perceptions about the investment return on their child's education, since most jobs in "formal employment" in Burkina Faso require a level of education beyond primary school and in particular French skills.[24] This parental perception measure is based on everything the parent knows about the child and about the labor market, whether right or wrong, and was asked about every child in the household. For each child, the parents responded whether that particular child had a 'small', 'medium', 'large', or 'very large' chance of future success in formal employment. Parents considered 25 percent of these children to have a 'small' chance of future success and only 8 percent to have a 'very large' chance of future success. Parents viewed most children (67 percent) to have a 'medium' (38 percent) or 'large' (29 percent) chance of future success.

To corroborate these parent perceptions, we also consider externally validated measures about a child. We use the Raven's CPM and the WISC Digit Span to measure a child's cognitive ability; both are tests that do not require formal schooling to be able to answer the questions. The Raven's CPM is a measure of fluid intelligence or problem solving ability. The test does not depend heavily on verbal skills, making it relatively "culture free" (Borghans, Duckworth, Heckman, and

[24] Schultz (2004) and Kazianga (2004) using Burkina Faso nationally representative data report substantial returns (9 to 16 percent) for an additional year of primary school, highlighting the importance of schooling in this context. Returns to secondary (14 to 26 percent) and tertiary (13 to 23 percent) schooling are even higher.

Weel 2008). In Figure 2.1 Panel A, we show two sample problems from the Raven's test (Raven, Raven, and Court 1998). The child respondent is asked to select the image that is missing in order to complete the picture. This type of question is novel to the children in Nahouri Province, thus providing a more natural or true measure of problem solving skills.

We ask 18 questions from the Raven's CPM and on average, children in our sample answer 4.9 questions correctly.[25] Younger children answer fewer questions correctly than older children (the average number correct for children age 5 is 2.8 and for children age 15 is 7.6).[26] To control for this relationship between age and raw test scores, we calculate a z-score for each child measured as the child's raw test score minus the average score for the same age children divided by the standard deviation of test scores for children of that age.[27] Therefore, the mean of the Raven's z-score is zero and the standard deviation is one for each age and across all ages.[28]

The WISC Digit Span is a measure of working memory and ability to concentrate and has both a forward and backward component. The respondent repeats a string of numbers to the enumerator and is scored by whether or not they repeat the full string correctly as shown in Figure 2.1 Panel B (Weschler 1974). In the Digit Span Forward, the child must repeat the string of numbers exactly as stated by the enumerator. The string of numbers increases in length as the child answers correctly. With the Digit Span Backward, similar strings of numbers are to be repeated in the reverse order from that stated by the enumerator until the child can no longer continue. We calculate a total combined score of the forward and backward digit spans, and the children have an average score of

[25] During extensive pretesting of the Raven's test, results were consistent whether children were asked the entire set of 36 questions or only the odd-numbered questions, so to save interview time we only administered the 18 odd-numbered questions (Sets A, Ab, and B).
[26] The average number of questions answered correctly for children ages 6, 7, 8, 9, 10, 11, 12, 13, and 14 is respectively 2.8, 3.6, 4.4, 5.1, 5.3, 5.6, 6.1, 6.5, and 6.4.
[27] We did not use the international Raven's norming standards since we asked a subset of the Raven's test and what is most important here is how the children in rural Burkina Faso compare to each other, not internationally.
[28] Note that in Section 5.2, we estimate alternative specifications to test the robustness of using the Raven's age-adjusted z-score instead of the raw test scores.

7.6 correct answers out of a total possible 32.[29] As with the Raven, we calculate a WISC Digit Span age-adjusted z-score to control for age effects.

In Table 2.1, we present summary statistics about children's schooling. Few households in rural Burkina Faso ever enroll all of their children. Only 54 percent of children are enrolled in the 2007-2008 school year. Fifty-six percent of households experience variation in enrollment among their children age 5 to 15, while 17 percent enroll no children and only 27 percent of households currently enroll all of their primary school-aged children. If we consider whether a child has ever been enrolled in school rather than current enrollment during 2007-2008, then 59 percent of children in the sample have ever been enrolled and 54 percent of households experience variation across their children in whether a child has ever been enrolled. Given these low enrollment rates, on average these children only have completed 1.8 years of school. Child labor rates are high, with children heavily involved in agriculture and household production.

In addition to examining the relationship between parent perceptions, child ability and school enrollment, we explore three alternative schooling-related outcomes (on-time start, grade progression, and discretionary school expenses) where sibling rivalry might matter. In Burkina Faso, parents typically enroll their children starting at age seven, so we construct a variable to indicate if children started school by this age or if they were delayed. The 'on-time start' variable shows that only 40 percent of primary school-aged children start school by age seven, with the rest either starting at a later age or never attending school. Fifty-four percent of households have variation across their children in whether each child started school by age seven. Second, we consider grade progression through school, which we calculate by dividing the child's highest grade attended by the number of years since the child started school.[30] The grade progression measure ranges from zero to one, with higher numbers indicating quicker progress towards completing primary school. Third, for

[29] Our regression results are robust to keeping the forward and backward digit span scores separate.
[30] For children who never attended school, they are assigned a grade progression measure of zero.

each child we calculate the total schooling-related discretionary expenditures during the 2007-2008 school year. While school in Burkina Faso has relatively low registration fees (904 FCFA on average per year, about $2.18 using the June 2008 exchange rate of 415 FCFA = $1 USD), there are additional expenditures expected of each family when they enroll their child. These can include purchasing uniforms, contributions for the school cafeteria, and transportation costs. Spreading resources evenly across children to pay the fixed costs associated with schooling may not be possible in the presence of liquidity constraints. However, for some of these expenditures, such as school supplies and parent association fees, parents have discretion in the amount spent each year on a child who is enrolled, and these discretionary expenditures are important in this setting as well as in developing countries that currently have free schooling. In our sample, the total average cost of sending a child to school is 3867 FCFA per school year (about $9.32), with parents spending on average 845 FCFA (about $2.04) on these discretionary items per child (about 22 percent of total educational expenses).

4. Empirical Identification Strategy

4.1 Econometric Specification

Studies of sibling rivalry in education typically use counts of the number of siblings and the number of sisters that a child has to explain different schooling outcomes (attendance, enrollment, attainment) as follows:

(1) $$e_{ih} = \omega_0 S_{ih} + \omega_1 F_{ih} + \alpha_0 X_{ih} + \alpha_1 Z_h + \varepsilon_{ih}$$

where e_{ih} is the educational outcome for child i in household h, S_{ih} is a count of the number of siblings the child has, F_{ih} is a count of the number of female siblings the child has, X_{ih} is a vector of individual characteristics such as age and gender that might influence parental investments, Z_h is a vector of household characteristics, and ε_{ih} is a random, idiosyncratic error term. The interpretation of ω_0 is the change in e_{ih} associated with an additional male sibling. The interpretation of ω_1 is the change in e_{ih} associated with the thought experiment of converting a sibling from a male to a female. $\omega_0 + \omega_1$ is then the change in e_{ih} associated with adding an additional female sibling. This approach takes current family size and composition as given at the time the parents make the enrollment decision.

To better understand the parental schooling investment decision, we expand on the sibling rivalry model in Equation 1 to control for previously unobserved characteristics about the child (his ability) and his home environment (his siblings' ability) that might influence the parent's decision. We employ two empirical approaches to estimate this relationship. First, we estimate the following household or sibling fixed effects logit regression that will control for all household level characteristics that are constant across siblings:

(2) $$e_{ih} = \beta_0 A_{ih} + \alpha_0 X_{ih} + \lambda_h + \eta_{ih}$$

where e_{ih} and X_{ih} are defined as above, A_{ih} is a direct measure of observed child ability, λ_h is the household fixed effect that captures all characteristics about the household that are constant across siblings, and η_{ih} is the child specific idiosyncratic error term. In Equation 1 and previous sibling rivalry papers, child ability was part of the error term, ε_{ih}, but in our analysis we are able to directly control for its effect on educational outcomes.[31] This within family estimate compares a child's own ability to the average ability of all the other children in the household to examine if parents compare a child's ability to the average ability of his siblings when making human capital investment decisions.

While the household fixed effects estimation compares own ability to average sibling ability, the second approach we adopt is to be more specific about the functional form of the sibling ability term and to include direct measures of sibling ability in the regression. This approach has the

[31] Note that in the household fixed effects specification, household characteristics, Z_h, and number of siblings, S_{ih}, will drop out of the specification because there is no variation across children within the household. In the household fixed effects regressions, we also drop the number of sisters variable, F_{ih}, because it is constant within a given household for children of the same gender.

additional advantage that we can include the same variables as in the sibling rivalry literature (in Equation 1) and allows us to examine how the relevant coefficients vary when controlling for a child's own ability and his sibling's ability. We estimate the following extended Equation 1 sibling rivalry regression:

(3) $$e_{ih} = \beta_0 A_{ih} + \beta_1 h(A_{-ih}) + \omega_0 S_{ih} + \omega_1 F_{ih} + \alpha_0 X_{ih} + \alpha_1 Z_h + \mu_{ih}$$

where $h(A_{-ih})$ is a measure of the ability of the other children *(-i)* in household *h* with varying functional forms that we discuss in detail below, and the other variables are as defined above. The error term, μ_{ih}, measures the child specific idiosyncratic part of ε_{ih} not captured by a child's own ability, A_{ih}, or his sibling's ability, $h(A_{-ih})$. The coefficients β_0 and β_1 respectively give an estimate of the impact of child *i*'s own ability and his sibling's ability on child *i*'s enrollment.

We use several alternative measures of sibling ability, $h(A_{-ih})$, including both absolute measures (highest sibling ability) and relative measures (whether there are any siblings with a higher ability score and dummies for the number of siblings with higher ability scores). Absolute measures provide insight into the role of the level of sibling ability in a household. Having siblings with high ability might raise overall enrollment levels in a family, or it might represent competition for the child. It could be that the average level of sibling ability affects a child's enrollment differently than the ability level of the household's 'best' sibling (with the highest ability). If sibling rivalry influences parents deciding who to send to school, then parents might consider how a child compares in ability to his siblings rather than considering the child's ability on its own, and relative sibling ability measures might be more informative. In our sample, 40 percent of the overall variation in ability arises from within family variation across siblings, while 60 percent is between families.

4.2 Potential Threats to Identification Strategy

Since schooling potentially affects cognitive ability, reverse causality is the primary problem we face. We attempt to address this in two ways. First, we estimate robustness specifications in which we limit the sample of children to Grades 2 and lower or to Grade 1 and lower. The decision to use this grade cutoff point is based on a regression of the Raven's age adjusted z-score on grade in school, and the coefficients for Grades 1 and 2 are close to zero (0.05 and 0.09 respectively) and not statistically significant. The coefficients for Grade 3 and 4 are slightly larger (0.14 and 0.12 respectively), but only the Grade 3 coefficient is statistically significant, while the Grade 4 coefficient is not statistically significant. We interpret this lack of relationship between the lower grades and ability test scores as evidence that children in Grade 2 and lower have not yet received enough schooling to influence their cognitive ability test scores. Based on this information and to be conservative in our robustness specifications, we select Grades 1 and 2 as the cutoff levels. Second, we restrict the sample to young children ages 5 to 7 (and 5 to 6) who are not enrolled in year 1 but for whom we have ability measures in year 1 and look at their enrollment in year 2. This eliminates any potential effect of schooling on the ability measures as these children were not enrolled at the time of taking the ability test.

5. Empirical Results

5.1 Sibling Rivalry, Parent Perceptions, and Child Ability

Since we are building upon the sibling rivalry literature, we begin our analysis estimating Equation 1 that uses the standard observable family composition characteristics, number of siblings, number of sisters, and birth order. Results of this regression are presented in Table 2.2, column 1. We find evidence of resource dilution and sibling rivalry consistent with the literature. We find that the number of siblings has a negative effect on enrollment (resource dilution) while the number of sisters has a positive effect (sibling rivalry). Holding constant the number of sisters, the addition of a male sibling is correlated with 2.5 percentage points (or 4.6 percent) lower likelihood of attending school. An additional female sibling has no impact on whether the child attends school. Subsequently, switching from a male to female sibling corresponds to a 2.2 percentage point higher likelihood of enrollment, or 4.1 percent of the base enrollment level. Birth order has a positive but not statistically significant coefficient indicating younger siblings are more likely to be enrolled, as is consistent with the literature in developing countries.

As discussed in Section 4.1, other factors about the child besides these observable demographic characteristics are likely to influence the parent's schooling investment decision. Parents know more about their children's characteristics than simply their gender and sibling composition. Since it is the parents' perceptions of their child's ability or potential for future success, whether correct or not, that informs and affects their decision about educational investment, we first examine the relationship between school enrollment and these parents' perceptions about each of their children.[32] We estimate both a household fixed effects logit as in Equation 2 and an extended sibling rivalry logit regression as in Equation 3, and in columns 2 to 5 of Table 2.2 we present the corresponding results.[33] We find a positive relationship between what parents think about a child and his current enrollment. On the other hand, perceptions of the child's siblings in the same age group have a negative relationship with the child's enrollment, suggesting parents make educational investment decisions based not only on what they think of one child but also what they think of that child's siblings.

The household fixed effects specification presented in Table 2.2 column 2 shows that children with one level higher parental perceptions compared to the average perceptions of their siblings have an 18.4 percentage point higher probability of enrollment, which corresponds to a 34.1 percent higher enrollment level. In columns 3 to 5, we explicitly estimate the relationship between parent perceptions about the child's siblings and a child's enrollment. Controlling for direct

---

[32] The parent perception variable takes values of 0 to 3, where 0 means a child has a small chance of future success, 1 a medium chance, 2 a large chance, and 3 a very large chance. Parents on average report that their children have a medium chance of success, with the variable having a mean of 1.2 and a standard deviation of 0.9.

[33] Parent perceptions may be influenced by the child's enrollment status, and therefore the results presented in this table should not necessarily be interpreted as causal.

measures of parent perceptions of siblings, the parental perceptions of the child are still positively correlated with the child's enrollment and statistically significant at the one percent level. One level higher parent perceptions is correlated with a 10.9 to 16.2 percentage point higher likelihood of being enrolled. However, parental perceptions of a child's siblings are negatively correlated with the child's enrollment status. Compared to the household fixed effects specification in which the parental perceptions of the child are compared to the average of his siblings, an alternative is to make the comparison with the parental perceptions of the 'best' sibling. Results in column 3 show that children whose sibling with the highest perception in the family has a one level higher value have a 6.5 percentage point lower likelihood of enrollment.

Relative sibling perceptions might be more relevant than absolute sibling perceptions since it is possible that having a sibling whom the parents think of more highly than oneself matters more than the overall perception level of one's siblings. Column 4 uses an indicator of whether the child has any sibling with parental perception higher than himself while column 5 uses indicators for whether the child has one, two, or three or more siblings with higher parental perceptions. Children having any sibling with better parent perceptions have a 7.0 percentage point lower probability of enrollment. Children with three or more siblings with higher parental perceptions have a 14.6 percentage point lower probability of being enrolled, corresponding to a 27 percent lower enrollment, and the coefficient is significant at the five percent level.

While the relationship between parental perceptions of a child and his schooling is strong, it does not eliminate the role of sibling composition. Having more brothers is correlated with lower enrollment, while having more sisters instead of brothers and holding the number of siblings constant is correlated with higher enrollment. Birth order is also important; younger siblings have a 2.3 to 2.5 percentage point higher probability of being enrolled (columns 3 to 5). Consistent with inter-generational education transmission and wealth effects, better educated parents and wealthier households have children who are more likely to be enrolled.

While parental perceptions about their child's chance of future success are correlated with the child's current school enrollment, these perceptions may or may not be accurate or well-informed. There may also be significant differences across households in how parents perceive their own children and what factors they take into account in formulating perceptions. To further explore these issues, we incorporate an externally validated measure of the child's cognitive ability using the Raven's CPM test. These tests were administered during the baseline survey to every child age 5 to 15 regardless of their current enrollment status and provide a consistent measure of child ability across children in all households. There is a strong positive relationship between the ability measure and parent perceptions. Higher ability children are viewed by their parents to have a higher chance of future success. However, after controlling for gender and age, the ability measure only explains about 20 percent of the variation in parental perceptions.

In Table 2.3, we estimate the relationship between child ability (as measured by the Raven's age adjusted z-score) and current school enrollment using a household fixed effects logit as described in Equation 2 and a logit regression with alternative sibling ability measures as described in Equation 3.[34] The household fixed effects logit results in column 1 indicate that a child with one standard deviation higher own ability compared to the average of his siblings has a 16.5 percentage point higher likelihood of being currently enrolled, corresponding to 30.6 percent of the base enrollment. The coefficient is significant at the one percent level. This is evidence parents take into account a child's cognitive ability in deciding enrollment, and the magnitude of the effect is large.

[34] All regressions include child gender and age dummies, and the regressions estimating Equation 3 also include village fixed effects, parent schooling, household assets, and family demographic composition measures. Results presented in Table 3 are consistent when using the number of siblings and the number of sisters age 5 to 15 rather than the number of siblings and sisters age 0 to 15. Correlation among the error terms of children in a given village experiencing the same enrollment environment might bias the standard errors downward, so in all regressions we cluster the standard errors by village.

When considering how parents make this enrollment decision, one approach would be for them to compare a child's ability with the average ability of his siblings, and this is captured in the household fixed effects specification. An alternative that takes into account the non-linear relationship between siblings' abilities would consider the impact of the sibling with the highest ability. Another approach would include relative measures indicating if the child has any sibling with a higher ability measure or whether the child has one, two, or three or more siblings with higher ability measures.[35] Controlling for these direct measures of sibling ability (in columns 2 to 4), the child's own ability is still positively correlated with the child's enrollment and statistically significant at the one percent level. One standard deviation higher own ability is correlated with 15.7 to 27.2 percent higher likelihood of enrollment compared to the base enrollment level. Having one's 'best' sibling have a one standard deviation higher ability is correlated with a 6.8 percentage point lower enrollment rate (column 2), and the coefficient is significant at the one percent level. Likewise, having any sibling with a higher ability is correlated with 11 percentage points lower likelihood of being enrolled (column 3), and this effect is magnified if there are two siblings with higher abilities (16.1 percentage points). Both coefficients are significant at the one percent level.

Including child ability and sibling ability measures does not significantly alter the family demographic composition variables. The sign and level of statistical significance are consistent with the initial regression presented in Table 2.2 column 1, while the magnitude of the coefficient for the number of siblings and number of sisters is somewhat reduced. It is worth noting that the relationship between a child's own ability and current school enrollment is four to eight times larger than the corresponding relationship between the standard demographic composition variables and enrollment. These sibling ability rivalry results are consistent with the parental perceptions

[35] Results are robust to additional sibling ability measures including median sibling ability, the number of siblings with a higher ability, dummies for whether a child's ability is highest or lowest in the household, and whether the child has any siblings who have ability measures one-half or one standard deviation higher.

regressions in Table 2.2 and indicate that part of what is driving the relationship between parental perceptions and the school enrollment decision is the child's ability.[36]

Having explored the relationship between child ability and the extensive margin of school enrollment, we next turn to the intensive margin of educational expenditures. This allows us to rule out the interpretation that the results presented in Tables 2.2 and 2.3 reflect solely the desire of the child to attend school. If higher ability children have more motivation to attend school and this reduces the parents' cost of effort to make the child attend, then we would not expect the parents to make additional discretionary monetary investments in these children. We focus on expenses for school supplies and parent association voluntary fees because these have a discretionary component, whereby parents have some leeway in how much they spend on each of their children.[37] For the regressions presented in Table 2.4, we restrict the sample of 4,635 children age 5 to 15 living in households with multiple siblings to only the 2,511 children who are currently enrolled in school. We estimate a similar series of regressions as in Table 2.3 (household fixed effects in column 1 and then including alternative sibling ability measures in columns 2 to 4). Results in column 1 indicate that within a given household, children with a one standard deviation higher ability receive 170 FCFA more in discretionary expenditures, representing 20.1 percent of mean discretionary expenses, and the coefficient is significant at the one percent level. Controlling directly for alternative functional forms of sibling ability in columns 2 to 4 does not alter the positive relationship between a child's own ability and educational expenses, with coefficients ranging from 112 to 139 FCFA.

---

[36] We also estimate the regressions separately by child gender and find no strong gender difference. Sibling rivalry appears to be more important for girls than boys, but we cannot reject the equality of coefficients between the genders. Similarly we cannot reject that the role of own ability or parent perceptions are the same for both genders. We also estimate the regressions broken down by poverty level, defining poor households to have log assets below the mean, below the median, or in the bottom quintile, and while the estimates for poor families are larger, we cannot reject that poor and non-poor families have the same level of sibling rivalry.

[37] School registration fees are not considered since all enrolled children have to pay the same fees. School meal fees, lodging fees, uniforms, and transportation expenses are the other educational expenses that are not included as these have much less variation across siblings within a household.

Finally, children with two siblings of higher ability have 136 FCFA lower educational expenditures, corresponding to 16.1 percent of discretionary educational expenses.

5.2 Robustness Checks

To test the robustness of our results, we present four tables of regressions where we explore different educational outcomes, use two approaches to address potential reverse causality issues between schooling and cognitive ability, and use alternative cognitive tests to measure child ability. First, in Table 2.5, we present results for alternative schooling outcomes including ever enrolled in school, on-time school start, and grade progression through school. Results are consistent with those in Table 2.3 for current enrollment. We use household fixed effects as well as the relative measure of whether a child has one, two, or three or more siblings of higher ability.[38] Relative to the base levels, in the household fixed effects specifications (columns 1, 3, and 5), children with one standard deviation higher own ability are 29.5 percent more likely to be ever enrolled, 39 percent more likely to start school on time, and 28.1 percent more likely to progress through school. Children with one sibling of higher ability have lower probability of these outcomes (12 percent lower level of ever being enrolled, 11 percent lower level of starting school on time, and 8 percent lower level of grade progression). Negative effects are larger for children who have two siblings of higher ability (27 percent lower level of ever being enrolled, 24 percent lower level of starting school on time, and 20 percent lower level of grade progression).

Second, in Table 2.6, we attempt to address the potential reverse causality of schooling affecting a child's cognitive ability by limiting the regression sample to children that are in Grade 2, Grade 1, or not enrolled (columns 1 to 4) and children in Grade 1 or not enrolled (columns 5 to 8) because the regression evidence discussed previously indicates that children in these grades have not yet received enough schooling to influence their cognitive ability test scores. Results for this

---

[38] We also estimate regressions including the highest sibling ability and whether the child has any sibling of higher ability and find consistent results, but due to space limitations we present the limited set of results.

restricted sample are consistent with those in Table 2.3. Household fixed effects logit regressions in columns 1 and 5 indicate that within a given household, relative to the base enrollment levels, a child with one standard deviation higher ability is respectively 33 and 36 percent more likely to be enrolled.[39] Children with two siblings of higher ability have a 6.2 or 3.9 percentage point lower probability of enrollment (columns 4 and 8 respectively), corresponding to 17.7 and 17.0 percent of the base level of enrollment.

Third, in Table 2.7, we further address any potential reverse causality between schooling and cognitive ability by using the ability measure of young children who are not enrolled in 2007-2008 to measure the effect on schooling in 2008-2009. This approach eliminates any potential effect of schooling on the ability measure as these children had never been enrolled at the time of taking the ability test, so the comparison is for the children not enrolled in year 1 to their siblings that are not enrolled. In columns 1 to 4, we first consider only children ages 5 to 7 and not enrolled in year 1 since many children in Burkina Faso are not enrolled at this young age. Then in columns 5 to 8, we further restrict the sample to only children ages 5 to 6 and not enrolled in year 1 to remove any potential concern that the not-enrolled seven year olds are somehow different than other seven year old children.[40] The household fixed effects logit regressions in columns 1 and 5 indicate that within a given household, a young child with a one standard deviation higher ability measured in year 1 is respectively 19.2 and 20.4 percentage points more likely to be subsequently enrolled in year 2.[41] The coefficient in column 1 is statistically significant at the five percent level. While the coefficient in column 5 is not statistically significant at standard levels, there are only 52 children in the regression as the household fixed effects logit is only identified from households with multiple children ages 5

---

[39] Mean enrollment for the sample of children in Grades 2, 1 or not enrolled (columns 1 to 4) is 0.35, while for children in Grade 1 or not enrolled (columns 5 to 8), average enrollment is 0.23.

[40] In 2007-2008, 74 percent of children ages 5 to 7 were not enrolled, and of these children 31 percent are then enrolled in 2008-2009. For children 5 to 6 years old, 89 percent of them were not enrolled in 2007-2008, and of these children, 28 percent are then enrolled in the subsequent year.

[41] Mean enrollment in Year 2 for the sample of children ages 5 to 7 (columns 1 to 4) is 0.29, while for children ages 5 to 6 (columns 5 to 8) average enrollment in Year 2 is 0.27.

to 6 who were not enrolled in year 1. Young children 5 to 6 years old who are not enrolled in year 1 and who have two siblings of higher ability who also are not enrolled in year 1 subsequently have a 17.3 percentage point lower probability of enrollment in year 2 (column 8).

Fourth, in Table 2.8, we present two alternative measures of a child's cognitive ability. To allay any concerns that transforming the Raven's scores into age-adjusted z-scores might have introduced bias, in columns 1 to 4, we estimate regressions using the Raven's raw test score. Results are consistent with those in Table 2.3. In the household fixed effects specification, within a given household, a child with a one standard deviation higher Raven's raw score (3.35 questions), has an 18.4 percentage point higher likelihood of being enrolled. In columns 5 to 8, we also employ an alternative measure of cognitive ability, the WISC Digit Span, to examine the relationship with current enrollment and find results consistent with using the Raven's test. Children with a one standard deviation higher own WISC z-score have a 17 to 22 percentage point higher probability of enrollment, representing 32 to 41 percent of the mean enrollment level. Children with two siblings having a higher WISC z-score have a 14 percentage point lower probability of enrollment (26 percent of the mean level of enrollment).

6. Conclusions

In this paper, we find strong evidence of sibling rivalry when parents make educational investment decisions in rural Burkina Faso. However, in contrast with previous research that generally focuses on easily observable demographic characteristics to measure sibling rivalry, we use measures of a child's own cognitive ability and different specifications of his siblings' abilities to test for how parents make schooling investment decisions. We examine both the extensive margin (school enrollment and grade progression) as well as the intensive margin of discretionary school expenditures. Own ability has a positive effect on educational outcomes, after controlling for individual and family characteristics and when using a family fixed effect specification. We find that within a given household a child with one standard deviation higher ability compared to the average ability of their siblings is 31 percent more likely to be enrolled. Regardless of how we measure sibling ability, we find evidence of sibling rivalry, and our results are particularly strong when we consider relative measures of sibling ability. The magnitude of these impacts is large. For a child that has one higher ability sibling the probability of enrollment declines by 16 percent and having two higher ability siblings lowers enrollment by 30 percent. Our findings are robust to using alternative objective measures of cognitive ability and the parent's perceptions of a child's chance of future success and to addressing issues about the potential reverse causality of schooling influencing child ability measures.

Our results can likely be generalized to other developing countries that have not yet achieved universal primary or secondary education and in which parents are deciding whether to send their children to school in a given year. This paper explores the context in which the decision is made during primary school, whereas in other countries the choice may occur later in a child's education. A more complete understanding of how parents make the educational investment decision is useful for policymakers. Our findings that high ability children within a family are more likely to be enrolled and receive more educational resources suggest that parents focus on getting the most talented children through higher levels of education, rather than spreading some education evenly amongst all of their children. If fixed costs for schooling and non-convexities in the education production function are important factors in the decision to not invest in education equally across all children, then supply-side schooling interventions (such as building schools, reducing class size, or school inputs like textbooks or uniforms) to raise the schooling of all children and achieve the Millennium Development Goals might not be effective if they are not large enough to overcome these non-convexities. These types of policies might raise the schooling of the more talented children rather than the schooling of all children, and so to increase overall education rates, demand side policies might be necessary. The results also point towards additional benefits of early childhood development programs that improve the cognitive ability of children and help them better compete with their siblings.

TABLES FOR CHAPTER 2

Table 2.1. Summary Statistics of Burkina Faso Social Protection Evaluation (BSPE) Data

Variable:	Mean	Standard Deviation	Percentage of Households with Variation
Household Size	8.88	3.81	
Number of Wives	1.47	0.95	
Number of Siblings	3.81	2.32	
Number of Sisters	1.79	1.50	
Number of Non-Biological Children in Household	0.41	0.89	
Male (Fraction Male)	0.54	0.50	
Age	9.41	2.99	
Birth Order	2.27	1.34	
Proportion Either Parent Ever Enrolled	0.13	0.34	
Log Household Assets	12.36	1.49	
Parent Perception of Chance Child Succeeds in Formal Employment			
Percentage 'Small' Chance	25		
Percentage 'Medium' Chance	38		
Percentage 'Large' Chance	29		
Percentage 'Very Large' Chance	8		
Raven's Raw Test Score	4.86	3.35	
Own Ability (Raven's age adjusted z-score)	-0.01	1.00	
WISC Raw Test Score	7.58	4.56	
WISC Age Adjusted Z-Score	-0.02	0.99	
Average Grades Completed	1.81	2.08	
Proportion Children Currently Enrolled	0.54	0.50	56%
Proportion Children Ever Enrolled	0.59	0.49	54%
Proportion Children with an On-Time Start	0.40	0.49	54%
Grade Progression	0.52	0.48	
Discretionary Education Expenditures (in FCFA)	845	1752	
Number of Households	1507		
Number of Children	4635		

Notes: All summary statistics are based on information for the 4635 children age 5 to 15 in Year 1 unless otherwise noted. Household assets are measured in FCFA (415 FCFA=$1) and the variable is created by taking the log of the sum of household durable goods and livestock. Parent perceptions of the chance their child succeeds in formal employment ranges from 0 to 3, with 0 indicating a small chance and 3 indicating a very large chance, own ability is measured using the Raven's Colored Progressive Matrices and normed by age (z-score), timely start indicates if the child started school by age 7 or younger, grade progression in school is the child's grade in school divided by number of years since the child started attending school and ranges from 0 to 1, discretionary education expenditures is the sum of per child expenses for school supplies and parent association fees in FCFA. Summary statistics for grade progression and average grades completed are based on only 4476 and 4633 children respectively due to missing grade data. Data source: Burkina Faso Social Protection Evaluation (BSPE) data from 2008.

Table 2.2. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between Current School Enrollment, Sibling Rivalry, and Parent Perceptions

Dependent Variable: Current Enrollment	(1)	(2)	(3)	(4)	(5)
Parent Perceptions of Child's Chance of Success in Formal Employment		0.184***	0.162***	0.111***	0.109***
		[0.025]	[0.023]	[0.019]	[0.019]
Highest Sibling Perception			-0.065***		
			[0.019]		
Higher Sibling Dummy (1 if any sibling with a higher perceived chance of success)				-0.070**	
				[0.033]	
One Higher Sibling Dummy (1 if only 1 sibling with a higher perceived chance of success)					-0.053
					[0.036]
Two Higher Sibling Dummy (1 if 2 siblings with a higher perceived chance of success)					-0.075
					[0.048]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with a higher perceived chance of success)					-0.146**
					[0.063]
Number of Siblings	-0.025***		-0.025***	-0.026***	-0.025***
	[0.009]		[0.009]	[0.009]	[0.009]
Number of Sisters	0.022**		0.021**	0.022**	0.022**
	[0.010]		[0.010]	[0.010]	[0.010]
Birth Order	0.014		0.025**	0.023**	0.025**
	[0.009]		[0.010]	[0.010]	[0.010]
Male	0.031*	0.040**	0.026	0.027	0.027
	[0.018]	[0.019]	[0.019]	[0.019]	[0.019]
Parent Schooling (Either parent ever enrolled=1)	0.181***		0.171***	0.166***	0.166***
	[0.040]		[0.045]	[0.044]	[0.044]
Log Household Assets	0.018*		0.018*	0.017*	0.017*
	[0.010]		[0.010]	[0.010]	[0.010]
Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects?	Yes	No	Yes	Yes	Yes
Household Fixed Effects?	No	Yes	No	No	No
Number of Children	4635	3210	4536	4536	4536

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Columns 1, 3, 4, and 5 present marginal effects for logit regressions. Column 2 presents marginal effects from a household fixed effects conditional logit regression. Regressions are restricted to children age 5 to 15, and number of siblings and number of sisters are for all siblings and sisters in the household. Regression sample includes 4635 children, of which 4536 have parent perception measures and 3210 have siblings with differing outcomes and parent perception measures. Data source: Burkina Faso Social Protection Evaluation (BSPE) data from 2008.

Table 2.3. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between Current School Enrollment and Child Ability

Dependent Variable: Current Enrollment	(1)	(2)	(3)	(4)
Own Ability (Raven's age adjusted z-score)	0.165***	0.147***	0.095***	0.085***
	[0.017]	[0.014]	[0.016]	[0.016]
Highest Sibling Ability		-0.068***		
		[0.013]		
Higher Sibling Dummy (1 if any sibling with an ability > own ability)			-0.110***	
			[0.024]	
One Higher Sibling Dummy (1 if only 1 sibling with an ability > own ability)				-0.086***
				[0.025]
Two Higher Sibling Dummy (1 if 2 siblings with an ability > ability)				-0.161***
				[0.032]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with an ability > own)				-0.177***
				[0.041]
Number of Siblings		-0.019**	-0.023***	-0.018**
		[0.009]	[0.009]	[0.009]
Number of Sisters		0.018*	0.020**	0.019**
		[0.010]	[0.010]	[0.010]
Birth Order		0.028***	0.025***	0.029***
		[0.009]	[0.009]	[0.010]
Male	0.026	0.019	0.019	0.018
	[0.020]	[0.020]	[0.020]	[0.020]
Parent Schooling (Either parent ever enrolled=1)		0.182***	0.181***	0.180***
		[0.041]	[0.041]	[0.041]
Log Household Assets		0.018*	0.019**	0.020**
		[0.010]	[0.010]	[0.010]
Age Fixed Effects?	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	Yes	Yes
Household Fixed Effects?	Yes	No	No	No
Number of Children	2861	4635	4635	4635

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Column 1 presents marginal effects from a household fixed effects conditional logit regression. Columns 2 to 4 present marginal effects for logit regressions. Regressions are restricted to children age 5 to 15, and number of siblings and number of sisters are for all siblings and sisters in the household. Regression sample includes 4635 children, with 2861 having siblings with differing enrollment outcomes. Own and sibling ability are measured using the Raven's Colored Progressive Matrices and normed by age (z-score). Data source: Burkina Faso Social Protection Evaluation (BSPE) data from 2008.

Table 2.4. OLS Regressions Estimating Relationship between Discretionary Education Expenditures and Child Ability, Only Enrolled Children

Dependent Variable: Discretionary education expenditures on supplies and parent associations fees (FCFA)	(1)	(2)	(3)	(4)
Own Ability (Raven's age adjusted z-score)	169.53***	139.23**	123.26*	112.35
	[54.60]	[55.71]	[70.35]	[75.67]
Highest Sibling Ability		-36.25		
		[40.94]		
Higher Sibling Dummy (1 if any sibling in household with an ability > own ability)			-10.74	
			[65.65]	
One Higher Sibling Ability Dummy (1 if only 1 sibling in household with an ability > own ability)				26.72
				[70.90]
Two Higher Sibling Ability Dummy (1 if 2 siblings in household with an ability > ability)				-136.03*
				[76.93]
Three or More Higher Sibling Ability Dummy (1 if 3 or more siblings in household with an ability > own)				-46.74
				[140.12]
Number of Siblings		27.23	24.06	27.63
		[34.28]	[33.40]	[35.33]
Number of Sisters		-69.66*	-68.48*	-69.65*
		[40.02]	[39.92]	[40.18]
Birth Order		-19.96	-24.38	-16.57
		[38.86]	[39.80]	[37.71]
Male	-82.35	-62.57	-61.95	-61.85
	[84.72]	[67.57]	[68.16]	[67.92]
Parent Schooling (Either parent ever enrolled=1)		257.47**	256.14**	255.01**
		[116.486]	[116.79]	[116.99]
Log Household Assets		12.42	11.95	12.54
		[25.83]	[25.58]	[25.93]
Age Fixed Effects?	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	Yes	Yes
Household Fixed Effects?	Yes	No	No	No
Number of Children	1994	2511	2511	2511

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions are restricted to children ages 5 to 15 who are currently enrolled in school. Discretionary education expenditures are the sum of per child expenses for school supplies and other parent association fees in FCFA, with a mean of 845 FCFA. The regression in column 1 includes household fixed effects and the sample is restricted to children in households with at least 2 enrolled children. Columns 2 to 4 include village fixed effects and the sample is restricted to children who are currently enrolled. Own and sibling ability are measured using the Raven's Colored Progressive Matrices and normed by age (z-score). Data source: Burkina Faso Social Protection Evaluation (BSPE) data from 2008.

Table 2.5. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between Alternative Schooling Outcomes and Child Ability

Dependent Variable:	Ever Enrolled (1)	Ever Enrolled (2)	On Time Start (3)	On Time Start (4)	Grade Progress (5)	Grade Progress (6)
Own Ability (Raven's age adjusted z-score)	0.174***	0.079***	0.156***	0.071***	0.146***	0.052***
	[0.017]	[0.014]	[0.018]	[0.014]	[0.015]	[0.010]
One Higher Sibling Dummy (1 if only 1 sibling with an ability > own ability)		-0.073***		-0.044**		-0.043***
		[0.023]		[0.020]		[0.016]
Two Higher Sibling Dummy (1 if 2 siblings with an ability > own ability)		-0.162***		-0.094***		-0.106***
		[0.028]		[0.027]		[0.019]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with an ability > own ability)		-0.167***		-0.130***		-0.120***
		[0.041]		[0.034]		[0.029]
Number of Siblings		-0.016*		-0.008		-0.012**
		[0.009]		[0.008]		[0.006]
Number of Sisters		0.016		0.016		0.017***
		[0.011]		[0.010]		[0.006]
Birth Order		0.024**		-0.007		0.016**
		[0.010]		[0.010]		[0.007]
Male	0.040*	0.033*	0.010	0.001	0.048**	0.028*
	[0.022]	[0.019]	[0.023]	[0.017]	[0.021]	[0.014]
Parent Schooling (Either parent ever enrolled=1)		0.203***		0.160***		0.120***
		[0.047]		[0.039]		[0.027]
Log Household Assets		0.023**		0.027***		0.014**
		[0.010]		[0.008]		[0.006]
Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	No	Yes	No	Yes
Household Fixed Effects?	Yes	No	Yes	No	Yes	No
Number of Children	2751	4635	2716	4635	2584	4476

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Cols 1 and 3 present marginal effects for a household fixed effects conditional logit regression. Cols 2 and 4 present marginal effects for logit regressions. Regressions are restricted to children age 5 to 15. On-time start indicates if the child started school by age 7 or younger, grade progression is grade in school divided by number of years since the child started school and ranges from 0 to 1. Own and sibling ability are measured using the Raven's CPM and normed by age.

Table 2.6. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between Current School Enrollment and Child Ability, Restricted to Children in Grades 2 or Lower

Dependant Variable: Current Enrollment	Grade 2, 1 or Not Enrolled				Grade 1 or Not Enrolled			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own Ability (Raven's age adjusted z-score)	0.114***	0.065***	0.048***	0.044***	0.083**	0.026***	0.021*	0.018
	[0.026]	[0.012]	[0.014]	[0.015]	[0.040]	[0.010]	[0.011]	[0.012]
Highest Sibling Ability		-0.025**				-0.007		
		[0.010]				[0.009]		
Higher Sibling Dummy (1 if any sibling with an ability > own ability)			-0.035				-0.010	
			[0.021]				[0.016]	
One Higher Sibling Dummy (1 if only 1 sibling with an ability > own ability)				-0.024				-0.000
				[0.023]				[0.017]
Two Higher Sibling Dummy (1 if 2 siblings with an ability > own ability)				-0.062**				-0.039**
				[0.026]				[0.020]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with an ability > own ability)				-0.053				-0.012
				[0.044]				[0.034]
Number of Siblings		-0.021**	-0.023***	-0.021**		-0.017***	-0.018***	-0.017***
		[0.008]	[0.008]	[0.009]		[0.006]	[0.006]	[0.006]
Number of Sisters		0.001	0.002	0.002		0.005	0.005	0.005
		[0.009]	[0.009]	[0.009]		[0.006]	[0.006]	[0.006]
Birth Order		0.027***	0.026**	0.028***		0.020**	0.019**	0.020***
		[0.011]	[0.011]	[0.011]		[0.008]	[0.008]	[0.008]
Male	0.036**	0.008	0.009	0.008	0.044	-0.001	-0.001	-0.000
	[0.041]	[0.019]	[0.019]	[0.019]	[0.048]	[0.014]	[0.014]	[0.014]
Household Characteristics?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household Fixed Effects?	Yes	No	No	No	Yes	No	No	No
Number of Children	1409	3118	3118	3118	730	2548	2548	2548

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Sample in columns 1 to 4 includes 3118 children in Grades 2 or lower, with 1409 having siblings with differing enrollment outcomes. Sample in columns 5 to 8 includes 2548 children in Grade 1 or lower, with 730 having siblings with differing enrollment outcomes. Columns 1 and 5 present marginal effects from a household fixed effects conditional logit regression. Columns 2 to 4 and 6 to 8 present marginal effects for logit regressions. Regressions are restricted to children age 5 to 15, and number of siblings and number of sisters are for all siblings and sisters in the household. Mean enrollment for the sample of children in Grades 2 or lower is 0.35 while for children in Grade 1 or lower is 0.23.

Table 2.7. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between School Enrollment in Year 2 and Child Ability Measured in Year 1

Dependent Variable: Current Enrollment Year 2	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own Ability (Raven's age adjusted z-score)	0.192**	0.060**	0.037	0.031	0.204	0.050	0.009	0.005
	[0.092]	[0.028]	[0.028]	[0.028]	[0.156]	[0.034]	[0.033]	[0.032]
Highest Sibling Ability		-0.038*				-0.069**		
		[0.021]				[0.029]		
Higher Sibling Dummy (1 if any sibling with an ability > own ability)			-0.043				-0.080	
			[0.040]				[0.050]	
One Higher Sibling Dummy (1 if only 1 sibling with an ability > own ability)				-0.020				-0.049
				[0.043]				[0.048]
Two Higher Sibling Dummy (1 if 2 siblings with an ability > own ability)				-0.122				-0.173*
				[0.076]				[0.101]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with an ability > own ability)				-0.135				-0.152
				[0.094]				[0.095]
Number of Siblings		-0.034	-0.038*	-0.033		-0.029	-0.033	-0.030
		[0.021]	[0.021]	[0.021]		[0.022]	[0.023]	[0.023]
Number of Sisters		0.010	0.013	0.012		0.002	0.005	0.002
		[0.024]	[0.024]	[0.025]		[0.030]	[0.031]	[0.031]
Birth Order		0.060**	0.060**	0.062**		0.076***	0.075***	0.079***
		[0.026]	[0.027]	[0.027]		[0.026]	[0.027]	[0.027]
Male	-0.183	-0.078*	-0.079*	-0.082*	-0.16	-0.082	-0.084	-0.088*
	[0.129]	[0.044]	[0.044]	[0.043]	[0.197]	[0.056]	[0.056]	[0.053]
Household Characteristics?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household Fixed Effects?	Yes	No	No	No	Yes	No	No	No
Number of Children	123	643	643	643	52	442	442	442

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Columns 1 and 5 present marginal effects from a household fixed effects conditional logit regression. Columns 2 to 4 and 6 to 8 present marginal effects for logit regressions. Regressions in columns 1 to 4 are restricted to children age 5 to 7 who were not enrolled during Year 1 and in columns 5 to 8 are restricted to children age 5 to 6 who were not enrolled during Year 1. Sample includes 643 children ages 5 to 7, with 123 having siblings with differing enrollment outcomes. Mean enrollment in Year 2 for the sample of children ages 5 to 7 (columns 1 to 4) is 0.29, while for children ages 5 to 6 (columns 5 to 8) average enrollment in Year 2 is 0.27. Sibling ability measures are for all siblings not enrolled during Year 1.

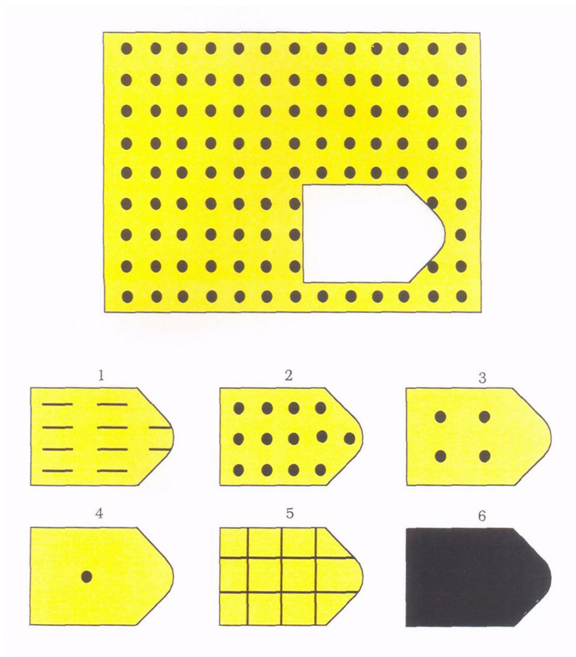
Table 2.8. Marginal Effects from Logit and Conditional Logit Regressions Estimating Relationship between Current School Enrollment and Alternative Child Ability Measures

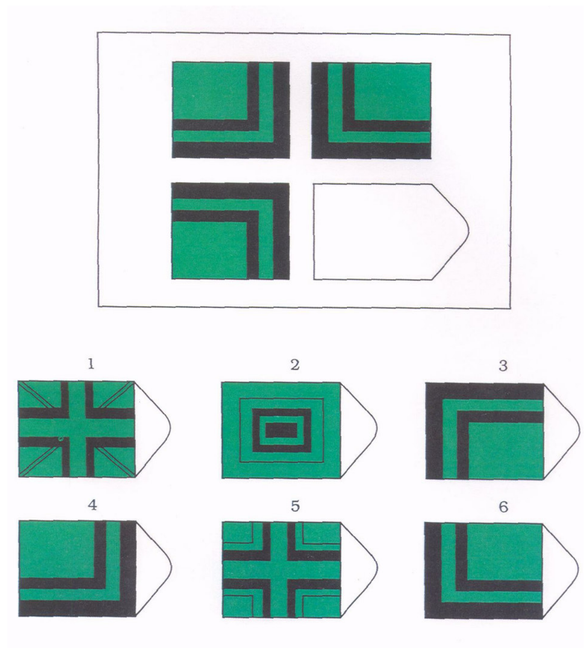
Dependant Variables: Current Enrollment	Raven's (raw score)				WISC Digit Span (z-score by age)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own Ability [Raven's raw, WISC age adjusted z-score]	0.055***	0.045***	0.035***	0.035***	0.221***	0.221***	0.180***	0.171***
	[0.006]	[0.005]	[0.005]	[0.005]	[0.028]	[0.022]	[0.024]	[0.025]
Highest Sibling Ability [Raven's raw score, WISC age adjusted z-score]		-0.010***				-0.043**		
		[0.004]				[0.018]		
Higher Sibling Dummy (1 if any sibling has an ability > own ability) [Raven's raw, WISC age adjusted z-score]			-0.085***				-0.115***	
			[0.022]				[0.031]	
One Higher Sibling Dummy (1 if only 1 sibling with an ability score > own score) [Raven's raw, WISC age adjusted z-score]				-0.080***				-0.100***
				[0.024]				[0.029]
Two Higher Sibling Dummy (1 if 2 siblings with an ability score > own score) [Raven's raw, WISC age adjusted z-score]				-0.104***				-0.138***
				[0.032]				[0.043]
Three or More Higher Sibling Dummy (1 if 3 or more siblings with an ability score > own score) [Raven's raw, WISC age adjusted z-score]				-0.095***				-0.191***
				[0.037]				[0.052]
Number of Siblings		-0.025***	-0.026***	-0.025***		-0.020**	-0.020**	-0.015*
		[0.009]	[0.009]	[0.009]		[0.009]	[0.009]	[0.009]
Number of Sisters		0.020**	0.020**	0.020**		0.014	0.014	0.014
		[0.010]	[0.010]	[0.010]		[0.011]	[0.011]	[0.010]
Birth Order		0.031***	0.030***	0.032***		0.022**	0.022**	0.027**
		[0.010]	[0.010]	[0.010]		[0.011]	[0.011]	[0.011]
Male	0.025	0.020	0.021	0.021	0.045**	0.018	0.020	0.019
	[0.020]	[0.020]	[0.020]	[0.020]	[0.021]	[0.020]	[0.020]	[0.020]
Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village Fixed Effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household Fixed Effects?	Yes	No	No	No	Yes	No	No	No
Number of Children	2861	4635	4635	4635	2843	4463	4463	4463

Notes: Robust standard errors in brackets, clustered at village level. * significant at 10%; ** significant at 5%; *** significant at 1%. Columns 1 and 5 present marginal effects from a household fixed effects conditional logit regression. Columns 2 to 4 and 6 to 8 present marginal effects for logit regressions. Columns 1 to 4 calculate ability measures using the Raven's CPM raw score. Columns 5 to 8 calculate ability measures using the WISC Digit Span normed by age (z-score). Sample sizes vary due to missing WISC Digit Span data. All regressions are restricted to children age 5 to 15 and also include household level controls for parent schooling and assets. Data source: Burkina Faso Social Protection Evaluation (BSPE) data from 2008.

Figure 2.1. Example Problems from the Raven's Colored Progressive Matrices and WISC Digit Span Tests

Panel A : Raven's Colored Progressive Matrices





Correct Response: Option 2

Correct Response: Option 3

Panel B: WISC Digit Span

Question	Correct Response
Digit Span Forward:	
"8-2"	"8-2"
"5-1-7-4-2-3-8"	"5-1-7-4-2-3-8"
Digit Span Backward:	
"8-2"	"2-8"
"1-6-5-2-9-8"	"8-9-2-5-6-1"

CHAPTER 3
RISK AND PROTECTIVE FACTORS FOR SCHOOL DROPOUT IN MEXICO AND CHILD[42]

1. Introduction

Determinants of educational attainment have been well studied in both developed and developing countries. Poverty level is often cited as an explanation for poor educational achievement: poor youth are more likely to leave school early than non-poor youth. However, the mechanism through which poverty contributes to low educational achievement is not fully understood. There are many potential explanations for why young people might leave school early, including having to work, becoming a parent, engaging in violent activity, and repeating grades until they are either forced to or choose to leave school. Poor youth are also more likely to be engaging in these activities or other risky behaviors than are non poor youth. Similarly, there are many other factors present in young peoples lives (and are not typically observed by researchers) that could be driving the relationship between poverty and educational achievement. This study aims to better understand the wide variety of factors in young people's lives that might be able to explain how poverty may affect the decision to leave school prior to completing a secondary degree.

The specific needs of young people are an increasingly debated subject in the development community. A large literature has emerged to inform the debate in Latin America, principally by mapping youth behaviors, positing policy and program interventions to address the behaviors, and, to a lesser extent, to identifying the (economic) factors driving young people's decisions (Lloyd 2005, World Bank 2006). A primary shortcoming of the research is that it treats youth as a homogenous group and reports average behaviors, thereby not capturing the complexities of the youth population across its many dimensions.

---

[42] This work is co-authored with Wendy Cunningham at the World Bank

Even if we do understand the heterogeneity in the youth population, appropriate policy requires an understanding of why young people make the decisions they do. The evidence for Latin American youth is limited. The role of economic incentives and budget constraints affecting decision-making by young people in developing countries has been examined (World Bank 2006 and the sources within), building on Gruber's (2001) work in the US. A second line of work considers the broader context in which young people form their preferences and make their decisions (World Bank 2003, 2007), building on the public health and psychology research in the US. While this literature gives good insights as to how to modify behaviors, it is silent on to whom prevention programs should be targeted.

This paper uses a regression framework to look at various factors related to early school dropout, finding that while parent education is important, risk and protective factors are strongly correlated as well. Many of these typically unobservable factors are also strongly correlated to six additional risky behaviors. We then construct a more complete picture of the heterogeneous youth population in LAC by quantifying and describing the at-risk youth population, focusing on the relationship between early school dropout and risk. We use a non-parametric methodology and special cross sectional youth surveys from Mexico and Chile that permit us to identify the factors that today's risk-taking youth had in their childhoods. Despite our inability to determine causality, these factors are potential candidates to use in identifying who should be the target of prevention policies and programs.[43]

The paper finds that over 20 percent of 18 to 24 year olds in Chile and 40 percent in Mexico have influences in their early and current lives that predispose them to negative behaviors, they have engaged in these behaviors, and they are suffering the consequences. Of these young people with high levels of risk in their lives, 77% in Chile and 70% in Mexico have left school prior to

[43] Causality can only be determined using panel data that includes information on behaviors and household factors, which do not exist in LAC, to the best of our knowledge.

completing secondary education. Conversely, 40 percent of Chileans age 18 to 24 and 16 percent of Mexicans of the same age do not display any of these risks, behaviors, or consequences. Of these youth with low risk, 93 to 94 percent completed secondary schooling. In both the stepwise regressions and the cluster analysis, several key factors emerge as important in addition to poverty, gender, age, and living in a rural area. These factors include having a good relationship with parents and peers, strong connection with local governmental institutions and schools, urban residence, younger age, and spirituality

There are six sections following this introduction. Section 2 provides a brief review of the literature. Section 3 presents the conceptual framework. Section 4 describes the methodology and Section 5 discusses the data. The results are presented in section 6 and section 7 concludes.

2. Review of the Literature

There is a large literature that looks at the determinants of education in developing countries. One thread in this literature (e.g. Horowitz and Souza (2011)) finds that poverty has a causal impact on the educational achievement and has informed policy makers. Many social programs, such as cash transfer programs which are widespread in LAC, assume that improving a family's wealth will improve educational attainment. Positive shocks to household income has been found to affect school attendance in Mexico (Skoufias and Parker 2001), Brazil (Duryea, Edwards, and Ureta 2003), and Colombia (Attanasios, Meghir, and Santiago 2005). Another thread (e.g. Carniero and Heckman (2002)) argues that unobservable characteristics of the child and his or her development and family that affect educational achievement are also correlated with poverty, thus driving the relationship between wealth and schooling.

Understanding determinants of behaviors other than schooling and employment are relatively recent in the field of economics. Perhaps the most extensive study in the economic literature is Gruber (2001), which investigates the determinants and implications of nine different behaviors among US youth – school dropout, smoking, risky driving, sexual activity, suicide, marijuana use, crime, alcohol use, and mis-nutrition – using both cross sectional and time series data. This study expands the rational addiction model developed by Becker and Murphy (1988) to allow for youth-specific characteristics identified in the psychology and human development literature, namely myopia, time inconsistent preferences, and projection bias.[44] The study concludes that youth respond to incentives such as age-specific legal penalties, prices, and income and that the marginal cost to additional risk taking is small once participation in risky behavior has begun.

Recent studies have tested these conclusions in the context of developing countries, finding similarities and differences with the US[45]. Cunningham and Maloney (forthcoming) found that negative shocks to household income affect labor force entry. Similarly, household poverty increases youth violence in Brazil (World Bank 2007), Colombia (Duque, Klevens, and Ramirez 2003) and the US (Grogger 1998, Mocan and Reese 1999).

A shortcoming of the economic literature is the absence of a discussion about preference formation and that early experiences vary across the youth population, leading to heterogeneity in decision-making during the youth years.[46] In contrast, an extensive literature in the public health and psychology fields start from the assumption that preference formation and constraints before the youth years, as well as during them, explain a significant portion of the variation in youth behaviors. The ecological risk framework posits that youth are a product of individual (personal), micro-, and macro-environmental factors (Bronfenbrenner 1979). The individual factors are those skills, behaviors, and ideas that are "hardwired", rather than formed, such as rage, optimism, or general health. The micro factors include preferences taught and formed by the family, peers, community, and local institutions and the constraints imposed by the same, including household poverty. The

[44] "Projection bias" is understood as today's preferences may not be representative of future adult preferences.
[45] See Lloyd ed. (2005), World Bank (2006), and Attanasio et al. (2005) for example.
[46] Gruber (2001) discusses preferences, but it focuses on the time-inconsistency of preferences between the youth period and adulthood. It does not investigate why different youth have different preferences.

macro factors include more general influences and constraints, such as gender/race discrimination, armed conflict, poverty and economic inequality. These factors are commonly classified into two groups: the set of personal, macro, and micro factors that increase the risk of negative behaviors (risk factors) and the set that prevent youth from engaging in negative behaviors, commonly called protective factors. Each person has a set of risk and protective factors that influence preference formation, constraints, and thus behaviors.

Most of the empirical testing of the ecological risk framework depends on correlations between current risky behaviors and personal, micro, and macro factors of the youth population. The literature in the US primarily focuses on identifying factors related to single risky behaviors (e.g. substance use).[47] Recent US literature has begun to take advantage of longitudinal data available in the US to demonstrate that many of those factors correlated with risk taking behaviors are actually causal factors.[48]

The few studies that have tested the ecological risk framework using data from Latin American and Caribbean find similar results to those in the US, however most do not consider schooling behaviors. Blum (2004) uses data collected in 11 Caribbean countries and finds that a positive relationship with a caring adult, whether in the family or in school, is a key factor that is

---

[47] There is an extensive literature in the US which is too large to cover here. The US National Library of Medicine and National Institute of Health maintains a web page (http://www.ncbi.nlm.nih.gov/sites/entrez), that lists hundreds of published articles that have used this methodology. A few of the papers that motivated this paper include: Resnick et al. (2004) look at risk and protective factors related to youth violence; Blum et al. (2002) describe the ecological risk framework and provide empirical evidence for three risky behaviors (weapon related violence, ever had sexual intercourse and ever used cocaine); Scal et al. (2003) look at risk and protective factors related to smoking; Zweig et al (2002) identify methods of predicting risk profiles using risk and protective factors such as psychosocial development, school and family characteristics using OLS and multinomial logit regressions with cross sectional data; and Bernat and Resnick (2006) provides a comprehensive review of the resiliency framework and additional empirical support for this framework in promoting healthy youth development.

[48]These studies, which also appear on the NIH web site, use the National Longitudinal Survey of Adolescent Health (ADDHEALTH) through the University of North Carolina, Carolina Population Center, surveys youth in grades 7-12 with the first wave of interviews in 1994 (http://www.cpc.unc.edu/projects/addhealth/design). Follow-up waves were in 1996 and 2002, the latter enabling more detailed analysis. This survey is school based and asks about risk and protective factors as well as behaviors. The National Longitudinal Survey of Youth (NLSY) 1997 from the Bureau of Labor Statistics, surveyed males and females born in 1980-1984 (http://www.bls.gov/nls/nlsy97.htm) and focuses primarily on educational and employment outcomes. The previous NLSY was from 1979.

positively correlated with less risky sexual behavior (sexual debut and condom use), contraception, pregnancy and childbearing. Youth in the English-speaking Caribbean who have lower risky behaviors[49] are also those who are more connected to parent/family, attending religious services, feel little rage, have not been abused or witnessed parental violence, do not have family members who have attempted suicide, have mentally healthy parents, and have households free of illegal substances (World Bank 2003). A series of papers by Brook, et. al. (2001, 2002a, 2002b) find similar correlations between household factors, poverty, and community factors for explaining substance use and violence among Colombian youth, while Hutz and Silva (2003) find that young Brazilian men who have been incarcerated for violence are disproportionately the sons of poor, uneducated, and violent fathers.

A few studies in LAC do consider behaviors related to education. Lloyd (2005) identifies the positive effects of schooling and health on transitions to risk-free adulthood using DHS surveys from developing countries. World Bank (2007) finds that low self-esteem, spirituality, school connectedness, abuse in the home, abuse in the community, connectedness to institutions, poverty and gender are the factors most associated with risky behaviors and negative outcomes in Brazil.[50] However, these studies do not consider the relationship between education related behaviors and other behaviors.

Young people who engage in one risky behavior often engage in many; i.e. negative behaviors often co-occur (Zweig et. al. 2002). The US data show that most of the risky behavior is being undertaken by a small set of young people who are engaging in multiple risky behaviors (Lindberg et. al. 2000, Bartlett et. al. 2005, Husler et. al. 2005, Brener et. al. 1998, and Zweig et. al. 2001). Similarly, smoking, drug and alcohol use, early sexual initiation, violence, and delinquencies

[49] Behaviors and outcomes studied include: perception of general health, ever had sexual intercourse, ever attempted suicide, violent behavior, problems due to alcohol and drugs.
[50] Behaviors and outcomes studied include: grade repetition, early labor force entry, early sexual initiation, risk taking sexual practices, alcohol use tobacco use, illegal drug use, violence and suicide attempt.

co-occur in the Caribbean (Ohene et. al. 2005) and in Brazil (World Bank 2007). The co-ocurrence is not surprising given the common set of risk and protective factors correlated with different negative behaviors. However, the co-occurrence may also be due to causal factors between behaviors. For example, youth in Peru, the Dominican Republic, Honduras, Brazil, and Argentina report that early school dropout results primarily from early work as well as substance abuse, violence and pregnancy and a myriad of other risk factors. Further, the respondents point to early dropout and early parenthood as underlying causes of their current unemployment or employment in low-quality jobs (IDDI 2006, Dasso 2006, Weiss 2006, World Bank 2006b).

3. Conceptual Framework

We formalize the ecological risk framework in the following model. A person *i* has a set of behaviors, B_i, that are determined by a vector of risk factors, r_i, and protective factors, p_i, determined at the individual (I_i), micro (c_i), and macro (M_i) levels.

$$B_i = f\left(r\left(I_i, c_i, M_i\right), p\left(I_i, c_i, M_i\right)\right) \qquad (1)$$

If an element in any of the vectors I_i, c_i, or M_i leads to a positive behavior in B_i, it will take a positive value in $p\left(I_i, c_i, M_i\right)$ and a 0 in $r\left(I_i, c_i, M_i\right)$. Likewise, an element that leads to risky behavior in B_i will take a 0 value in $p\left(I_i, c_i, M_i\right)$ and a positive value in $r\left(I_i, c_i, M_i\right)$. A weighted average of the risk (r_i) and protective factors (p_i) specific to each person will predict the behavior elements in the vector B_i. Behaviors include elements such as unprotected sex, school truancy, or substance use.

The outcomes of these behaviors are a function of the behaviors, the individual, micro, and macro environments, and luck (δ). The outcome, O_i, is given by

$$O_i = f(B_i, I_i, c_i, M_i, \delta) \quad (2)$$

Outcomes can be good (not dropping out of school early, youth participation) or negative (school dropout, exclusion). Risk and protective factors are included in the function since they can magnify or mitigate the outcomes of a behavior. We assume a distribution of δ that is constant across people, but instrumental in determining if behavior B_i becomes outcome O_i.

We assume that both risk and protective factors are constant during a person's youth. For example, we assume that family risk factors do not vary extensively over a youth's age range. This assumption is necessary since we do not have longitudinal data that would allow variance in factors over time, however the model can be expanded to allow for time varying factors and behaviors.

Equations 1 and 2 can be used to link the concepts of risk/protective factors, behaviors, and outcomes to levels of risk and give insight to policy. A person with high values of the elements in $r(I_i, c_i, M_i)$, low values of the elements in $p(I_i, c_i, M_i)$, and a B_i and O_i that displays few negative behaviors and negative outcomes, is defined as being type I risk. More generally, a person classified as type I risk has many risk factors and few protective factors, indicating that they have a predisposition to engage in negative behaviors, but the person has not undertaken any risky behaviors. Prevention programs would be targeted to this group.[51] A person with high values of the individual elements in $r(I_i, c_i, M_i)$, low values of the individual elements in $p(I_i, c_i, M_i)$, a B_i that displays many negative behaviors, and an O_i, with few negative outcomes would be classified as type II risk, where risk factors are present, protective factors are few, and the young person has engaged in risk-taking behaviors without having experienced any negative consequences. These

[51] See Blum (1998) for a discussion of resiliency based intervention programs.

youth are at-risk of suffering consequences and can thus benefit from prevention or second chance programs. Youth with positive values in the O_i vector are those who are suffering the consequences of their behaviors and are thus categorized as type III risk. These youth tend to have high risk factors, poor protective factors, and many negative behaviors, many of which might be identified prior to their suffering the consequences of their risky behaviors. They are candidates for second chance (remedial or rehabilitation) programs.

4. Methodology

4.1. Stepwise Regression Analysis

This paper uses step-wise regression analysis to identify those variables with the highest explanatory power for early school dropout and other positive youth behaviors. The regressions are estimated using the forward stepwise method, which is an iterative process in which a set of independent variables (the risk and protective factors in this study) is identified a priori. A variable is then randomly selected from that set and included in the regression. If the newly included variable is significant at the 1% level, it is maintained in the regression and another randomly selected variable is added to the regression. If any variable loses significance, it is dropped in the next round. Once all the variables in the initial set have been tested, the regression will only contain those variables that are significant at the 1% level, which can be interpreted as the sub-set of variables that are most correlated with the dependent variable. The regressions were also run using the backward stepwise method as a robustness check, with similar results. We run three different models, depending on the nature of the dependent variables: Ordinary Least Squares (OLS) for continuous dependent variables, Logit for binary dependent variables and Ordered Logit for ordered ordinal variables, where higher numbers correspond with a better outcome.

We estimate the key correlates for seven behavioral outcomes in Mexico and six outcomes in Chile. We run separate analyses for each country since the data are sufficiently different to not allow pooling. However, to the extent possible, similar variables are used in our analysis of Chile and Mexico. We estimate separate regressions for men and women to ensure that differences between genders would not confound results.[52]

4.2. Cluster Analysis

Cluster analysis is a means to identify correlations across large data sets without imposing, a priori, a structure on the data. Observations are grouped based on minimizing a distance measure between each variable for each observation, i.e. the observations in a cluster share a set of common variables. By comparing the mean values of various variables across clusters, we can characterize each cluster.

Ward's method (minimum-variance) was selected since it provided the most distinct and interpretable clusters.[53] Ward's linkage cluster analysis uses the error sum of squares criteria and is a commonly used agglomerative hierarchical method.[54] The variance is minimized by calculating the sum of squared errors from the mean of the cluster for each of the m variables for each observation:

$$W = \sum_k \sum_j \sum_i (x_{ijk} - \bar{x}_{jk})^2 \qquad (3)$$

$i = 1,....,n$ observations, $j = 1,...,m$ variables, and $k = 1,...,l$ clusters

Initially, each of the n observations forms its own cluster. The first merge is identified by calculating the sum of squares for each pair of cluster. The pairing with the smallest sum of squares

52 Bagby and Cunningham (2007) find large differences between male and female youth typologies, including different incidence of certain behaviors and different exposure to/levels of risk and protective factors.

53 There are many different ways to perform cluster analysis, and no particular method is considered the best.

54 An attractive feature of the Ward's method is that it performs well with groups that are of unequal size, which, as will be shown in the results, strongly characterizes the data. See Everitt et al. (2001) for a theoretical discussion of Cluster Analysis and Ward's criterion. See Cunningham and Maloney (2001) for an application.

is identified and those clusters are joined, leaving n-1 clusters. The second grouping calculates the sum of squared errors again and pairs the two clusters that have the smallest value, leaving n-2 clusters. The process is repeated until the optimal number of clusters has been reached.

Three tools were used to determine the optimal number of clusters. First, stop commands following two possible rules (Calinsky-Harabasz and Duda-Hart) were used to find criterion for each cluster possibility.[55] The Calinski and Harabasz method suggests the optimal number of clusters (g) that maximizes an index C(g) which uses the pooled within-cluster covariance matrix (W) and the between-cluster covariance matrix (V), where

$$C(g) = [\text{trace } (V)/(g-1)]/[\text{trace } (W)/(n-g)] \qquad (4)$$

The Duda and Hart method maximizes D(g), where

$$D(g) = Je(2)/Je(1) \qquad (5)$$

Je(2) is the sum of the within cluster sum of squared distances between the objects and centroid if the cluster is split into two and Je(1) is the within cluster sum of squared distances. The local criteria calculated in equations 4 and 5, C(g) and D(g), are then combined with test statistics for each clustering option to suggest the optimal number of clusters. Larger values of both methods indicate that the clusters are more distinct from each other while lower values indicate that the clusters are not very different from each other and therefore are artificially sub-divided.

Second, dendrograms were used to select among the multiple "right" clusters that the other methods may give. Dendrograms graphically depict the hierarchical relationship between the clusters by showing the order in which clusters are merged as well as the distance between the clusters. At each level of the cluster formation process, a dendrogram can be generated to view the relationships between the clusters. The dendrogram changes as clusters are grouped and un-grouped, thus enabling the researcher to optimally choose the clustering level.

---

[55] These two methods are implemented in STATA. They were identified as the two best methods available (out of 30) by Milligan and Cooper (1985) and are discussed in Everitt et al. (2001).

Third, once the optimal number of clusters is suggested and the dendrograms generated, the clusters themselves are observed and the means of the variables are compared across clusters. The differences found between the different clusters are used to ultimately determine the optimal number of clusters. For instance, if 5 clusters were suggested, 6 clusters were investigated to see if there was an interpretable difference. If not, 5 clusters were used, however if so, 6 clusters were used. This process could then repeat.

The cluster analysis is performed for 10 cohorts, identified by age, gender, and country. The sample for each country is divided ex-ante by age and sex since, when pooling the sample, these two variables dominated the clusters to such an extent that the risk and protective factors of interest played a very small role. In addition, we do not want to confound the effects of different factors and behaviors amongst youth that are at very different points in their lives. We therefore separate out the age cohorts for young people that are at different points in the age profile. Six cohorts from Mexico are analyzed: female ages 12 to 14, female ages 15 to 17, female ages 18 to 24, and males in each of the three age groups. Only the four older cohorts from Chile are analyzed since youth age 12 to 14 were not included in the sample.

While the objective of a cluster analysis is to identify which variables move together, a decision was made to treat some variables endogenously and others exogenously. For example, a hypothesis is that poverty status is a good indicator for a youth being "at-risk". If we use this variable to create the clusters, it is possible that poverty is such a strong factor that it drives the clusters and renders the other variables meaningless. Thus, for these type variables, we carried out the analysis treating them as both endogenous and exogenous and found little difference. We thus report only the results for treating them exogenously.

The advantage of cluster analysis is that the only priors required are in the variable construction, such that they range between 0 and 1. For continuous variables, the value was normalized. Binary variables were assigned a 0 or 1 and discrete variables were assigned a value between 0 and 1 based on the ordering of the responses. A variable takes a value of 1 the closer it is to the variable being described. For example, the variable "abuse" takes a value of 1 if there is abuse in the household and a 0 if there is not while the variable "connected" takes a value of 1 if the respondent reports that they reach out for help all of the time if they have problems, a value of 0 if they never reach out, and a value in between depending on how frequently they reach out. Variables that could not be ordered in a logical way were not included in the analysis.

5. Data Description

5.1. Data

As youth departments and governments become more sophisticated in their efforts to understand the lives of young people, several LAC countries have developed specialized youth surveys. We use two surveys in this paper: the 2003 National Youth Survey (*Encuesta Nacional de Juventud*) from Chile and the 2000 National Youth Survey from Mexico. Youth are defined as being between the ages of 12 and 24 for this analysis, consistent with World Bank (2006). In addition to having information about youth education and employment, these cross-sectional data contain rarely available information on perceptions, family background, attitudes, and behaviors, thus providing a rare insight into the youth populations in these countries.

Chile initiated its National Youth Survey, *Encuesta Nacional de Juventud* (ENJ), in 1994 and has repeated it every three years. We use the 2003 data, which was the most recent that could be accessed at the time of the analysis. The survey contains information about risk factors – household poverty, ethnicity, family cohesion, neighborhood violence, and social exclusion – and protective factors - trust in institutions, connectedness, good relationship with parents, and mental health. The behaviors and outcomes are limited to schooling, employment, sexual health, and participation in activities. There is no information about drug use or violent behavior. We limit the sample to those aged 15-24, giving us a sample size of 5321.[56]

Mexico's National Youth Survey, *Encuesta National de Juventud* (ENJ) was carried out in 2000.[57] The sample is nationally representative and was performed in two stages: first, the entire household was surveyed and asked basic household characteristics, and later, youth ages 12-29 were asked a separate set of youth-specific questions. Only those aged 12 to 24 were included in this analysis, resulting in a sample of 37,979 respondents.[58]

As with Chile, the Mexican dataset includes information on several risk and protective factors and behaviors. However, the survey also includes information about attitudes towards alcohol and drugs, parental residence in the household, parental response to their children's behaviors, attitudes toward school (a proxy for school connectedness) and various proxies for family poverty. Unfortunately, this dataset does not have information about connectedness with other adults, abuse in home (aside from the form used to respond to misbehavior), community violence, ethnicity, or sexual behavior for those aged 12-14. To the extent possible, similar variables are used in our analysis for both countries, although the Chilean data are more robust with the protective factors and the Mexican data include more risk factors.

---

[56] For this analysis, 195 observations out of 7,189 (2.7% of sample) for youth aged 15 to 29 were dropped due to missing data. We then restricted the sample to youth aged 15 to 24 dropping another 1674 observations. In many cases, missing responses could be coded based on responses to related questions so as to maintain a larger sample size. For instance, if someone does not respond as to whether they attend church, after they have already indicated that they do not believe in God, then we assume that they do not attend church.

[57] The survey was repeated in 2005, but the data were not available at the time this paper was under preparation.

[58] Of the almost 60,000 youth in the original sample, about 10,000 youth aged 12-29 were not surveyed the second time, and were dropped from the sample used for the analysis. The reasons for not interviewing these youth were tracked in the dataset: they did not want to participate, were not at home at the time of the interview and would not return within the week, were on vacation, were working or at school in another city, were disabled, and other. Comparing the poverty variables (education level and monthly earnings of heads of households) and rural means of this dataset before and after dropping the data showed no significant difference at α=.01. An additional 5% of the observations were dropped in creation of the variables; the resulting sample was not statistically different from the original. Finally, the data were restricted to youth aged 12 to 24, thus further decreasing the sample size by 8903.

Three groups of variables are used as posited by the ecological risk framework: protective factors, risk factors, and behaviors/outcomes.[59] The empirical findings from the US, Latin America, and Caribbean studies guided our selection of risk and protective factors to include in the Mexico and Chile analysis. The twelve risk factors in the analysis include low socioeconomic status (parental education level);[60] rural residence, indigenous ethnicity, low healthcare access, social exclusion, weak family cohesion, physical or psychological abuse by a parent, household substance abuse, poor parental response to bad behavior, positive parental influence on smoking and alcohol, perceived high neighborhood violence; and experienced discrimination. The fifteen protective factors considered include trust in institutions (government & community); live with both parents; connectedness (overall, with mother, with father, with other adult); engage in activities with parents; church attendance and spirituality; school quality; feeling of preparedness for the future; optimism towards employment; and sense of wellbeing. The data permit us to analyze secondary school dropout as well as six additional positive behaviors/outcomes: not inactive,[61] older age at first job[62], safe sexual behavior (not sexually active, using contraception if sexually active), older age at first pregnancy/parenthood, participation in activities, and healthy attitudes towards alcohol. Table 3.1 shows which variables are used in the Mexico analysis and which are used for Chile, defines the variables, and discusses the methodology for the creation of composite proxy variables.

[59] The classification of a variable as a "risk factor" or "protective factor" may seem like an artificial distinction. For example, poverty is a risk factor while lack of poverty may be a protective factor. However, there are some risk factors without corresponding protective factors and vice versa. Rather than enter the debate in this paper, we simply classify a factor as "risk" or "protective" based on how the question was asked in the survey.

[60] Various proxies for poverty were tested, including household earnings, household luxury/durable goods ownership, and socio-economic indicators generated by the government. All had similar results as the "parental education" variable, but the parental education was used since the first alternative is a poor measure due to earnings being only a temporal measure of wealth, there was not sufficient variance in the second measure, and the algorithm for the third variable was not available.

[61] Inactivity is defined as not being in school or work and not searching for a job.

[62] In an environment where many children are engaging in child labor activities, considering the age at which the children start work for the first time is informative. We must note however that we must be careful to distinguish between those that delay work so that they can stay in school and those older youth that are not able to find a job and are unemployed. For this reason, we look both at age of first job and inactivity.

All variables were normalized in the range 0 to 1. Behaviors were normalized such that a 1 indicates good behavior and a 0 indicates "risky" behavior. For factors, a high level of a risk factor would get a 1 and a high level of a protective factor would be 1. So for example, a risk factor variable that takes on three values would be assigned a 1 for the worst situation (for example, household abuse), a 0 for the best situation (for example, no household abuse), and a 0.5 for intermediate (for example, the threat of household abuse).

5.2. Descriptive Statistics

In both Chile and Mexico, a significant percentage of youth are dropping out of school early. Nearly 30 percent of Mexican and 14 percent of Chilean youth in the samples dropped out of school prior to completing high school (Table 3.2). While about 30 percent of Mexican and Chilean youth are inactive after age 18, 12.8 percent of Mexican and 6.7 percent Chilean youth are inactive before age 18.

Chileans have earlier, and more risky, sex than Mexicans. While only 13 percent of Mexicans age 15 to 17 report having had their first sexual experience, 27 percent of Chileans in this age group report the same. Half of the sexually active Mexicans use contraception; Chileans are 12 percentage points less likely to use protection than are Mexicans. 2.5 percent of Mexicans age 15-17 have had their first child while nearly 4 percent of Chileans in the same age group are parents (Table 3.2). Nearly 30 percent of Mexican youth and 80 percent of Chilean youth report having ever been involved in activities outside of school; this percentage decreases with age. Finally, only 3.4 percent of the sample of youth aged 12-24 in Mexico can justify getting drunk and 1 percent justify drug use.

Many of these behaviors by Mexican and Chilean youth co-occur, as is also reported by young people interviewed across Latin America (IDDI 2006, Dasso 2006, Weiss 2006, World Bank 2006b) and found in empirical work in the US (Bartlett et. al. 2005), Caribbean (World Bank 2003), and Brazil (World Bank 2007). For example, inactive youth disproportionately leave school before completing their secondary education as demonstrated by the high correlation between not dropping out early and not being inactive (0.42 in Mexico and 0.38 in Chile). Risky and early sexual behaviors are also positively correlated with early school leaving (0.35 in Mexico and 0.27 in Chile), early working (0.14 in Mexico and 0.29 in Chile), and youth inactivity (0.21 in Mexico and 0.30 in Chile) (Tables 3.3a and 3.3b). There is a positive correlation between participation in activities and positive youth behaviors in both countries (Tables 3.3a and 3.3b).

Many Mexican and Chilean youth have risk factors in their lives, which may underlie their negative behaviors. Over half of the Mexican sample and one-third of the Chilean sample are from families whose parents have no more than a primary education and thus can be considered poor. In Chile, about 13 percent of the sample is considered rural[63] while in Mexico about 25 percent is rural. About 11 percent of the Chilean sample self-identifies as indigenous. Fortunately, abuse and substance abuse in the home are not very prevalent; six percent of Chileans report suffering abuse in the home and 8.5 percent report substance abuse in their homes. Approximately half of Mexican youth and 18 percent of Chilean youth report social exclusion. The incidence increases with age in Chile and decreases with age in Mexico (Table 3.2).

These risk factors are correlated with many negative behaviors. Figure 3.1a and 3.1b shows the strong correlation between school dropout and parent education. Young people with better educated parents (a proxy for lower poverty level) are less likely to drop out before completing secondary school. They also engage in riskier behaviors as shown by the negative correlation coefficients in Tables 3.3a and 3.3b. Living in a rural area is negatively correlated secondary school completion and with five additional behavior variables – not inactive, later age of initial job, later age of first parenthood and participation in activities. Figures 3.2a and 3.2b suggest that rural living has

[63] The sampling was done in communities with at least 2000 inhabitants, so a rural indicator means that the respondent comes from a community with between 2000 and 5000 inhabitants. An urban respondent lives in a community of at least 5000 people.

different implications for men's and women's behaviors. Rural men generally initiate sexual activity later than do urban men, particularly in Mexico, while rural women begin their sexual lives earlier than do urban women. In Mexico, rural men and urban women are more likely to use contraception but in Chile rural women are more likely to use contraception. Being indigenous is not highly correlated with most behavior variables (Table 3.3b). Indigeneity has a significantly negative correlation with older age at first job, but the magnitude is small (-0.05). Table 3.3b shows that abuse and substance abuse are negatively correlated with all positive youth behaviors in Chile. Poor parental response to misbehavior (such as hitting their child, insulting their child, or accusing their child in front of others) is significantly negatively correlated with all seven behaviors in Mexico. Social exclusion is negatively correlated with positive behaviors in Chile (Table 3.3b), while being socially excluded is positively correlated with not dropping out of school early, starting work at an older age, safe sexual behavior and later age at first parenthood in Mexico (Table 3.3b).[64]

Protective factors are also prevalent in the lives of young Mexicans and Chileans. More than 70 percent of Mexican youth live with both parents (Chilean youth were not asked this question). Mexican and Chilean youth report a high level of personal connections with caring adults. Mexican youth regularly discuss a wide range of topics with their parents including school, work, politics, and religion. While Chileans report good relationships with their parents, ten percent of Chileans also note important relationships with other adults. Fewer than 7 percent of Mexican youth are verbally or physically abused by their parents. Nearly 95 percent of youth Chileans believe in a god, and 66 percent of Mexican youth attend church with some frequency. School quality, a proxy for school connectedness, is high in Mexico. Youth are generally content with their schools in Mexico, with

[64] The opposite trends in the two countries is likely due to the definition that "social exclusion" takes in each survey. In Chile, social exclusion is defined as not having friends to spend time with, whereas the Mexico survey asks who the respondent spends his or her free time with. If the respondent answers "friend" or "boyfriend/girlfriend", he or she is not considered to be socially excluded. However, most who were coded as socially excluded replied that they spend their free time with parents, siblings, and other family. Thus, these respondents may have a rich social life such that they spend time with friends while in school or at their sports club while they are with family during their unstructured free time.

small percentages of youth reporting dissatisfaction in school attributes such as the physical environment, teacher preparedness, and teacher attendance (Table 3.2). However, Mexican youth trust government institutions (6.5% trust politicians, 16.1% trust judges, and 12.3% trust the police) less than they trust local institutions (66.2% trust teachers, 70.8% trust doctors, 15.9% trust shop owners, 12.5% trust union leaders, and 61.8% trust priests). Similarly, in Chile, youth trust local institutions more than government institutions (Table 3.2). Youth have positive outlooks towards life and the future. Nearly 70 percent of Mexican youth report that they are very happy (compared to 1.7 percent reporting they are not happy at all), In Chile, 87 percent of youth feel optimistic about future work possibilities.

Living with both parents, having a positive relationship with parents, and spirituality are highly correlated with positive youth behaviors in both Chile and Mexico (Tables 3.3a and 3.3b). Trust in governmental institutions has a positive and significant correlation with many behaviors in Chile and Mexico, but the magnitude of the relationship is not very large. The school quality composite variable that focuses primarily on teacher and school characteristics is positively and significantly correlated with many behaviors by Mexican youth but the magnitude is small (Table 3.3a). A sense of well-being is weakly, but positively, correlated with Mexican youth behaviors (Table 3.3a).[65]

6. Results

6.1. Factors correlated with risky behaviors

Many of the same risk and protective factors identified in the US and existing Latin American and Caribbean literature explain the variance in school dropout, inactivity, age of first job, safe sexual behavior, age of first pregnancy/parenthood and involvement in activities in Chile and

[65] Based on the literature from the US, mental health is very important in a youth's life, and one would expect these correlations to be larger.

Mexico. The explanatory power of these sets of variables for our sample ranged from (0.01 R^2 – 0.25 R^2), and correctly predicted 68% - 90%[66] of the behaviors. The variables are most successful in explaining the variance in staying in school, safe sexual behavior, and later age of first pregnancy/parenthood for boys and girls in Chile and Mexico and female inactivity in both countries. They explain very little of the variance in involvement in activities in both countries, in later labor force entry and in attitudes toward alcohol in Mexico, and in inactivity among Chilean males. Table 3.4a presents the coefficient estimates from the stepwise regressions for Mexico and Table 3.4b presents the results for Chile. A positive coefficient indicates that the independent variable is correlated with a "good" behavior, i.e. not leaving school prior to completing the secondary level.

Three micro-factors - household poverty, relationship with the family, and use of/relationship with local institutions - repeatedly emerge as key explanatory factors in both countries for school dropout and all additional behaviors, i.e. significant at the 1% level in the stepwise regressions. Consistent with previous literature, household poverty, proxied by parental education level, explains a significant amount of the variance in all the behaviors considered for both genders in both countries. Young men and women who live in poorer households leave school earlier, are more likely to be inactive, start working earlier, engage in riskier sexual activity, have their first pregnancy (females only) at a younger age, and engage in fewer activities than those in wealthier households. This correlation emerges even though we have controlled for connectedness with parents and others. They are, however, less likely to justify drunkenness compared to youth from wealthier households in Mexico. However, even after having controlled for household poverty, many other protective and risk factors are strongly correlated with these behaviors.

---

[66] This is only for the binary variables "not inactive" and "school completion".

The importance of personal connections with peers, parents or other adult figures and positive behaviors found in the US, Latin American, and Caribbean literature is confirmed for the cases of Mexico and Chile. A positive relationship with parents as well as living with both parents explain some of the variance in all seven behaviors and is positively correlated with not dropping out of school early, not being inactive, older age at first job, less risky sexual activity, older age at first parenthood, and participation in activities. Males with a good relationship with their fathers or mothers stay in school, are not inactive, and participate in activities. Girls' relationships with their father is related to staying in school, not working and greater participation in activities while the relationship with the mother is also related to safe sexual practices. Although living with both parents is important for all seven behaviors, the actual relationship with parents is also important, as shown by controlling for the former. This relationship is largely an emotional connection, since engaging in activities with parents did not emerge as a strong protective factor in the analysis.

Abuse in the family weakly emerges as a deterrent to positive behaviors and is not strongly correlated with school dropout. Physical abuse in the household explains the variance of early working and early and unsafe sexual behavior for females but it does not emerge as a potential explanatory variable for male behaviors. It does not emerge as an explanatory variable for school dropout. Parental responsiveness to youth behaviors, whether good or bad, is a good explanatory variable for both males and females. Poor parental response to misbehavior – which may range from a lack of parental responsiveness (no connectedness) to verbal abuse to physical abuse – is negatively correlated with an older age at first job, safe sexual behavior, later age at first parenthood and a healthy attitude towards alcohol for both genders in Mexico, but is not correlated with schooling. On the other hand, poor parental response to good behavior is negatively correlated with school completion, an older age at first job and participation in activities for both genders.

Having a connection with peers is important in explaining variance in Chile, but the relationship is less clear in Mexico. Chilean youth who do not have a group of friends with whom they spend time engage in a range of negative behaviors, including school dropout, inactivity, early labor force entry by women, unsafe sexual behavior, earlier parenthood and less participation in activities (males and females) –, i.e. more social exclusion and worse behaviors go together. The Mexican data do not allow us to measure social exclusion directly, since it only tells us with whom Mexican youth spend their free time. Not spending free time with friends or a girlfriend/partner means greater school completion among Mexican males and participating in fewer in activities among Mexican females; it also means an older work age, safer sexual behavior, an older age at first parenthood and a healthier attitude among males and females. What this may be reflecting is that Mexican youth spend their free time (outside of school) with parents and siblings which, as discussed above, have a strong, positive correlation with good behaviors. Thus the composite variable that we call "social exclusion" may actually be another proxy for family connections among Mexican youth.

Trust in government institutions emerges as an explanatory factor of variance in Chilean behaviors, but it plays a much lesser role for explaining Mexican behaviors. Chilean males who trust in government institutions are older when they take their first job and have safer sexual behaviors while females are less likely to complete school. In Mexico, females that trust in government figures are more likely to be younger when first becoming a parent. Of course, this may be endogenous such that the trust is developed once these behaviors are established. This difference between Chile

– where the variable emerges for several behaviors – and Mexico where it is virtually absent may reflect the difference in the access to, coverage of, and quality of institutions in the two countries.[67]

School quality – which may be a proxy for school connectedness – is positively correlated with school- and labor market- related behaviors in Mexico (the variable is not included in the Chile data set). Those who perceive high school quality are less inactive, and are older when they begin working but are more likely to not complete secondary school. This particularly emerges for girls. It is also important for explaining responsible attitudes toward alcohol, however the link with sexual behaviors is not strong.

Age and spirituality are the only individual factors that consistently emerged in the regression analysis. Age plays a part in explaining the variance in all seven behaviors considered. It is negatively correlated with all the behaviors in Chile and with all behaviors except "older age at first job" for males in Mexico. Young people who identify themselves as being more spiritual also have safer sexual behaviors and participate in more activities. Young Chilean males who are spiritual are more likely to complete school, while in Mexico it is correlated with younger labor force entry. Young Mexican women are particularly influenced by spirituality; however the exact role is not clear.

Mental health is not correlated with school dropout and plays a small role in explaining the variance of the other six behaviors in Mexico and Chile. Good mental health – proxied by feeling prepared for future employment, feeling optimistic about job prospects, or having a sense of well-being – emerge as key explanatory variables for not being inactive in Mexico and Chile, and for older age at first job for Mexican females and younger age at first parenthood for Mexican males.

[67] Or, it may reflect the different definitions of "government institutions" it the two surveys. In Chile, it is defined as higher level institutions – congress, city government, the judicial system, political parties – while in Mexico it is more personal – police, judges, and politicians.

Among the macro-factors, gender and rural/urban residence are key explanatory factors for many behaviors while ethnicity plays a lesser role. For all seven behaviors in both countries, gender is one of the key variables that emerged in the stepwise regressions[68] and risk and protective factors affect the schooling and work outcomes differently for males and females. The R^2 for each female regression is about equal to or much higher than that for males, suggesting that the regressions better explains the variance in girls' behaviors than boys' behaviors. A large set of micro variables, ranging from the family to the community, were useful in explaining girls' behavior, but not that of boys. In Mexico, positive female behaviors were strongly and positively correlated with living with both parents and with a good relationship with the mother, while in Chile the relationship with the mother is positively correlated with better behaviors by girls. These factors played a lesser role for boys. Finally, certain factors affected female behavior in an opposite manner than they affected male behaviors, especially among those variables that were measured only for Mexico. For example, older females are less likely to be working and not in school (defined as "inactive") while there is no strong work/age relationship for males. This is clearly a result of gender roles where women increasingly allocate their time toward homecare as their families grow.

Being from rural areas helps to explain who leaves school earlier, who goes to work at a younger age and who is inactive in both Chile and Mexico. This is not surprising given the familial and greater informal nature of rural labor markets compared to those in urban areas. Notably, these results emerge even when controlling for household poverty and ethnicity (Chile). In spite of the unconditional positive correlation between rural residence and younger marriage or risky sexual activity discussed above, being from rural areas is not a key explanatory variable for risky sexual activity (proxied by earlier age of sexual initiation and use of contraception) in either country (with the exception of Mexican females) or for parenthood in Chile.

[68] The regression estimates of the pooled sample can be obtained from the authors

Being indigenous does not emerge as a key explanatory factor for any of the behaviors in Chile. While 11% of the sample is indigenous, it appears that other factors better explain the variance in behaviors. This seems contrary to evidence from elsewhere that suggests that ethnic disparities are important for explaining the seven behaviors discussed in this paper (Rew and Horner 2003 for the US, Koller et. al. 2004 for Brazil). However, we cannot conclude that indigeneity is not a valuable characteristic for understanding youth behavior since it is possible that other variables that consistently emerge as key explanatory factors are correlated with being indigenous and thus pick up the variance. For example, there is a high correlation (significant at the 1% level) between indigeneity and household poverty (low parental education level). Also, feelings of being discriminated against, which may be ethnic, racial, spatial, or based on other variables, emerges as a key explanatory factor for earlier age of employment for boys and girls.

6.2. Risk profiles of youth in Mexico and Chile

Table 3.5 presents an overview of the clusters arising from the cluster analysis. More than half of young people in these two countries can be considered at-risk. Of the youth that have dropped out prior to completing secondary school 70 percent are most at-risk and suffering the consequences of negative youth behaviors. In Mexico, nearly one-third of youth age 12-24 are suffering the consequences of negative youth behaviors – early school dropout, adolescent mothers, not working – while 17 percent of Chilean youth are in this situation. They also come from the poorest families and have the fewest social bonds.

Another one-quarter of Mexican and Chilean youth are engaging in negative behaviors and on their way to the worst-off category. A younger group in Mexico (20 percent) and Chile (8 percent) are not engaging in risky behaviors, but they have factor in their lives that suggest that they may be graduating to these more harmful groups before long since they also lack the social supports and mental health that the no-risk group boasts.

Figures 3.2a and 3.2b show the average behavior and factor levels for each cluster type. Consistent with the stepwise regressions, good behaviors are positively correlated with protective factors in both Mexico and Chile. As suggested by the ecological framework, the opposite is observed when considering the correlation between risk factors and positive behaviors. First considering type III, we see these clusters have lower averages for those variables that might be considered protective factors and high values for those variables that might be considered risk factors. Youth with "no" risk in their lives have a low presence of risk factors and a higher level of protective factors and good behaviors than those with type II or III risk levels. For youth with type I risk behaviors are better than for types II and III risk and comparable to type 0 risk. Risk factors are worse than those with type 0 risk, but better than those with types II and III risk. Protective factors in Mexico are comparable to those with no risk; however those who are classified as type I risk in Chile have the fewest protective factors of any risk type, suggesting that this is a temporary phase with potentially worse to come. Youth with type II risk have worse behaviors and outcomes compared to those with types 0 and I risk, however are not nearly as bad as those with type III.

Risk and protective factors that are driving some of the differences between the clusters are consistent with those that arose in the stepwise regressions for each risky behavior. The clusters themselves are presented in the Appendix. Some factors repeatedly arise in the best-off or worst-off clusters. Parent education is strongly correlated with risk level, as seen in Figures 3.2 and 3.3. In addition, having a positive relationship with the family is a recurring protective factor across the clusters. The clusters show that youth who live(d) with both parents have a lower incidence of all risky behaviors. Those youth who feel connected to a parent, i.e. those who feel that they can relate to a parent, the parent cares for them, they can depend on the parent, etc, also have lower risky behavior and negative outcomes than youth who do not feel connected. Connection with non-family members can partially compensate for absent parental connection, as shown by youth who are connected to non-family members having less risky behavior than those who do not have connections with anyone. Conversely, family abuse, substance abuse and lack of family cohesion cluster with victimization in Chile, while non-positive feedback from parents is correlated with risky behaviors in Mexico.

Positive institutional factors are also correlated with positive behaviors. School quality matters in Chile, and an individual's relationship with his/her community (trust, feeling of school quality) is positively correlated with voter registration. Low spirituality/ church attendance clusters loosely with the risk type III groups. Surprisingly, social exclusion is rarely correlated with negative behaviors. Instead, the "loner" youth usually showed up as sub-groups in the less at-risk clusters.

Ethnic identity is not strongly correlated with risk level, although indigenous youth are over-represented among clusters defined as risk type III (Table 3.5b). A closer look at the indigenous youth in the Chile sample shows that while indigenous youth have worse risk factors and behaviors, they have better protective factors.[69] The more at risk clusters are about five percentage points more indigenous than the average, but the negative relationship with positive behaviors is slight as seen in Figure 3.2b.[70]

7. Conclusion

Based on unique youth surveys in Chile and Mexico, we find evidence that while early school drop out is strongly correlated with socioeconomic status (proxied by parent education), it is also strongly correlated with other risky behaviors (which are also strongly correlated with socioeconomic status) and several risk and protective factors. Of the risk and protective factors

[69] Further details available upon request.
[70] We must remember, however, that this is a self-reported survey, with the youth indicating if they identify with a particular indigenous group. See Hall and Patrinos (2006) for documentation of the limitations of using self-reported ethnicity.

investigated in this study, good relationship with parents and peers (social inclusion), strong connection with local government and school institutions, urban residence, younger age, and feelings of spirituality cluster together and emerge as key explanatory factors for school dropout in the stepwise regressions. These factors are also positively correlated with other youth behaviors examined for Mexico and Chile– not inactive, later age of first job, safe sexual behavior, later age of sexual initiation /parenthood, participation in activities, and healthy attitudes toward alcohol.

Gender and age are also important variables. The incidence of risk factors, protective factors, and behaviors differ by gender and age in the clusters and the set of factors that is most important in explaining behaviors differed by gender and across age profiles. Most notably, the relationship with government institutions has differing effects on women's and men's behaviors (Chile) while local institutions (school) seem particularly relevant for women. Spirituality also emerges as a more influential factor for women than for men.

Young people in Mexico and Chile who engage in one positive behavior tend to engage in other positive behaviors, while negative behaviors (or outcomes) cluster together. This is not surprising since a common set of protective factors continuously emerge as important explanatory variables for each behavior in the stepwise regressions. Thus, strengthening positive factors and minimizing negative factors, has the potential to influence a range of youth behaviors.

These findings, the first of their kind for Mexico and Chile, are consistent with those from other LAC countries - including the English-speaking Caribbean, Brazil, and Honduras - as well as studies using US data. Unlike other studies, though, self identified indigeneity and household abuse do not emerge as key explanatory factors for Chile, and the variables that proxy positive mental health do not emerge strongly in the analysis for either country.

Despite the limitations in identifying causality, our ability to determine the key factors important for a range of behaviors can be useful to policymakers. While poverty alleviation and education are common policy entry points, programs that tap into other areas of a young people's lives to ensure the presence of positive influences have the potential for a significant impact in preventing risky youth behaviors in LAC. For example, young people without family support can benefit from mentoring programs. Or, certain models of "after" school clubs are effective for affecting a range of behaviors. Or, those from poor households can benefit from cash transfers that reward positive behaviors. In fact, there are a range of evidence-based program and policy interventions at the individual-, micro-, and macro-levels that prevent negative youth behaviors and provide second opportunities to those who have made poor decisions (Cunningham et. al. 2008).

Our findings show that typically unobservable factors behaviors that are strongly correlated with poverty are also strongly correlated with school dropout. While our findings are not causal, they lend support to the argument that it may be these unobservable variables that are driving the relationship between poverty and school dropout. We advocate that studies looking at school dropout take these additional factors and behaviors into account.

FIGURES AND TABLES FOR CHAPTER 3

Table 3.1. Construction of Variables

	Risk Type	Chile	Mexico	Cluster analysis	Factor analysis
Behaviors/Outcomes					
No early school dropout (i.e. completed secondary school)	II	X	X	X	X
Not inactive – either in school nor working	II, III	X	X	X	X
Older age when started working	II, III	X	X	X	X
Safe sexual behavior (not sexually active, contraception use if active)	II	X	X		X
Older age at first pregnancy/parenthood	III	X	X[a]	X	X
Participate in extracurricular activities	II	X	X	X	X
Attitude towards alcohol (respondent can justify getting drunk)	II		X[a]		X
Literate (can read and/or write a message)	III		X	X	
In school	II, III	X	X		
Years of education completed	I, III	X	X		
Working (Not Working[c])	II, III	X	X		
Low number of sexual partners in the past year	II		X[a]		
Older age at onset of sexual activity	II	X	X[a]	X	
Has at least one child (Does not have at least one child[c])	III	X	X		
Married (Not Married[c])	I, III	X	X	X	
Registered to vote, planning on voting in the next election, desires to vote when of age (if under 18 years old)	III	X	X[b]	X	
Has not been a victim of a crime (proxy for criminal activity)	II	X		X	
Attitude towards drugs (respondent can justify using drugs)	II		X[a]		
Protective Factors					
Trust in governmental institutions	I	X[d]	X[a,e]	X	X
Trust in community institutions	I	X[f]	X[a,g]	X	X
Connected (whether youth reaches out – for talk or help- to someone when they have problems)	I	X	X	X	X

Table 3.1 (cont.)

Living with both, one or no parents	I		X	X	X
Positive relationship with father	I	X[h]	X[i]	X	X
Positive relationship with mother	I	X[j]	X[k]	X	X
Connected with an adult other than parents	I	X		X	X
Church attendance	I	X	X	X	X
Spiritual influence in beliefs, opinions and attitudes	I		X		X
School quality[l]	I		X	X	X
Feeling optimistic about future work	I	X			X
Feeling prepared for future employment	I	X			X
Sense of wellbeing (level of happiness reported)	I		X[a]		X
Positive vew on youth employment	I	X			X
Communication with parents (talk to parents when facing a personal problem)	I		X		
Activities with parents	I		X		X
Risk Factors					
Low parental education level (proxy for poverty)	I	X	X	X	X
Have felt discriminated against	I	X			X
Limited access to healthcare	I		X	X	X
Rural residence (versus urban)	I	X	X		X
Social exclusion	I, III	X	X	X	X
Poor family cohesion	I	X		X	X
Physical/verbal abuse in the home	I	X		X	X
Substance abuse in the home	I	X		X	X
Level of perceived violence in the neighborhood	I	X		X	X
Indigenous (self-identifying as indigenous)	I	X			X
Parental influence regarding smoking and alcohol[n]	I		X	X	X
Parental response to misbehavior[o]	I		X		X
Parental response to good behavior[p]	I		X		
Household ownership of durable/luxury goods[m]	I		X		
Monthly earnings of heads of household	I		X		

[a] for respondents age 15 or older
[b] for respondents age 18 or older

Table 3.1 (cont.)

[c] for respondents under age 18
[d] indicates level of confidence in government, congress, city government, political parties, judicial system
[e] indicates level of confidence in politicians, judges, the police, and the military
[f] level of confidence in hospitals, the Catholic Church, schools, universities, and family
[g] indicates level of confidence in teachers, doctors, shop owners, union leaders and priests
[h] quality of relationship with father on various attributes (communication, demonstration of love or affection, understanding and help with problems, respect for private life of youth, the time spend with father)
[i] variety of topics that the youth communicates with the father about (school, politics, religion, sexual relations, work, and other topics)
[j] indicates quality of relationship with mother on various attributes (communication, demonstration of love or affection, understanding and help with problems, respect for private life of youth, the time spend with mother)
[k] variety of topics that the youth communicates with the mother about (school, politics, religion, sexual relations, work, and other topics)
[l] rank of the overall quality of the youth's current/past school as reported by youth (physical building, scholastic materials, teachers preparation, content of courses and teachers assistance)
[m] includes radio recorder, CD burner, TV, cable, VCR, game console, telephone, computer, internet, car/truck/van - could be used instead of household education level to indicate economic class
[n] indicates level of control parents attempt to control children's behaviors (do they forbid smoking/drinking, grant periodic permission, allow the child to make his/her own decision)
[o] how parents respond when child bothers/angers them (0=by talking with their child, 0.5 = punishing, 1 = beating/hitting, insulting, accusation in front of others, stop talking)
[p] indicates frequency with which parents use positive feedback (words of encouragement, hug/kiss, give a gift, concede to something) when child does something good/correct (0 = always, 0.5 = sometimes, 1 = never)

Table 3.2. Descriptive Statistics

	Mexico				Chile		
Actual percentages	All ages	12 to 14	15 to 17	18 to 24	All ages	15 to 17	18 to 24
Average age	17.19	13.0	16.0	20.8	18.9	15.95	20.8
Percent female	53.5	50.65	52.9	55.9	53.0	50.5	54.6
Percent indigenous	-	-	-	-	10.8	11.5	10.4
Percent rural	24.7	28.1	26.0	21.6	13.2	13.3	13.1
Behaviors and Outcomes							
Share dropping out of school before completing high school	29.9	10.0	33.0	42.3	14.2	8.4	17.8
Share in school	45.3	85.9	61.8	27.5	60.5	91.4	41.0
Share idle/inactive	20.3	8.9	17.7	29.9	27.3	6.7	35.9
Share working	41.9	17.2	35.1	35.1	21.4	4.7	32.0
Share having sex	52.1	-	13.3	56.0	59.1	26.9	79.4
Share of the sexually active using protection	50.6	-	54.0	52.5	61.9	59.9	62.3
Share reporting at least 1 child	19.4	-	2.5	28.9	18.6	3.9	27.9
Share married	15.7	0.4	4.1	32.9	4.8	0.1	7.7
Risk and Protective Factors							
Share reporting physical or psychological abuse in home	-	-	-	-	6.3	6.2	6.4
Share reporting problems arising from substance abuse in home	-	-	-	-	8.5	7.0	9.5
Share without access to medical services	49.4	50.2	50.7	48.0	-	-	-
Share attending church weekly	9.3	11.3	10.3	7.3	21.6	20.9	18.1
Share attending church at least once in the past month	66.3	72.0	65.9	62.5	-	-	-
Share believing in God	-	-	-	-	94.8	95.1	94.6
Share optimistic about future work	-	-	-	-	87.7	88.2	87.3
Share reporting being happy (sense of well-being)	98.2	-	98.0	98.4	-	-	-
Share reporting social exclusion	53.2	66.8	53.9	43.1	17.7	10.1	22.5
Share with parents who have a primary degree or less	56.1	56.3	58.0	55.0	38.1	38.8	37.7
Share reporting they have felt discriminated against	-	-	-	-	54.7	54.2	55.1

Table 3.3a. Correlation Coefficients Significant at the 1% Level, Mexico

		Not inactive	No school dropout	Older age when started working	Safe Sexual behavior	Older age at first pregnancy/ parenthood	Participate in activities	Attitude towards alcohol
Behaviors/Outcomes	No early school dropout	0.42						
	Older age when started working	-0.04	0.24					
	Safe Sexual behavior	0.21	0.35	0.14				
	Older age at first pregnancy/child	0.35	0.32	0.07	0.53			
	Participate in extracurricular activities	0.12	0.15	-0.06	0.10	0.13		
	Negative attitude towards alcohol	-0.10	-0.04	0.06	0.04	-0.06	-0.05	
Risk factors	Low parental education	-0.10	-0.21	-0.11	0.03	-0.03	-0.12	0.08
	Rural residence	-0.12	-0.08	-0.02	--	-0.04	-0.06	0.11
	Negative parental response to misbehavior	-0.06	-0.08	-0.05	-0.13	-0.15	-0.03	-0.02
	Social exclusion	--	0.11	0.11	0.18	0.05	--	0.09
Protective factors	Living with both parents	0.27	0.35	0.13	0.49	0.49	0.08	-0.03
	Positive relations with father	0.13	0.12	0.02	0.09	0.15	0.16	-0.03
	Positive relations with mother	0.08	0.09	0.03	0.09	0.12	0.15	
	Church attendance	-0.02	0.06	0.06	0.10	0.05	0.12	0.08
	Spiritual influence in beliefs/opinions/attitudes	--	--	-0.07	0.03	0.04	0.15	--
	Trust in government institutions	--	--	--	0.02	--	0.03	0.02
	Trust in community institutions	--	--	0.02	--	--	0.02	0.02
	School quality	0.05	0.06	0.05	0.04	0.04	--	0.04
	Well-being	0.03	0.05	0.05	-0.03	--	0.04	0.02

-- not significant at the 1% level

Table 3.3b. Correlation Coefficients Significant at the 1% Level, Chile

		Not inactive	No school dropout	Older age when started working	Safe sexual behavior	Older age at first pregnancy/ parenthood	Participate in activities
Behaviors/Outcomes	No early school dropout	0.38					
	Older age when started working	0.13	0.22				
	Sexual behavior	0.30	0.27	0.29			
	Older age at first pregnancy/ child	0.36	0.40	0.13	0.41		
	Participate in extracurricular activities	0.19	0.17	--	0.14	0.19	
Risk factors	Low parental education level	-0.16	-0.23	-0.14	-0.08	-0.09	-0.10
	Rural residence	-0.11	-0.11	-0.10	--	--	-0.05
	Physical/Verbal abuse in the home	-0.04	-0.06	-0.05	-0.05	-0.04	--
	Substance abuse in the home	-0.05	-0.07	-0.08	-0.10	-0.04	--
	Social exclusion	-0.26	-0.23	-0.09	-0.17	-0.31	-0.21
	Indigenous	--	--	-0.05	--	--	--
Protective factors	Good relationship with father	0.10	0.13	0.11	0.13	0.14	0.07
	Good relationship with mother	0.13	0.16	0.10	0.15	0.15	0.06
	Spiritual influence in beliefs, opinions and attitudes	--	0.07	0.05	0.16	0.05	0.20
	Trust in governmental institutions	--	--	0.04	0.09	--	0.04
	Trust in community institutions	--	--	0.06	0.09	--	--
	Optimism towards future work opportunities	0.09	0.05	--	--	0.05	0.05
	Preparation for work	--	--	-0.06	-0.09	-0.04	--

-- not significant at the 1% leve.

Table 3.4a. Key Correlates of Positive Youth Behaviors, Mexico

	No school dropout[a]		Not Inactive[a]		Older age at first job[b]		Safe Sexual behavior[b]		Older age at first parenthood[b]		Participation in activties[b]		Healthy attitude towards alcohol[b]	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Low parental education level	-2.83 (0.14)	-2.41 (0.12)	-0.89 (0.19)	-1.21 (0.12)	-1.14 (0.08)	-0.81 (0.08)	--	-0.36 (0.11)	-0.75 (0.16)	-0.72 (0.12)	-0.69 (0.10)	-0.82 (0.09)	0.65 (0.10)	1.07 (0.12)
Rural residence	-0.66 (0.06)	-0.84 (0.06)	-0.38 (0.09)	-0.90 (0.06)	-0.32 (0.05)	0.23 (0.05)	--	-0.17 (0.06)	--	-0.22 (0.06)	-0.38 (0.06)	--	0.56 (0.07)	0.80 (0.09)
Live with both parents	0.54 (0.07)	0.88 (0.05)	--	1.32 (0.05)	0.47 (0.05)	0.40 (0.04)	0.98 (0.06)	2.12 (0.05)	1.89 (0.08)	2.22 (0.06)	--	0.30 (0.06)	--	-0.39 (0.08)
Poor parental response to misbehavior	--	--	--	--	-0.13 (0.05)	-0.20 (0.04)	-0.13 (0.05)	-0.22 (0.05)	-0.21 (0.07)	-0.31 (0.05)	--	--	-0.20 (0.05)	-0.17 (0.07)
Poor parental response to good behavior	-0.61 (0.12)	-0. 73 (0.11)	--	--	-0.50 (0.09)	-0.45 (0.08)	--	--	--	--	-0.61 (0.11)	-0.55 (0.11)	--	--
Positive relationship with father	0.49 (0.18)	0.91 (0.16)	0.89 (0.23)	0.48 (0.16)	--	--	--	--	-0.70 (0.25)	--	0.76 (0.17)	0.87 (0.16)	--	--
Positive relationship with mother	1.56 (0.19)	1.43 (0.15)	--	1.35 (0.15)	--	-0.63 (0.11)	--	0.73 (0.14)	1.48 (0.26)	1.14 (0.15)	0.90 (0.18)	1.48 (0.16)	--	--

Table 3.4a (cont.)

Social exclusion	0.25	--	--	--	0.17	0.10	0.63	0.60	1.19	0.38	--	-0.13	0.33	0.27
	(0.05)				(0.04)	(0.04)	(0.05)	(0.05)	(0.11)	(0.06)		(0.04)	(0.05)	(0.06)
Spiritual influence	--	--	--	0.30	-0.46	-0.32	0.21	0.32	--	0.49	0.95	0.93	--	--
				(0.08)	(0.06)	(0.06)	(0.08)	(0.08)		(0.09)	(0.08)	(0.07)		
Church attendance	--	--	--	--	--	--	0.43	0.62	--	0.57	0.56	0.81	0.39	--
							(0.07)	(0.08)		(0.08)	(0.08)	(0.07)	(0.08)	
Trust in governmen t institutions	--	--	--	--	--	--	--	--	--	-0.26	--	--	--	--
										(0.10)				
Access to healthcare	-0.23	-0.31	-0.42	-0.44	-0.24	--	--	--	--	--	--	-0.13	--	--
	(0.05)	(0.04)	(0.07)	(0.05)	(0.04)							(0.05)		
Parental influence	--	0.19	--	0.16	-0.18	--	-0.17	--	--	--	--	0.25	-0.37	-0.22
		(0.05)		(0.05)	(0.04)		(0.05)					(0.05)	(0.06)	(0.07)
School quality	--	-0.27	--	0.28	0.52	0.23	0.39	--	--	--	--	--	0.55	0.44
		(0.10)		(0.10)	(0.08)	(0.08)	(0.09)						(0.11)	(0.13)
Connected	--	--	--	0.31	0.30	0.27	0.38	0.95	1.35	1.10	-0.27	--	--	--
				(0.07)	(0.09)	(0.06)	(0.11)	(0.07)	(0.16)	(0.07)	(0.10)			
Activities with parents	0.41	0.33	--	--	--	0.39	--	--	--	--	--	--	0.37	--
	(0.12)	(0.11)				(0.08)							(0.02)	
Age	-0.89	-1.07	--	-2.17	1.09	-0.18	-5.64	-4.16	-5.51	-3.31	-0.41	-0.90	-0.45	-0.13
	(0.15)	(0.13)		(0.14)	(0.12)	(0.11)	(0.15)	(0.14)	(0.24)	(0.16)	(0.14)	(0.14)	(0.15)	(0.18)
Positive sense of wellbeing	--	--	0.56	-0.26	--	0.30	--	--	-0.40	--	--	--	--	0.50
			(0.13)	(0.09)		(0.07)			(0.15)					(0.12)

[a] Logit; [b] Ordered Logit. Standard errors in parentheses.

-- indicates that the variable is not significant at the 1% level and was dropped from the regressions during the iterative process.

Table 3.4b. Key Correlates of Positive Youth Behaviors, Chile

	No school dropout		Not inactive[a]		Older age at first job[b]		Safe sexual behavior[c]		Older age at parenthood[b]		Participate in activities[c]	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Low parental education level	-4.18 (0.70)	-2.76 (0.31)	-2.02 (0.26)	-1.47 (0.23)	-0.28 (0.03)	-0.11 (0.03)	-0.79 (0.18)	-1.09 (0.18)	--	-0.11 (0.02)	-0.83 (0.16)	-1.06 (0.18)
Rural residence	-0.46 (0.17)	--	--	-0.63 (0.14)	-0.09 (0.02)	--	--	--	--	--	--	--
Positive relationship with father	0.61 (0.19)	--	--	--	--	0.07 (0.02)	--	--	--	0.06 (0.01)	--	--
Positive relationship with mother	--	1.59 (0.21)	0.76 (0.25)	0.75 (0.19)	--	--	--	0.98 (0.18)	--	0.10 (0.02)	--	--
Abuse in household	--	--	--	--	--	-0.07 (0.02)	--	-0.67 (0.15)	--	--	--	--
Substance abuse in the home	--	--	--	--	--	--	-0.70 (0.14)	--	--	--	--	--
Social exclusion		-1.54 (0.12)	--	-1.29 (0.10)	--	-0.07 (0.01)	0.37 (0.13)	-0.98 (0.10)	-0.04 (0.01)	-0.16 (0.01)	-0.85 (0.12)	-1.20 (0.11)
Trust in governmental institutions	--	-0.87 (0.24)	--	--	0.07 (0.03)	--	0.78 (0.15)	--	--	--	--	--
Connectedness	0.49 (0.16)	0.60 (0.17)	--	--	--	--	--	--	--	--	--	--
Connection with adult other than parent		--	--	--	--	--	--	--	--	--	0.43 (0.13)	--
Perceived violence in community	--	--	--	--	--	--	--	0.74 (0.18)	--	--	--	--
Poor family cohesion	--	--	--	--	-0.11 (0.02)	--	--	--	--	--	--	0.54 (0.13)

Table 3.4b (cont.)

Age	-3.5 (0.33)	-2.62 (0.31)	-2.66 (0.28)	-4.01 (0.25)	-0.41 (0.03)	-0.47 (0.03)	-5.57 (0.23)	-5.25 (0.21)	-0.15 (0.01)	-0.32 (0.02)	-1.19 (0.18)	-1.37 (0.20)
Spirituality/church attendance	0.70 (0.22)	--	--	--	--	--	0.75 (0.12)	0.58 (0.11)	--	--	1.19 (0.11)	1.39 (0.11)
Optimism towards future work	--	--	--	0.41 (0.13)	--	--	--	--	--	--	--	--
Have felt discriminated against	--	--	--	--	-0.04 (0.01)	-0.03 (0.01)	--	--	--	--	0.25 (0.07)	--
Positive outlook towards work	--	--	0.17 (0.12)	0.04 (0.10)	--	--	--	--	--	--	--	--

[a] Logit; [b] OLS [c] Ordered Logit. Standard Errors in parentheses

-- indicates that the variable is not significant at the 1% level and was dropped from the regressions during the iterative process

Table 3.5a. Mexico Cluster analysis overview

	Type III	Type II	Type I	None
Percent of total sample	33.1	25.1	20.2	21.6
Percent of early dropouts	69.7	17.9	8.5	3.9
Percent of 12-14 year olds	14.1	33.1	24.3	28.5
Percent of 15-17 year olds	38.9	11.9	27.5	21.7
Percent of 18-24 year olds	43.4	26.8	13.1	16.7
Percent of males	31.9	23.2	22.2	22.7
Percent of females	34.2	26.7	18.4	20.7
Average Parental Education (poverty proxy)	Primary graduate, some secondary	Primary graduate, some secondary	Secondary graduate	Secondary graduate
Percent of rural	42.9	24.4	18.1	14.6
Percent of urban	29.9	25.3	20.8	23.9

Table 3.5b. Chile Cluster Analysis Overview

	Type III	Type II	Type I	None
Percent of total sample	16.8	28.0	8.7	46.5
Percent of early dropouts	77.3	19.0	0.5	3.2
Percent of 15-17 year olds	8.3	20.4	15.9	55.5
Percent of 18-24 year olds	22.2	32.8	4.2	40.8
Percent of males	23.7	19.1	13.2	44.1
Percent of females	10.8	4.7	35.9	10.8
Average Parental Education (poverty proxy)	Primary Graduate	Some Secondary Education	Some Secondary Education	Secondary Graduate
Percent of rural	27.2	29.5	7.0	36.3
Percent of urban	15.3	27.7	9.0	48.0
Percent of indigenous	21.7	26.3	8.7	43.3
Percent of non-indigenous	16.3	28.2	8.7	46.9

Figure 3.1a. Relationship between Early Dropout/Inactivity and Parent Education in Mexico

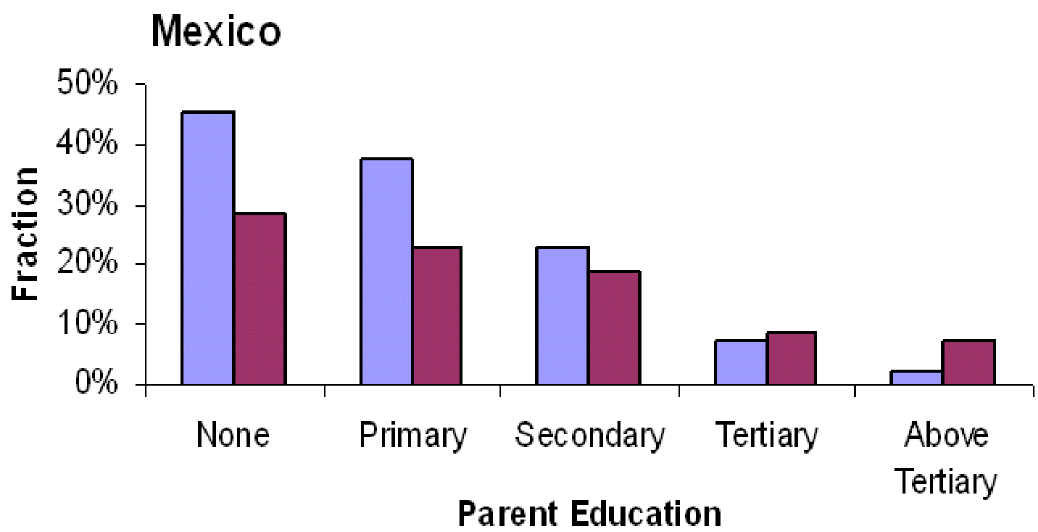


Figure 3.1b. Relationship between Early Dropout/Inactivity and Parent Education in Chile

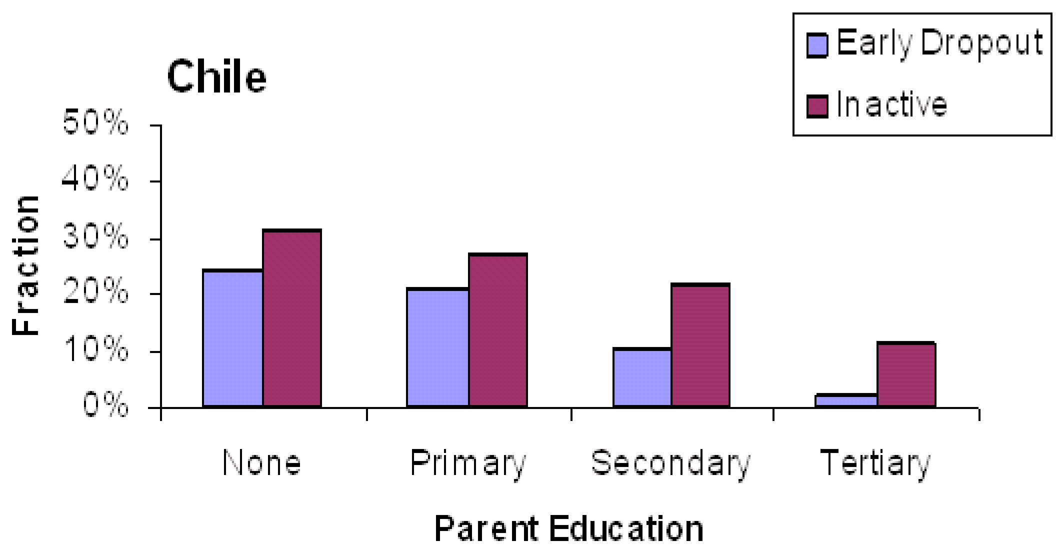


Figure 3.2a. Cluster Group Means – Mexico

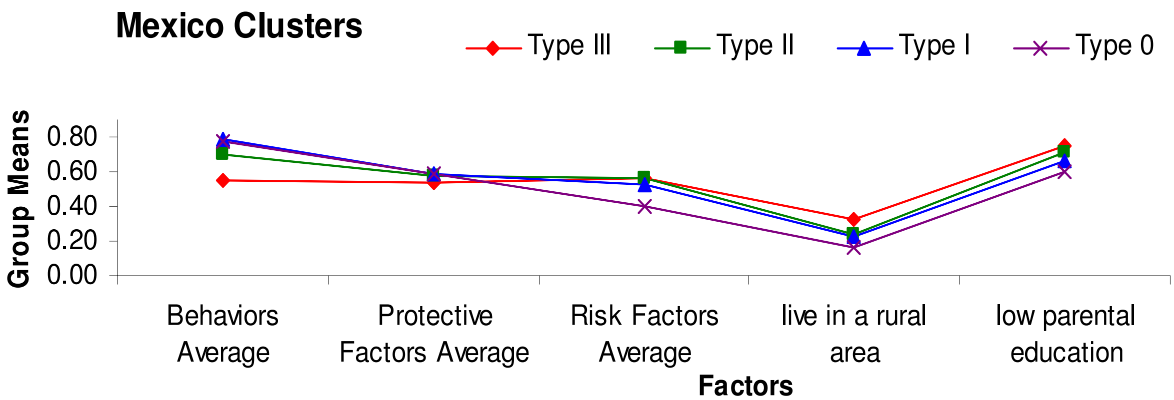


Figure 3.2b. Cluster Group Means – Chile

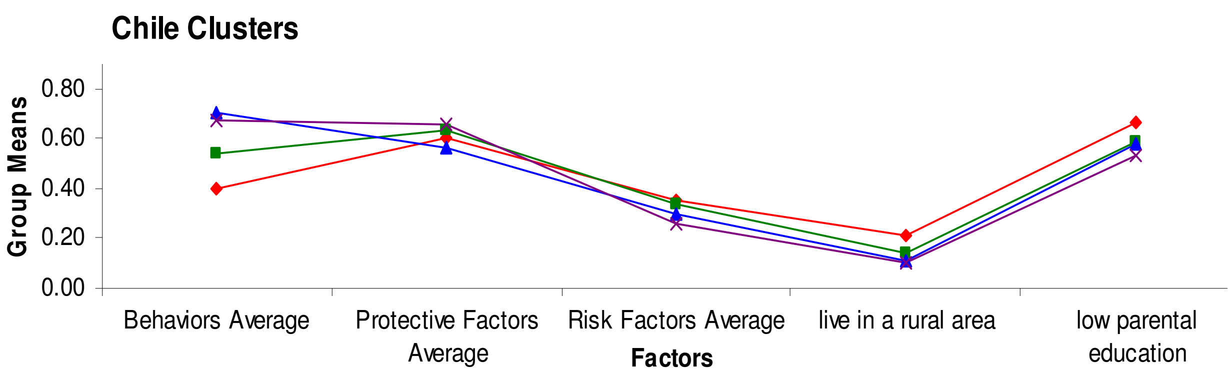


Figure 3.3a. Relationship between Risk Level and Parent Education in Chile

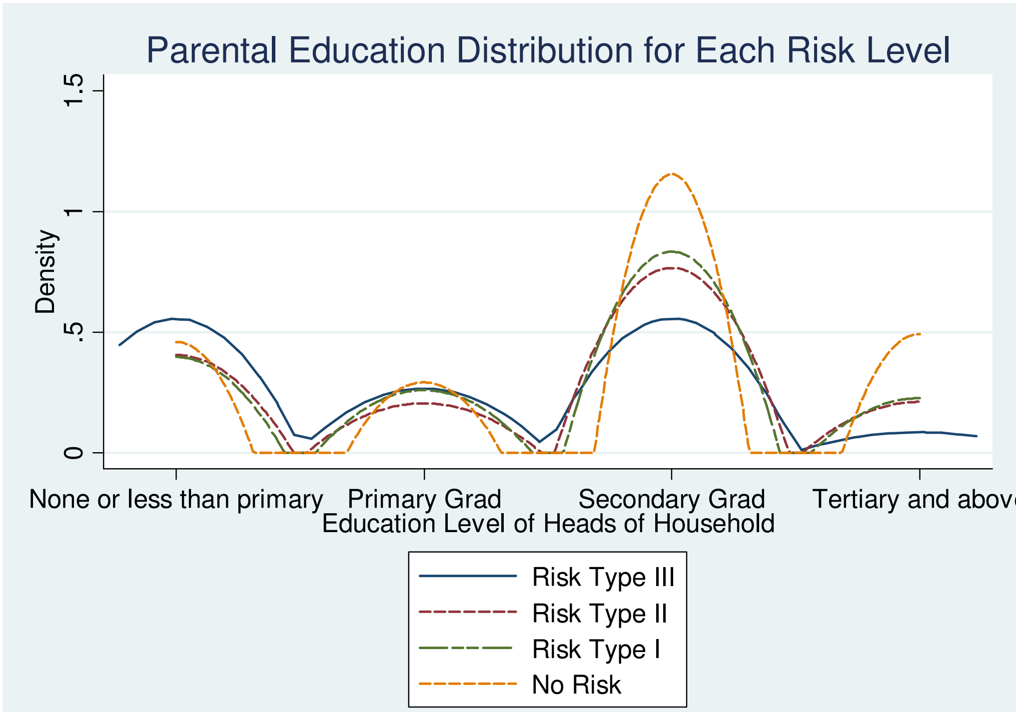
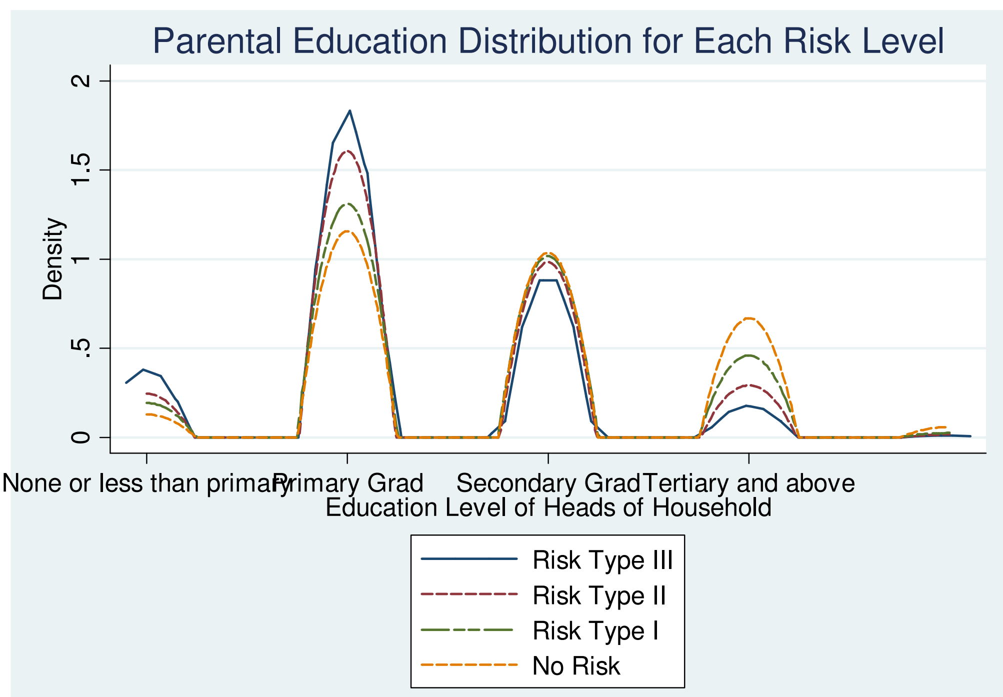


Figure 3.3b. Relationship between Risk Level and Parent Education in Mexico



APPENDIX

A.1. Cluster analysis results: Mexican Males Aged 12 to 14

Cluster name	**at-risk**	**very early workers**	**advantaged loners**	**advantaged youth**
Age	13.25	13.07	12.87	13.05
Behaviors Average	**0.41**	**0.69**	**0.83**	**0.77**
older age when first started working	0.42	0.11	1.00	0.62
not idle	0.49	1.00	1.00	1.00
no early school dropout	0.01	1.00	1.00	1.00
Literacy	0.91	1.00	1.00	1.00
not married	0.99	1.00	1.00	1.00
participate in activities	0.14	0.29	0.17	0.23
not working	0.51	0.58	0.99	0.81
in school	0.01	1.00	1.00	1.00
years of education completed	0.20	0.28	0.27	0.28
Protective Factors Average	**0.54**	**0.58**	**0.60**	**0.58**
relationship with father	0.37	0.42	0.40	0.40
relationship with mother	0.42	0.48	0.46	0.46
connected	0.96	0.84	1.00	1.00
live with both parents	0.87	0.90	0.93	0.90
church attendance	0.36	0.41	0.39	0.39
school quality	0.75	0.81	0.83	0.82
communication with parents	0.49	0.52	0.55	0.51
spiritual influence	0.15	0.22	0.15	0.17
Risk Factors Average	**0.57**	**0.57**	**0.54**	**0.41**
parental influence (alcohol & smoking)	0.44	0.38	0.35	0.37
social exclusion	0.58	1.00	1.00	0.00
limited access to healthcare	0.63	0.52	0.45	0.47
parental response to misbehavior	0.38	0.31	0.29	0.32
parental response to good behavior	0.58	0.52	0.51	0.53
rural	0.55	0.49	0.45	0.43
parental education	0.80	0.70	0.67	0.67
household ownership of goods	0.84	0.79	0.79	0.77
monthly earnings household heads	0.85	0.80	0.78	0.78
n	758	1193	1893	1989
Risk Type	III	II	I	-
Percent of sample	13.0%	20.5%	32.5%	34.1%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.2. Cluster analysis results: Mexican Females Aged 12 to 14

Cluster name	**at-risk**	**early workers**	**advantaged loners**	**advantaged youth**
Age	13.33	12.98	12.97	13.13
Behaviors Average	**0.43**	**0.79**	**0.83**	**0.80**
older age when first started working	0.66	0.74	1.00	0.77
not idle	0.28	1.00	1.00	1.00
No early school dropout	0.01	1.00	1.00	1.00
literacy	0.92	1.00	1.00	1.00
not married	0.97	1.00	0.99	1.00
participate in activities	0.13	0.24	0.24	0.25
not working	0.73	0.87	0.99	0.89
in school	0.01	0.99	1.00	1.00
years of education completed	0.20	0.27	0.29	0.29
Protective Factors Average	**0.56**	**0.59**	**0.61**	**0.59**
relationship with father	0.36	0.39	0.42	0.40
relationship with mother	0.45	0.50	0.52	0.51
Connected	0.95	0.93	1.00	1.00
live with both parents	0.86	0.88	0.93	0.90
church attendance	0.45	0.44	0.42	0.45
school quality	0.77	0.82	0.81	0.81
communication with parents	0.45	0.50	0.55	0.48
spiritual influence	0.18	0.17	0.17	0.18
Risk Factors Average	**0.66**	**0.65**	**0.48**	**0.48**
parental influence (alcohol & smoking)	0.40	0.48	0.06	**0.34**
social exclusion	0.73	1.00	**1.00**	0.00
limited access to healthcare	0.64	0.69	0.00	**0.43**
parental response to misbehavior	0.30	0.27	0.26	0.27
parental response to good behavior	0.59	0.51	0.47	0.51
Rural	0.59	0.49	0.35	0.40
parental education	0.81	0.71	0.60	0.65
household ownership of goods	0.86	0.81	0.76	0.76
monthly earnings household heads	0.87	0.81	0.74	0.75
N	908	2723	982	1376
Risk Type	III	II	I	-
Percent of sample	15.2%	45.5%	16.4%	23.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.3. Cluster analysis results: Mexican Males Aged 15 to 17

Cluster name	**at-risk**	**becoming at risk**	**loners**	**advantaged youth**
age	16.16	15.89	15.88	15.97
Behaviors Average	**0.64**	**0.84**	**0.85**	**0.84**
older age when first started working	0.25	0.41	0.52	0.48
not idle	0.74	1.00	1.00	1.00
no early school dropout	0.06	1.00	1.00	1.00
literacy	0.97	1.00	1.00	1.00
older age at onset of sexual activity	0.85	0.92	0.90	0.88
older age at first pregnancy	0.98	1.00	1.00	1.00
not married	0.97	1.00	1.00	1.00
participate in activities	0.17	0.27	0.27	0.32
not working	0.30	0.63	0.73	0.73
in school	0.06	0.99	0.99	1.00
years of education completed	0.28	0.39	0.42	0.42
safe sex	0.85	0.93	0.91	0.90
low number of sexual partners in past year	0.96	0.98	0.98	0.97
does not have a child	0.99	1.00	1.00	1.00
attitude towards drugs	0.97	0.97	0.96	0.95
attitude towards alcohol	0.88	0.89	0.91	0.85
Protective Factors Average	**0.54**	**0.58**	**0.59**	**0.58**
Connected	0.91	1.00	1.00	1.00
live with both parents	0.84	0.88	0.89	0.88
relationship with father	0.38	0.42	0.44	0.43
relationship with mother	0.44	0.47	0.50	0.49
trust in governmental institutions	0.35	0.36	0.36	0.38
trust in community institutions	0.58	0.60	0.60	0.61
church attendance	0.32	0.35	0.35	0.32
school quality	0.78	0.81	0.81	0.82
communication with parents	0.44	0.48	0.51	0.47
spiritual influence	0.19	0.19	0.18	0.18
sense of wellbeing	0.82	0.85	0.89	0.87
Risk Factors Average	**0.54**	**0.57**	**0.45**	**0.29**
parental influence (alcohol & smoking)	0.54	0.46	0.42	0.45
social exclusion	0.46	0.48	1.00	0.00
limited access to healthcare	0.61	1.00	0.04	0.00
parental response to misbehavior	0.29	0.27	0.21	0.24
parental response to good behavior	0.60	0.55	0.50	0.53
rural	0.51	0.49	0.35	0.29
parental education	0.78	0.72	0.59	0.56
household ownership of goods	0.82	0.78	0.73	0.69
monthly earnings household heads	0.85	0.82	0.74	0.70
n	1765	1120	681	869
Risk Type	III	II	I	-
Percent	39.8%	25.3%	15.4%	19.6%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.4. Cluster analysis results: Mexican Females Aged 15 to 17

Cluster names	**early workers**	**at-risk mothers & wives**	**at-risk idle dropouts**	**resilient loners**	**advantaged loners**	**advantaged youth**
age	16.20	16.42	16.07	15.85	15.85	15.92
Behaviors Average	**0.66**	**0.46**	**0.68**	**0.86**	**0.88**	**0.88**
older age when started working	0.26	0.43	0.65	0.62	0.71	0.67
not idle	1.00	0.21	0.02	1.00	1.00	1.00
no early school dropout	0.00	0.07	0.00	0.96	0.99	1.00
literacy	1.00	0.96	0.95	1.00	1.00	1.00
older age at onset of sexual activity	0.97	0.29	0.97	0.99	0.98	0.97
older age at first pregnancy	1.00	0.43	0.99	1.00	1.00	1.00
not married	0.99	0.34	0.99	1.00	0.99	1.00
participate in activities	0.16	0.07	0.15	0.25	0.27	0.31
not working	0.00	0.83	0.98	0.77	0.86	0.83
in school	0.00	0.06	0.00	0.95	0.99	1.00
years of education completed	0.27	0.27	0.26	0.38	0.42	0.42
safe sex	0.96	0.19	0.96	0.99	0.98	0.97
low number of sexual partners in past year	0.99	0.89	0.99	1.00	1.00	0.99
does not have a child	1.00	0.46	0.99	1.00	1.00	1.00
attitude towards drugs	0.98	0.97	0.97	0.97	0.97	0.96
attitude towards alcohol	0.94	0.94	0.94	0.95	0.93	0.89
Protective Factors Average	**0.57**	**0.49**	**0.57**	**0.57**	**0.62**	**0.60**
connected	1.00	0.85	0.96	0.82	1.00	1.00
live with both parents	0.85	0.27	0.87	0.85	0.91	0.86
relationship with father	0.43	0.41	0.45	0.48	0.52	0.45
relationship with mother	0.39	0.35	0.36	0.41	0.44	0.42
trust in governmental institutions	0.35	0.32	0.34	0.33	0.36	0.37
trust in community institutions	0.59	0.57	0.58	0.58	0.60	0.62
church attendance	0.20	0.16	0.18	0.22	0.20	0.21
school quality	0.82	0.79	0.78	0.82	0.82	0.80
communication with parents	0.51	0.46	0.46	0.53	0.57	0.54
spiritual influence	0.42	0.31	0.45	0.45	0.43	0.42
sense of wellbeing	0.81	0.85	0.79	0.82	0.89	0.86
Risk Factors Average	**0.54**	**0.53**	**0.57**	**0.62**	**0.43**	**0.35**
parental influence (alcohol & smoking)	0.43	0.43	0.41	0.40	0.38	0.39
social exclusion	0.60	0.39	0.68	1.00	1.00	0.00
limited access to healthcare	0.62	0.63	0.64	0.92	0.01	0.39
parental response to misbehavior	0.26	0.38	0.28	0.25	0.22	0.24
parental response to good behavior	0.59	0.58	0.58	0.52	0.47	0.49
rural	0.48	0.50	0.61	0.50	0.33	0.32
parental education	0.78	0.77	0.80	0.72	0.60	0.60
household ownership goods	0.83	0.84	0.84	0.80	0.74	0.72
monthly earnings HH heads	0.85	0.85	0.86	0.83	0.74	0.73

A.4 (cont.)

n	603	363	932	957	947	1178
Risk Type	III	III	III	I	I	-
Percent	12.1%	7.3%	18.7%	19.2%	19.0%	23.7%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.5. Cluster analysis results: Mexican Males Aged 18 to 24

Cluster name	**idle**	**drinkers**	**working dads**	**loners**	**resilient loners**	**advantaged students**
age	20.58	19.60	22.28	21.19	20.96	20.15
Behaviors Average	**0.47**	**0.56**	**0.64**	**0.58**	**0.63**	**0.69**
older age when started working	0.42	0.30	0.30	0.32	0.40	0.46
not idle	0.00	0.98	1.00	1.00	1.00	1.00
no early school dropout	0.27	0.51	0.11	0.00	0.25	1.00
literacy	0.99	0.90	1.00	1.00	1.00	1.00
older age at onset of sexual activity	0.66	0.69	0.50	0.62	0.71	0.68
older age at first pregnancy/parenthood	0.91	0.92	0.69	0.92	0.95	0.98
married	0.14	0.14	0.99	0.05	0.06	0.01
participate in activities	0.19	0.18	0.18	0.20	0.24	0.31
registered to vote	0.95	0.95	0.98	0.96	0.96	0.93
working	0.00	0.81	0.99	0.91	0.78	0.56
in school	0.00	0.30	0.07	0.20	0.38	0.74
years of education completed	0.36	0.34	0.37	0.35	0.44	0.69
safe sex	0.57	0.61	0.22	0.51	0.63	0.64
low number of sexual partners in past year	0.89	0.90	0.87	0.87	0.90	0.89
has a child	0.16	0.14	0.71	0.15	0.11	0.03
attitude towards drugs	0.82	0.84	0.86	0.82	0.85	0.77
attitude towards alcohol	0.73	0.06	1.00	1.00	1.00	1.00
Protective Factors Average	**0.54**	**0.53**	**0.53**	**0.55**	**0.55**	**0.59**
relationship with father	0.46	0.45	0.48	0.47	0.49	0.50
relationship with mother	0.40	0.40	0.43	0.42	0.43	0.48
connected	0.96	0.95	0.99	1.00	0.81	1.00
live with both parents	0.75	0.73	0.34	0.73	0.79	0.83
trust in governmental institutions	0.34	0.36	0.35	0.34	0.38	0.36
trust in community institutions	0.59	0.60	0.62	0.61	0.63	0.64
church attendance	0.30	0.28	0.31	0.30	0.31	0.31
school quality	0.79	0.76	0.78	0.80	0.80	0.81
communication with parents	0.47	0.45	0.48	0.48	0.49	0.53
spiritual influence	0.19	0.20	0.19	0.17	0.20	0.19
sense of wellbeing	**0.82**	**0.83**	0.91	0.85	0.84	0.89
Risk Factors Average	**0.58**	**0.56**	**0.50**	**0.51**	**0.60**	**0.45**
parental influence (alcohol & smoking)	0.66	0.66	0.68	0.74	0.73	0.72
social exclusion	0.37	0.36	0.10	0.00	1.00	0.16
limited access to healthcare	0.61	0.52	0.38	0.46	0.40	0.44
parental response to misbehavior	0.28	0.29	0.31	0.29	0.24	0.22
parental response to good behavior	0.59	0.60	0.59	0.60	0.56	0.54
rural	0.48	0.44	0.38	0.36	0.38	0.24
parental education	0.74	0.73	0.66	0.73	0.69	0.54
household ownership goods	0.78	0.80	0.80	0.76	0.75	0.66

A.5 (cont.)

monthly earnings household heads	0.84	0.83	0.80	0.82	0.81	0.71
n	743	1511	837	1778	1335	1142
Risk Type	III	III	III	II	I	None
Percent	10.1%	20.6%	11.4%	24.2%	18.2%	15.5%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.6. Cluster analysis results: Mexican Females Aged 18 to 24

Cluster name	**at risk, idle married mothers**	**at risk, idle, and excluded married mothers**	**poor working early dropouts**	**somewhat advantaged married moms**	**resilient poor**	**advantaged students**
age	21.35	21.09	21.27	21.48	19.02	20.27
Behaviors Average	**0.50**	**0.51**	**0.63**	**0.63**	**0.62**	**0.72**
older age when started working	0.53	0.56	0.47	0.46	0.48	0.63
not idle	0.00	0.07	1.00	1.00	0.95	0.89
no early school dropout	0.07	0.19	0.00	0.10	0.62	1.00
literacy	0.99	0.92	1.00	0.98	1.00	1.00
older age at onset of sexual activity	0.58	0.67	0.85	0.70	0.92	0.91
older age at first pregnancy/parenthood	0.60	0.68	0.87	0.76	0.97	0.96
married	0.63	0.40	0.12	0.45	0.03	0.05
participate in activities	0.11	0.11	0.18	0.18	0.22	0.29
registered to vote	0.98	0.98	0.97	0.98	0.97	0.97
working	0.00	0.05	0.83	0.82	0.60	0.40
in school	0.00	0.02	0.29	0.26	0.48	0.69
years of education completed	0.29	0.29	0.35	0.39	0.40	0.71
safe sex	0.36	0.50	0.75	0.54	0.88	0.86
low number of sexual partners in past year	0.90	0.93	0.96	0.93	0.98	0.98
has a child	0.68	0.56	0.23	0.40	0.05	0.06
attitude towards drugs	0.96	0.95	0.93	0.90	0.92	0.88
attitude towards alcohol	0.76	0.73	1.00	0.93	0.00	0.99
Protective Factors Average	**0.53**	**0.51**	**0.57**	**0.56**	**0.58**	**0.61**
relationship with father	0.43	0.43	0.46	0.45	0.46	0.51
relationship with mother	0.37	0.36	0.41	0.41	0.41	0.48
connected	1.00	0.64	0.85	1.00	1.00	0.95
live with both parents	0.26	0.52	0.74	**0.47**	0.80	0.80
trust in governmental institutions	0.35	0.35	0.35	0.35	0.34	0.36
trust in community institutions	0.61	0.61	0.62	0.62	0.61	0.63
church attendance	0.37	0.39	0.39	0.36	0.41	0.39
school quality	0.80	0.76	0.81	0.81	0.80	0.82
communication with parents	0.49	0.48	0.54	0.55	0.54	0.60
spiritual influence	0.18	0.20	0.21	0.22	0.22	0.25
sense of wellbeing	0.89	0.82	0.82	0.90	0.82	0.89
Risk Factors Average	**0.51**	**0.65**	**0.59**	**0.42**	**0.54**	**0.46**
parental influence (alcohol & smoking)	0.47	0.49	0.55	0.53	0.50	0.59
social exclusion	0.00	0.97	0.80	0.01	0.52	0.42
limited access to healthcare	0.51	0.65	0.57	0.14	0.54	0.36
parental response to misbehavior	0.36	0.33	0.28	0.30	0.26	0.21
parental response to good behavior	0.58	0.59	0.55	0.55	0.54	0.49
rural	0.45	0.54	0.40	0.26	0.40	0.24

A.6 (cont.)

parental education	0.71	0.75	0.74	0.63	0.72	0.55
household ownership of goods	0.82	0.83	0.79	0.76	0.78	0.68
monthly earnings household heads	0.84	0.86	0.84	0.75	0.82	0.72
n	1810	2324	1693	999	850	1644
Risk Type	III	III	II	II	I	-
Percent	19.4%	24.9%	18.2%	10.7%	9.1%	17.6%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.7. Cluster analysis results: Chilean Males Aged 15 to 17

Cluster name	**at-risk**	**risky sex**	**doing ok**	**loners**	**advantaged youth - some sexual activity**	**advantaged youth**
age	16.31	16.01	15.90	15.95	15.78	15.72
Behaviors/Consequences Average	**0.53**	**0.77**	**0.82**	**0.78**	**0.84**	**0.86**
older age when first working	0.39	0.52	0.67	0.71	0.74	0.97
not idle	0.36	1.00	1.00	1.00	1.00	1.00
no early school dropout	0.00	1.00	1.00	0.98	1.00	1.00
older age at onset of sexual activity	0.60	0.63	0.83	0.73	0.86	0.98
older age at first pregnancy/child	0.97	0.99	1.00	0.98	1.00	1.00
not married	1.00	1.00	1.00	1.00	1.00	1.00
Participate in activities	0.09	0.16	0.18	0.11	0.23	0.14
not a victim	0.92	0.91	0.91	0.91	0.93	0.94
not working	0.64	0.92	0.95	0.92	0.98	0.99
in school	0.00	1.00	1.00	0.98	0.99	1.00
years of education completed	0.31	0.46	0.45	0.46	0.45	0.45
safe sex	0.58	0.62	0.82	0.72	0.87	0.98
does not have a child	0.96	0.99	1.00	0.97	1.00	0.99
desires to vote	0.53	0.55	0.63	0.44	0.64	0.59
Protective Factors Average	**0.59**	**0.62**	**0.75**	**0.50**	**0.76**	**0.66**
relationship with father	0.65	0.62	0.70	0.70	0.92	0.86
relationship with mother	0.85	0.88	0.79	0.86	0.95	0.93
Connected to other adults	0.05	0.02	0.93	0.00	0.00	0.02
Connected	0.72	0.96	1.00	0.00	1.00	1.00
trust in government	0.31	0.33	0.37	0.28	0.44	0.37
trust in community	0.74	0.79	0.83	0.74	0.84	0.81
Spirituality	0.26	0.21	0.34	0.23	0.97	0.22
Risk Factors Average	**0.35**	**0.31**	**0.27**	**0.26**	**0.25**	**0.24**
poor family cohesion	0.38	0.38	0.28	0.35	0.19	0.26
abuse in the home	0.09	0.09	0.00	0.01	0.00	0.01
substance abuse in the home	0.08	0.19	0.00	0.00	0.00	0.00
social exclusion	0.13	0.01	0.00	0.11	0.00	0.17
community violence	0.15	0.21	0.25	0.18	0.29	0.19
felt discriminated against	0.61	0.63	0.51	0.50	0.46	0.48
live in a rural area	0.27	0.13	0.19	0.08	0.13	0.11
indigenous	0.20	0.12	0.13	0.10	0.17	0.08
low economic class (nse)	0.85	0.71	0.74	0.71	0.67	0.63
low parental education	0.75	0.59	0.61	0.59	0.56	0.50
N	75	359	88	105	126	265
Risk Level	III	II	I	I	-	-
Percent of sample	7.4%	35.3%	8.6%	10.3%	12.4%	26.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.8. Cluster analysis results: Chilean Females Aged 15 to 17

Cluster name	at-risk	high risk but connected	loners	advantaged youth	connected
age	16.26	16.15	15.91	15.94	16.00
Behaviors/Consequences Average	**0.44**	**0.79**	**0.81**	**0.83**	**0.82**
older age when started working	0.59	0.78	0.84	0.86	0.88
not idle	0.07	1.00	1.00	1.00	0.99
no early school dropout	0.01	1.00	1.00	1.00	1.00
older age at onset of sexual activity	0.37	0.67	0.84	0.89	0.87
older age at first parenthood	0.61	0.99	0.96	0.99	0.99
not married	0.98	1.00	1.00	1.00	1.00
participate in activities	0.02	0.12	0.08	0.11	0.11
not a victim	0.92	0.91	0.94	0.93	0.92
not working	0.93	0.93	0.99	0.99	0.98
in school	0.00	1.00	1.00	1.00	0.99
years of education completed	0.31	0.45	0.47	0.46	0.47
safe sex	0.28	0.63	0.82	0.88	0.84
does not have a child	0.51	0.98	0.95	0.98	0.99
desires to vote	0.62	0.55	0.50	0.55	0.49
Protective Factors Average	**0.61**	**0.60**	**0.55**	**0.66**	**0.74**
relationship with father	0.59	0.55	0.57	0.75	0.71
relationship with mother	0.75	0.71	0.85	0.91	0.86
connected to other adults	0.12	0.20	0.02	0.00	1.00
connected	0.84	1.00	0.41	1.00	1.00
trust in government	0.38	0.24	0.30	0.31	0.29
trust in community	0.79	0.70	0.70	0.81	0.77
spirituality	0.28	0.36	0.32	0.42	0.42
Risk Factors Average	**0.39**	**0.48**	**0.34**	**0.27**	**0.29**
poor family cohesion	0.39	0.69	0.39	0.33	0.38
abuse in the home	0.14	1.00	0.05	0.00	0.00
substance abuse in the home	0.13	0.20	0.05	0.05	0.06
social exclusion	0.55	0.10	0.53	0.00	0.07
community violence	0.25	0.29	0.34	0.31	0.33
felt discriminated against	0.52	0.82	0.55	0.51	0.58
live in a rural area	0.25	0.18	0.08	0.12	0.10
indigenous	0.11	0.20	0.12	0.10	0.08
low economic class (nse)	0.85	0.73	0.73	0.69	0.71
low parental education	0.72	0.59	0.59	0.56	0.58
n	95	60	133	661	90
Risk Level	III	II	I	-	-
Percent of sample	9.1%	5.8%	12.8%	63.6%	8.7%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.9. Cluster analysis results: Chilean Males Aged 18 to 24

Cluster name	idle dropout (dads)	idle high school graduates	working (fathers) - socially excluded	connected	not well protected	advantaged	advantaged students
age	20.99	19.98	21.02	20.65	20.43	20.41	21.72
Behaviors/Consequences Average	**0.37**	**0.37**	**0.48**	**0.48**	**0.48**	**0.51**	**0.56**
older age when started working	0.38	0.60	0.48	0.49	0.53	0.57	0.56
not idle	0.62	0.00	0.80	0.79	0.79	0.93	0.81
no early school dropout	0.28	1.00	0.80	0.85	0.99	0.98	0.93
older age at onset of sexual activity	0.41	0.54	0.53	0.48	0.47	0.51	0.48
older age at first parenthood	0.90	0.97	0.90	0.96	0.96	0.95	0.98
married	0.00	0.00	0.08	0.04	0.02	0.05	0.00
participate in activities	0.08	0.11	0.07	0.16	0.14	0.12	0.12
registered to vote	0.02	0.01	0.14	0.17	0.01	0.01	1.00
not victim	0.93	0.92	0.92	0.90	0.90	0.91	0.91
working	0.53	0.00	0.57	0.42	0.38	0.39	0.38
in school	0.13	0.00	0.28	0.44	0.52	0.64	0.56
years of education completed	0.43	0.61	0.55	0.61	0.62	0.65	0.68
safe sex	0.22	0.37	0.35	0.33	0.30	0.35	0.32
has a child	0.22	0.06	0.25	0.12	0.09	0.13	0.06
Protective Factors Average	**0.58**	**0.66**	**0.62**	**0.78**	**0.50**	**0.63**	**0.66**
relationship with father	0.56	0.85	0.69	0.74	0.61	0.68	0.72
relationship with mother	0.81	0.86	0.85	0.87	0.79	0.87	0.87
connected to other adults	0.02	0.00	0.02	1.00	0.00	0.03	0.00
connected	0.75	1.00	0.81	1.00	0.00	1.00	0.85
trust in government	0.27	0.34	0.29	0.33	0.28	0.29	0.39
trust in community	0.72	0.81	0.73	0.75	0.72	0.74	0.76
spirituality	0.19	0.32	0.28	0.39	0.21	0.27	0.35
Risk Factors Average	**0.36**	**0.25**	**0.38**	**0.27**	**0.28**	**0.25**	**0.25**
poor family cohesion	0.42	0.25	0.31	0.31	0.41	0.31	0.29
abuse in the home	0.05	0.00	0.02	0.00	0.09	0.08	0.01
substance abuse in the home	0.42	0.00	0.04	0.04	0.15	0.03	0.08
social exclusion	0.09	0.00	1.00	0.13	0.01	0.01	0.00
community violence	0.18	0.18	0.23	0.19	0.20	0.20	0.18
felt discriminated against	0.55	0.47	0.56	0.56	0.55	0.54	0.59
live in a rural area	0.22	0.16	0.17	0.09	0.09	0.10	0.11
indigenous	0.14	0.10	0.17	0.12	0.09	0.09	0.16
low economic class (nse)	0.83	0.75	0.74	0.68	0.69	0.65	0.59
low parental education	0.71	0.62	0.58	0.59	0.55	0.53	0.47
n	237	111	169	117	137	576	134
Risk Level	III	III	III	II	I	-	-
Percent of sample	16.0%	7.5%	11.4%	7.9%	9.3%	38.9%	9.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

A.10. Cluster analysis results: Chilean Females Aged 18 to 24

Cluster name	early drop out mothers	resilient	spiritual	connected	advantaged students
age	21.74	20.80	20.83	20.86	20.49
Behaviors/Consequences Average	**0.31**	**0.44**	**0.45**	**0.48**	**0.56**
older age when first started working	0.54	0.57	0.60	0.60	0.71
not idle	0.12	0.55	0.36	0.63	0.93
no early school dropout	0.00	0.73	0.92	0.88	1.00
older age at onset of sexual activity	0.38	0.52	0.58	0.58	0.69
older age at first parenthood	0.50	0.75	0.78	0.80	0.93
married	0.30	0.05	0.18	0.09	0.01
participate in activities	0.02	0.04	0.04	0.07	0.07
registered to vote	0.05	0.05	0.05	0.09	0.16
not victim	0.94	0.91	0.93	0.93	0.93
working	0.12	0.24	0.24	0.33	0.29
in school	0.00	0.38	0.16	0.36	0.75
years of education completed	0.36	0.56	0.60	0.61	0.68
safe sex	0.12	0.30	0.34	0.39	0.53
has a child	0.88	0.46	0.45	0.39	0.15
Protective Factors Average	**0.58**	**0.59**	**0.61**	**0.72**	**0.64**
relationship with father	0.57	0.53	0.64	0.61	0.72
relationship with mother	0.72	0.73	0.84	0.82	0.86
connected to other adults	0.02	0.13	0.00	1.00	0.01
connected	0.86	0.89	0.79	1.00	1.00
trust in government	0.33	0.24	0.27	0.30	0.30
trust in community	0.75	0.73	0.75	0.74	0.76
spirituality	0.30	0.32	**0.37**	0.34	0.33
Risk Factors Average	**0.36**	**0.48**	**0.31**	**0.30**	**0.25**
poor family cohesion	0.32	0.59	0.31	0.30	0.30
abuse in the home	0.01	0.56	0.02	0.01	0.00
substance abuse in the home	0.00	0.67	0.01	0.01	0.00
social exclusion	0.58	0.33	0.46	0.34	0.02
community violence	0.32	0.29	0.27	0.29	0.29
felt discriminated against	0.52	0.70	0.55	0.58	0.51
live in a rural area	0.23	0.16	0.15	0.11	0.08
indigenous	0.13	0.10	0.08	0.09	0.10
low economic class (nse)	0.80	0.75	0.71	0.70	0.65
low parental education	0.66	0.62	0.58	0.56	0.52
n	209	210	582	160	622
Risk Level	III	II	II	II	-
Percent of sample	11.7%	11.8%	32.6%	9.0%	34.9%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise.

REFERENCES

Afridi, Farzana. 2010. "Child welfare programs and child nutrition: Evidence from a mandated school meal program in India," *Journal of Development Economics*, 92 (2010) 152-165.

Aizer, Anna. 2004. "Home Alone: Supervision After School and Child Behavior." *Journal of Public Economics*, 88(9), 1835-1848.

Alderman, Harold, John Hoddinott, and Bill Kinesy. 2006. "Long term consequences of early childhood malnutrition," *Oxford Economic Papers*, 58(2006): 450-474.

Attanasios, Orazio, Costas Meghir, and Ana Santiago. 2005. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progressa." Mimeo, Center for the Evaluation of Development Policies, Institute for Fiscal Studies, EWP05/01, London.

Attanasio, Orazio, Erich Battistin, Emla Fitzsimons, Alice Mesnard and Marcos Vera-Hernández. 2005. "How effective are conditional cash transfers? Evidence from Colombia" *Institute for Fiscal Studies Briefing Note No. 54*.

Ayalew, Tekabe. 2005. "Parental Preference, Heterogeneity, and Human Capital Inequality," *Economic Development and Cultural Change*, 53, 381-407.

Baranowski T, D. Sprague, JH Baranowski, and JA Harrison. 1991. "Accuracy of Maternal Dietary Recall for Preschool Children," *Journal of the American Dietetic Association* 91(6): 669-74.

Bartlett, Robin, Diane Holditch-Davis and Michael Belyea. 2005. "Clusters of Problem Behaviors in Adolescents," *Research in Nursing and Health*, 28, 230-239.

Becker, Gary S., and H. Gregg Lewis. 1973. "On the interaction between the quantity and quality of children," *Journal of Political Economics*, 82(2):279–288.

Becker, Gary and Kevin J. Murphy. 1988. "Theory of Rational Addition", *Journal of Political Economy*, 96, 675-700.

Becker, Gary S., and Nigel Tomes. 1976. "Child endowments and the quantity and quality of children," *Journal of Political Economics*, 84(2):S143–S162.

Behrman, Jere R. 1997. "Intrahousehold Distribution and the Family." In *Handbook of Population and Family Economics Volume 1A*, editors Mark R. Rosenzweig and Oded Stark. Amsterdam: Elsevier.

Behrman, Jere R., Robert A. Pollack, and Paul Taubman. 1982. "Parental Preferences and Provision for Progeny," *Journal of Political Economy 90*(1, Feb.), 52-73.

Behrman, Jere R., Mark R. Rosenzweig. 2004. "Returns to Birthweight," *The Review of Economics and Statistics*, 86(2): 586-601.

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. 1994. "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy,* 102(6), 1131-1174.

Bernat, Debra and Michael Resnick. 2006. "Healthy Youth Development: Science and Strategies", *Journal of Public Health Management Practice*, November (Supplement), S10-S16.

Binder, Melissa. 1998. "Family Background, Gender, and Schooling in Mexico." *Journal of Development Studies*, 35(2), 54-71.

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics*, 120(2), 669-700.

Blakely, Tony, Simon Hales, Charlotte Kieft, Nick Wilson, & Alistair Woodward. 2005. "The Global distribution of risk factors by poverty level", *Bulletin of the World Health Organization*, February, 83 (2).

Blum, Robert. 1998. "Healthy Youth Development as a Model for Youth Health Promotion", *Journal of Adolescent Health*, 22, 368-375.

_____,, Robert., Trisha Beuhring, Marcia Shew, Linda Bearinger, Renée Sieving, and Michael Resnick. 2000. "The Effects of Race/Ethnicity, Income, and Family Structure on Adolescent Risk Behaviors," *American Journal of Public Health*, Volume 90, No 12.

_____, R., Clea McNeely, and James Nonnemaker. 2002. "Vulnerability, Risk and Protection,"*Journal of Adolescent Health 31S, 28-39.*

_____, Robert. 2002. "Adolescent Health in the Caribbean." Unpublished paper, World Bank, Washington DC.

_____, R. and Marjorie Ireland. 2004. "Reducing Risk, Increasing Protective Factors: Findings from the Caribbean Youth Health Survey." *Journal of Adolescent Health* 35, 493-500.

_____, Marjorie Ireland and Sally-Ann Ohene. 2005. **"**The Clustering of Risk Behaviors Among Caribbean Youth**",** *Maternal and Child Health Journal*, 9(1), 91 - 100.

Borghans, Lex, Angela Lee Duckworth, James A. Heckman, and Bas ter Weel. 2008. "The Economics and Psychology of Personality Traits." NBER Working Paper 13810.

Brener, Nancy and Janet Collins. 1998. "Co-occurrence of Health Risk Behaviors Among Adolescents in the United States," *Journal of Adolescent Health*, 22, 209-213.

Bronfenbrenner, U.. 1979. *The ecology of human development.* (Harvard University Press: Cambridge, MA).

Brook, Judith S., David W. Brook, Mario de la Rosa, Martin Whiteman, Erica Johnson, and Ivan Montoya. 2001. "Adolescent Illegal Drug Use: the Impact of Personality, Family, and Environmental Factors" *Journal of Behavioral Medicine*, 24(2), 183-203.

Brook, David W., Judith S. Brook, Zohn Rosen, and Ivan Montoya. 2002. "Correlates of Marijuana Use in Colombian Adolescents: A Focus on the Impact of the Ecological/Cultural Domain," *Journal of Adolescent Health*, 31, 286-298.

Brook, David W., Judith S. Brook, and Tine Pahl. 2002. **"**The Longitudinal Relationship Between Drug Use and Risky Sexual Behaviors Among Colombian Adolescents" *Archives of Pediatric and Adolescent Medicine*, 156(11), 1101-1107.

Butcher, Kristen, and Anne Case. 1994. "The Effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, 104(3, August), 531-63.

CEPAL (Comisión Económica para América Latina y el Caribe). 2004. *La juventud en Iberoamérica: Tendencias y urgencias.* Santiago, Chile: División de Desarrollo Social, CEPAL.

Cunningham, Wendy and William Maloney. 2001. "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis", *Economic Development and Cultural Change*, 131-156.

Cunningham, Wendy, and William Maloney. (forthcoming) "School Dropout and Child Labor over the Business Cycle." Mimeo. Background paper for *Youth At Risk in Latin America and the Caribbean Regional Study,* World Bank, Washington, DC.

Cunningham, Wendy, Lorena Cohan, Sophie Naudeau, and Linda McGinnis. 2008. *Supporting Youth at-Risk: A Policy Toolkit for Middle Income Countries*, World Bank: Washington, DC.

Dahan, Momi and Alejandro Gaviria. 2003. 'Parental actions and sibling inequality." *Journal of Development Economics*, Vol 72, pp. 281-297.

Dammert, Ana. 2010. "Siblings, Child Labor, and Schooling in Nicaragua and Guatemala." *Journal of Population Economics*, Volume 23 , No. 1, 199-224.

Dasso, Elizabeth. 2006. "Voices of Youth: Youth consultations in Peru." Contribution to the *World Development Report* 2007. http://siteresources.worldbank.org/INTWDR2007/Resources/1489782-1137012196309/2112807-1146686942621/PERU_Final_Report_WDR_2007.pdf

Datar, Ashlesha, Arkadipta Ghosh, and Neeraj Sood. November 2007. "Mortality Risks, Health Endowments, and Parental Investments in Infancy: Evidence from Rural India," NBER Working Paper No. 13649.

Datar, Ashlesha, M. Rebecca Kilburn and David S. Loughran. 2010. "Endowments and Parental Investments in Infancy and Early Childhood," *Demography*, Volume 47-Number 1, February 2010: 145-162.

Duda, R. and P. Hart. 1973. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, NY.

Duque, L. F., J. Klevens, and C. Ramirez. 2003. "Cross-Sectional Survey of Perpetrators, Victims, and Witnesses of Violence in Bogotá, Colombia." *Journal of Epidemiology and Community Health* 57: 355–360.

Duryea, S., A. C. Edwards, and M. Ureta, eds. 2003. *Critical Decisions at a Critical Age: Adolescents and Young Adults in Latin America.* Washington, DC: Inter-American Development Bank.

Edmonds, Eric. 2007. "Understanding Sibling Differences in Child Labor." *Journal of Population Economics*, 19(4), 795-821.

Emerson, Patrick M. and Andre Portela Souza. 2008. "Birth Order, Child Labor, and School Attendance in Brazil." *World Development*, 36(9), 1647-1664.

ESA Consultores. 2001. "Ninos(as) y Adolescentes en una sociedad en desarrollo: los riesgos de la exclusión y la desviación social," Honduras.

Everitt, Brian, Sabine Landau and Morven Leese. 2001. *Cluster Analysis*, 4rd Edition, London: E. Arnold ; New York : Oxford University Press.

Garg, Ashish and Jonathan Morduch. 1998. "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana." *Journal of Population Economics*, 11(4), 471-493.

Glewwe, Paul and Elizabeth M. King. 2001. "The Impact of Early Childhood Nutritional Status on Cognitive Development: Does the Timing of Malnutrition Matter?**"** *World Bank Economic Review* 15: 81-113.

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

Glick, Peter and David E. Sahn. 2010. "Early Academic Performance, Grade Repetition, and School Attainment in Senegal: A Panel Data Analysis." *World Bank Economic Review*, 24(1), 93-120.

Grogger, Jeffrey. 1998. "Market Wages and Youth Crime." *Journal of Labor Economics* 16 (4): 756–791.

Gruber, Jonathan, Editor. 2001. *Risky Behavior Among Youths: An Economic Analysis*,_University of Chicago Press.

Hall, Gillette and Harry Patrinos. 2006. *Indigenous Peoples, Poverty and Human Development in Latin America*, New York: Palgrave Macmillan.

Horowitz, Andrew W. and André Portela Souza. 2011. "The Impact of Parental Income on the Intra-Household Distribution of School Attainment: A measurement strategy and evidence," *The Quarterly Review of Economics and Finance*, Vol. 51, pp 1-18.

Husler, Gebhard, Ronny Blakeney and Egon Werlen. 2005. "Adolescent Risk: The Co-Occurrence of Illness, Suicidality and Substance Use," *Journal of Youth and Adolescence*, Vol. 34, No. 6, 547-557.

Hutz, Claudio and Débora Frizzo M. Silva. 2003. "Developmental Issues and Delinquent Behavior of Brazilian Adolescents," *International Society for the Study of Behavioural Development*, No. 2, Serial No. 42, pp 6-8.

IDDI (Instituto Dominicano de Desarrollo Integral). 2006. "Informe de resultados de sesiones focales de grupos tematicos con jovenes en Republica Dominicana", Unpublished paper, IDDI.

Instituto Cidadania. 2004. *Projecto Juventude: Documento de Conclusão* (Brasilia, Brazil)

Jacoby, Hanan G. 2002. "Is there an intrahousehold 'flypaper effect'? Evidence from a school feeding program," *The Economic Journal*, 112 (January), 196-221.

Kazianga, Harounan. 2004. "Schooling Returns for Wage Earners in Burkina Faso: Evidence from the 1994 and 1998 National Surveys." Yale University Economic Growth Center Discussion Paper No. 892.

Kessler, Daniel. 1991. "Birth Order, Family Size, and Achievement: Family Structure and Wage Determination." *Journal of Labor Economics*, 9(4), 413-426.

Kim, Hisam. 2005. "Parental Investment Between Children With Different Abilities." Unpublished manuscript. University of Wisconsin, Madison, Department of Economics.

Koller, Silvia, J'aims Ribeiro, Elder Cerqueira-Santos, and Normanda Araújo de Morais. 2004. "Joventude Brasileira: Comportamentos de Risco, Fatores de Risco e de Proteçào." October 20, 2004.

Kooreman, Peter. 2000. "The Labeling Effect of a Child Benefit System," *American Economic Review*, 90 (3), 571-583.

Lindberg, Laura Duberstein, Scott Boggess, and Sean Williams. 2000. "Multiple Threats: The Co-Occurrence of Teen Health Risk Behaviors," *The Urban Institute*, Washington, DC.

Loughran, David S., Ashlesha Datar, M. Rebecca Kilburn. 2008. "The Response of Household Parental Investment to Child Endowments," RAND Working Paper WR-404-1, April 2008.

Milligan, G.W. and M.C. Cooper. 1985. "An examination of procedures for determining the number of clusters in a data set." *Psychometrika*. 50(2), 159-179.

Morduch, Jonathan. 2000. "Sibling Rivalry in Africa." *American Economic Association Papers and Proceedings*, 90(2), 405-409.

National Research Council and Institute of Medicine. 2005. *Growing Up Global: The Changing Transitions to Adulthood in Developing Countries,* ed. Cynthia B. Lloyd. Washington, DC: National Academies Press.

Ohene, Sally-Ann, Marjorie Ireland and Robert Blum. 2005. **"**The Clustering of Risk Behaviors Among Caribbean Youth**",** *Maternal and Child Health Journal*, 9(1), 91 - 100.

Ota, Masako and Peter Moffatt. 2007. "The Within-household Schooling Decision: A Study of Children in Rural Andhra Pradesh." *Journal of Population Economics*, 20(1), 223-239.

Parish, William and Robert Willis. 1993. "Daughters, Education, and Family Budgets: Taiwan Experiences." *Journal of Human Resources*, 28(4), 863-898.

Paxson, Christina and Norbert Schady. 2007. "Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health and Parenting." *The Journal of Human Resources*, XLII, 1, 49-84.

Pitt, M., M. Rosenzweig and MD. Nazmul Hassan. 1990. "Productivity, Health and Inequality in the Intra-household Distribution of Food in Low-Income Countries", *American Economic Review*, 80: 1139-56.

Ponce, Juan and Arjun S. Bedi. August 2008. "The Impact of a Cash Transfer Program on Cognitive Achievement: The Bono Desarollo Humano of Ecuador", IZA DP No. 3658

Raven, John E., John C. Raven, and John H. Court. 1998. *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1 General Overview*. Oxford: Oxford Psychologists Press.

Resnick, Michael D., Marjorie Ireland, and Iris Borowsky. 2004. "Youth Violence Perpetration: What Protects? What Predicts? Findings from the National Longitudinal Study of Adolescent Health", *Journal of Adolescent Health*, 35, 424.e1– 424.e10.

Rew, Lynn and Sharon D. Horner. 2003. "Youth Resilience Framework for Reducing Health-Risk Behaviors in Adolescents," *Journal of Pediatric Nursing*, Vol 18, No 6 (December).

Rosenzweig, Mark R. and T. Paul Schultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India." *American Economic Review*, 72(4), 803-815.

Rosenzweig, Mark R. and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy," Review of Economic Studies Limited (2009) **76**, 1149–1174.

Scal, Peter, Marjorie Ireland, and Iris Borowsky. 2003. "Smoking Among American Adolescents: A Risk and Protective Factor Analysis" *Journal of Community Health*, 28(2).

Schady, Norbert and José Rosero. July 2007. "Are Cash Transfers Made to Women Spent Like Other Sources of Income?" *Economics Letters*, Volume 101, Issue 3, December 2008, Pages 246-248.

Schultz, T. Paul. 2004. "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education." *Journal of African Economies*, 13 (Suppl. 2), 95-148.

Skoufias, E. and S. Parker. 2001. "Conditional Cash Transfers and Their Impact on Child Work and School Enrollment: Evidence from the PROGRESA Program in Mexico." *Econometria* 2: 45–96.

Strauss, J., G. Mwabu, and K. Beegle. 2000. "Intra-household Allocations: A Review of Theories and Empirical Evidence." *Journal of African Economies*, 9(0), Supplement 1, 83-143.

Strauss, John, and Duncan Thomas. 1995. "Human Resources: Empirical Modeling of Household and Family Decisions." In *Handbook of Development Economics Volume 3A*, editors Jere Behrman and T.N. Srinivasan. Amsterdam: Elsevier/North Holland.

Wechsler, David. 1974. *Wechsler Intelligence Scales for Children-Revised*, New York: The Psychological Corporation.

Weiss, Maria Jose. 2006. *Pernambuco Young Voices: A Contribution from the Brazilian and Northeastern Youth to the World Development Report,* Development and the Next Generation. Brasilia: World Bank. *http://siteresources.worldbank.org/INTWDR2007/Resources/1489782-1137012196309/2112807-1150737884547/vozes_jovens_ingles.pdf*

World Bank. 2003. *Caribbean Youth Development: Issues and Policy Directions*. Washington, DC. World Bank.

______. 2006a. *World Development Report 2007: Development and the Next Generation*, Washington, DC. World Bank.

———. 2006b. "Documento borrador consulta juvenil en Honduras", Unpublished. *http://siteresources.worldbank.org/INTWDR2007/Resources/1489782-1137012196309/2112807-1138980001537/CAS_Consultations_WDR_Preliminar_Report.pdf*

______. 2007a. *Youth At-Risk in Brazil*, Report No. 32310-BR, Washington, DC. World Bank.

______. 2007b. *Nutritional Failure in Ecuador, Causes Consequences and Solutions.* The World Bank, Washington, D.C.

Zweig, Janine M., Laura Duberstein Lindberg, and Karen Alexander McGinley. 2001. "Adolescent Health Risk Profiles: The Co-Occurrence of Health Risks Among Females and Males," *Journal of Youth and Adolescence*, Vol. 30, No. 6, 707-728.

Zweig, Janine M., Stacey D. Phillips, and Laura Duberstein Lindberg. 2002. "Predicting Adolescent Profiles of Risk: Looking Beyond Demographics", *Journal of Adolescent Health*; 31, 343–353.