Three Essays on the Economics of Skills, Health and Victimization

Duc Thanh Nguyen

A Thesis
in
The Department
of
Economics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Economics) at
Concordia University
Montréal, Québec, Canada

November 2022

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: **Duc Thanh Nguyen**

Entitled: **Three Essays on the Economics of Skills, Health and Victimization**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Economics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____________________ Chair
Dr. Mark K. Watson

____________________ External Examiner
Dr. Xingfei Liu

____________________ External to Program
Dr. Erin Barker

____________________ Examiner
Dr. Ian Irvine

____________________ Examiner
Dr. Jan Victor Dee

____________________ Supervisor
Dr. Tatyana Koreshkova

Approved by ____________________
Dr. Christian Sigouin, Graduate Program Director
Department of Economics

December 21st, 2022 ____________________
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Three Essays on the Economics of Skills, Health and Victimization

Duc Thanh Nguyen, Ph.D.

Concordia University, 2022

Adolescence is a critical development period characterized by biological, cognitive and social-emotional changes that profoundly affect the development of successful lives. In this period, youth is also susceptible to surrounding environments. This thesis explores: (i) the effects of cognitive and noncognitive (or social-emotional) skills in adolescence on education and earnings in adulthood, (ii) the development of the various dimensions of human capital during adolescence and (iii) the consequences that victimization among adolescents has on health and psychosocial outcomes. The thesis uses high-quality longitudinal data from the Young Lives survey in Vietnam that follows children into adulthood. Young Lives provides a rich data set on diversified aspects of children, their families and communities. The diversity of children in terms of various attributes and experiences allows for analyzing causal relations and the dynamics of child development over time in a low-resource setting.

Chapter 1 examines the effects of cognitive and noncognitive skills on college completion decisions and subsequent earnings. I explicitly embed a model of endogenous education decisions and subsequent earnings into a latent factor model. This approach allows for the identification of latent competencies to capture multiple skill dimensions more accurately and correct for measurement errors in observed measures of skills. It also allows for the isolation of the effects of these skills on earnings into components explained by schooling and productivity. Furthermore, this approach solves the endogeneity and reverse causality problems of skills, schooling and earnings by excluding education variables from earnings equations, introducing latent skills and using panel data with skills and outcomes observed at different

times. The findings indicate that both cognitive and noncognitive skills in adolescence are associated with college completion and better earnings in early adulthood. Both types of skills are important in directly determining earnings and indirectly determining earnings through their influence on schooling.

In Chapter 2, I analyze the process by which current levels of cognitive skills, noncognitive skills and health depend on past cognitive and noncognitive abilities, past health, parental cognitive and noncognitive abilities and parental investments. I estimate a dynamic production function model with endogenous parental investments to examine dynamic complementarities and interactions among different inputs and factors in forming child human capital. I use a maximum likelihood approach to estimate the joint distribution of latent factors, which are proxied by observable measures and dynamic CES production functions of human capital. My results show strong effects of parental investments on child cognitive skills, noncognitive skills and health and indicate that parental investments are driven by parental skills and resources. I find evidence that there are dynamic complementarities among the inputs in human capital production, implying that returns to investments are higher for children with better initial conditions. I also find evidence of high levels of self-productivity and the existence of cross-productivity from noncognitive skills and health to cognitive skills and from cognitive and noncognitive skills to health.

Chapter 3 uses a structural model combined with an instrumental variable strategy to deal with the endogeneity and measurement error issues of bullying to study the consequences of peer victimization on a range of health and psychosocial indicators. The findings indicate that peer victimization has strong effects on subjective well-being, alcohol consumption and emotional and mental distress of children. These results are consistent with evidence from both developed and developing countries that bullying has substantial consequences on health risks and psychosocial outcomes. I do not find evidence of associations between bullying victimization and self-rated health.

Acknowledgments

My completion of this Ph.D. program and thesis would have been impossible without the support of a rich network of mentors, friends, and family.

I want to express my deepest gratitude to my supervisor and mentor, Dr. Tatyana Koreshkova, for her continuous support and guidance every step along the way. I want to thank her for her patience, kindness, availability and encouragement. Her insightful advice and sharp comments helped me conceptualize and implement my research papers.

I also wish to offer my special thanks to the members of my thesis committee, Dr. Mark K. Watson, Dr. Xingfei Liu, Dr. Erin Barker, Dr. Ian Irvine and Dr. Jan Victor Dee for their precious time and their invaluable comments and suggestions.

My appreciation extends to my teachers and colleagues at the Department of Economics for creating an extraordinary academic and research environment. My special thanks go to Dr. Paul Gomme for organizing the Macroeconomics Group sections and giving insightful but fun remarks and to Dr. Damba Lkhagvasuren for his helpful comments, suggestions and encouragement.

I wish to thank the former and current Graduate Program Directors, Dr. Effrosyni Diamantoudi, Dr. Szilvia Pápai, Dr. Damba Lkhagvasuren, and Dr. Christian Sigouin and the former and current Chairs, Dr. Greg LeBlanc, and Dr. Jorgen Hansen for their invaluable support.

My sincere gratitude goes to all the staff members of the Department of Economics, especially Ms. Elise Melancon, Ms. Lucy Gilson, Ms. Lise Gosselin, Ms. Bonnie Janicki, Ms. Kelly Routly, Ms. Domenica Barreca, Ms. Sandra Topisirovic, Ms. Melissa Faisal, and Ms. Emilie Martel for their kindness, patience and help.

My special thanks also go to my colleagues at the National Economics Universiy in Hanoi, especially Dr. Chuong, Dr. Cuong, Dr. Tho, Mrs. Huong, Mr. Hung, Mrs. Lan, Dr. Van, Dr. Yen, Ms. Lien, Dr. Thanh, Dr. Loi, Dr. Duc for their great support and encouragement.

I wish to express my sincere thanks to Mr. Duc Hong Le, Mrs. Kim Yen Le and their family for their unconditional support and encouragement. I have been incredibly lucky to have them here in Montreal; they have always been ready and helped me with anything I need. I am also thankful to my family friends: Hiep Trang, Tuan Linh, Bach Phuong, Tri Phuong, Kevin, Anh Thien, Nguyen Tuyen.

I am also very grateful to my family – immediate, extended, and in-laws. They have been incredibly patient and supportive. My parents constantly encouraged and supported me in my academic pursuits from an early age and made great sacrifices to give me the best opportunities they could.

Finally, above all, I would like to express my thanks and all my deep gratitude to my wife Xuan Linh and two wonderful daughters – Khanh My and Jenny for their love, encouragement and support. My daughters are the inspiration and motivation for my thesis. To my wife: there are no words to describe the strength, encouragement and support you have given me throughout this long journey. You have made enormous sacrifices for our family and me.

♡Gửi tặng vợ và hai con yêu dấu - Khánh My, Jenny ♡

Contents

List of Figures xii

List of Tables xiv

1 Transition from School to Work: The Role of Cognitive and Noncognitive Abilities 1

1.1 Introduction . . . 1
1.2 Literature . . . 4
1.3 Data and Definition . . . 7
1.3.1 Cognitive Skills . . . 8
1.3.2 Noncognitive Skills . . . 10
1.3.3 Data on Education and Earnings . . . 12
1.4 Model . . . 12
1.4.1 Measurement System . . . 14
1.4.2 College Decision . . . 17
1.4.3 Labor Earnings . . . 17
1.4.4 Estimation . . . 19
1.5 Results . . . 20
1.5.1 Measurement System . . . 20
1.5.2 Effects of Skills on College Decision . . . 25
1.5.3 Effects of Skills on Earnings . . . 25
1.6 Conclusion . . . 32

Appendices **34**

Appendix A: Description of Variable Construction 34

Appendix B: Exploratory Factor Analysis 38

Appendix C: Model Estimation Procedure 41

Appendix D: Factor Distribution Moments 44

Appendix E: An Alternative Specification for the Measurement System 45

2 Parental Investment and Child Development **50**

2.1 Introduction 50

2.2 Literature Review 52

2.3 Model 53

2.3.1 Dynamics of Skill Formation 53

2.3.2 Measurement System 55

2.3.3 Parental Investments 58

2.4 Estimation 59

2.5 Data and Variables 60

2.5.1 Children's Measures: Cognitive Skills, Noncognitive Skills and Health 61

2.5.2 Parental Cognitive Skills and Noncognitive Skills 64

2.5.3 Parental Investments 66

2.6 Results 67

2.6.1 Measurement System 67

2.6.2 Determinants of Parental Investments in Children 71

2.6.3 Production Functions 74

2.6.3.1 Cognitive Skills 76

2.6.3.2 Noncognitive Skills 83

2.6.3.3 Child Health 88

2.7 Conclusion 94

Appendices **96**

Appendix A: The Construction of Measures of Child's and Parental Noncognitive Skills and Quality of Relationship 96

Appendix B: Factor Moments 101

Appendix C: Marginal Products of the CES Functions 104

Appendix D: Estimates of Production Functions without Endogenous Investments 106

3 The Consequences of Bullying Victimization on Health and Psychosocial Outcomes in Young Children **108**

3.1 Introduction 108

3.2 Literature Review 109

3.3 Data, Definition and Measures 112

3.3.1 Measures of Peer Bullying Victimization 112

3.3.2 Measures of Outcomes 115

3.3.3 Summary Statistics 117

3.4 Conventional Regressions and Endogeneity Issue 117

3.5 Empirical Strategy 120

3.5.1 Identification of Factors 120

3.5.2 Outcomes 125

3.5.3 Instrumental Variables 127

3.5.4 Estimation 128

3.6 Model Results 129

3.6.1 Measurement System 129

3.6.2 The Determinants of Victimization and its Consequences on Outcomes 130

3.7 Conclusion 134

Appendices **136**

Appendix A: Variable Description 136

Appendix B: Factor Distribution 140

Appendix C: Model Estimates without Endogenous Bullying Victimization 141

Bibliography **142**

List of Figures

1.1 Signals and Noises for the Measures 23

1.2 Estimated Cognitive Skill Distribution 24

1.3 Estimated Skill Distribution by Educational Level 24

1.4 Probability of College Completion by Deciles of the Skills 29

1.5 Earnings by Deciles of the Skills 30

1.6 Earnings by College Completion by Deciles of the Skills 31

B.1 Scree Plot - Cognitive Skills 39

B.2 Scree Plot - Noncognitive Skills 40

2.1 Cognitive Skills: Self-productivity 77

2.2 Cognitive skills: Cross-productivity 78

2.3 Cognitive Skills: Self-productivity ($\partial\Theta_t^C/\partial\Theta_{t-1}^C$) 79

2.4 Cognitive Skills: Cross-productivity from Noncognitive Skills to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^{NC}$) 80

2.5 Cognitive Skills: Cross-productivity from Health to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^H$) 80

2.6 Complementarity between Investments and Cognitive Skills 82

2.7 Complementarity between Investments and Cognitive Skills 82

2.8 Noncognitive skills: Self-productivity 84

2.9 Noncognitive Skills: Self-productivity ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{NC}$) 84

2.10 Noncognitive Skills: Cross-productivity from Cognitive Skills and Health . . 85

2.11 Noncognitive Skills: Cross-productivity from Cognitive Skills to Noncognitive Skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^C$) 85

2.12 Noncognitive Skills: Cross-productivity from Health to Noncognitive Skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{H}$) 86
2.13 Complementarity between Investments and Noncognitive Skills 87
2.14 Complementarity between Investments and Noncognitive Skills 88
2.15 Health: Self-productivity 89
2.16 Health: Self-productivity ($\partial\Theta_t^{H}/\partial\Theta_{t-1}^{H}$) 90
2.17 Health: Cross-productivity 90
2.18 Health: Cross-productivity from Cognitive Skills to Health ($\partial\Theta_t^{H}/\partial\Theta_{t-1}^{C}$) . . 91
2.19 Health: Cross-productivity from Noncognitive Skills to Health ($\partial\Theta_t^{H}/\partial\Theta_{t-1}^{NC}$) 91
2.20 Complementarity between Investments and Child Health 92
2.21 Complementarity between Investments and Child Health 93
B.1 Factor Distributions 101
3.1 Outcomes by Deciles of the Victimization Factor Distribution 134

List of Tables

1.1 Scores Used to Measure Cognitive and Noncognitive Skills 11
1.2 Descriptive Statistics 13
1.3 Measurement System 21
1.4 Probability of College Completion as a Function of Skills 26
1.5 Effects of Skills on Earnings 27
1.6 Treatment Effects 28
A.1 Description of Variable Construction 34
B.1 Factor Analysis/Correlation - Cognitive Skills (Principal Component Factors) 38
B.2 Factor Loadings (Pattern Matrix) and Unique Variances - Cognitive Skills . 38
B.3 Factor Analysis/Correlation - Noncognitive Skills (Principal Component Factors) 39
B.4 Factor Loadings (Pattern Matrix) and Unique Variances - Noncognitive Skills 39
D.1 Factor Means, Standard Deviation and Correlation 44
D.2 Mixture Component Means 44
E.1 Measurement System - Correlated Factors 47
E.2 Probability of College Completion as a Function of Skills - Correlated Factors 48
E.3 Effects of Skills on Earnings - Correlated Factors 49
2.1 Key Descriptive Statistics 61
2.2 Observed Variables in the Young Lives Surveys and Corresponding Latent Factors 62
2.3 Key Descriptive Statistics - Child Measures 65

2.4 Key Descriptive Statistics - Parental Skill Measures 66
2.5 Descriptive Statistics: Parental Investments 67
2.6 Measurement System 69
2.7 Estimates of Parental Investment Function 73
2.8 Estimates of Production Functions 75
2.9 Marginal Effects 76
A.1 Construction of Measures of Noncognitive Skills and Quality of Relationship 96
B.1 Factor Means, Standard Deviations and Correlations 102
B.2 Mixture Component Means 103
D.1 Estimates of Production Functions without Endogenous Investments 106
D.2 Marginal Effects - Production Functions without Endogenous Investments . 107
3.1 Measures of Bullying Victimization 114
3.2 Percentage of Children Experiencing Different Forms of Bullying Two or More Times (%) 115
3.3 Summary Statistics by Victimization Status 118
3.4 Conventional Regressions: Association between Overall Victimization Indicator and Outcomes 121
3.5 Conventional Regressions: Association between Different Types of Victimization and Outcomes 122
3.6 Estimated Parameters of Measurement System 130
3.7 Determinants of Bullying Victimization 131
3.8 Consequences of Bullying Victimization on Health and Psychosocial Outcomes 133
A.1 Description of Variables 136
B.1 Factor Means, Standard Deviations and Correlation 140
B.2 Mixture Component Means 140
c.1 Consequences of Bullying Victimization on Health and Psychosocial Outcomes without Instrumental Variables 141

Chapter 1

Transition from School to Work: The Role of Cognitive and Noncognitive Abilities

1.1 Introduction

Schooling has been widely used as a proxy for understanding the impact of skills on labor market outcomes. However, evidence has shown that educational attainment is insufficient to ensure labor market success since it does not necessarily guarantee the required knowledge and skills. Schooling generally raises wages only if it generates skills that create labor productivity and have returns in the labor market (Hanushek, 2002). It is especially true when skills vary widely for children with similar schooling levels (Hanushek and Woessmann, 2008; Singh, 2019). Identifying and understanding the effects of cognitive and noncognitive skills on college education and labor outcomes is vital to better inform policy designs for which skills are rewarded by the labor market and should be improved. However, one of the key challenges in assessing the impact of skills is that it is difficult to reliably capture multiple dimensions of skills with several imperfect candidate measures in surveys. This study aims to model latent skills as a source of unobserved heterogeneity to capture the dimensions of skills more accurately and then assess the effects of these skills on endogenous educational choices and subsequent earnings.

Recently, researchers have paid increasing attention to more accurate skill measures and their impacts on labor market outcomes. These skills can be classified into two categories, cognitive skills and noncognitive skills. Research has shown that cognitive and noncognitive skills play a critical role in educational attainment, labor market success as well as the development of successful lives (Almlund et al., 2011; Cunha and Heckman, 2009; Heckman et al., 2006; Urzúa, 2008; Cawley et al., 2001). Most researchers have focused on the role of cognition in explaining schooling and economic outcomes. Recently, the literature has focused on the impact of noncognitive skills on education and labor productivity, but evidence on the impacts of noncognitive skills on schooling and labor market outcomes is still much limited. While there is consensus on the important role of both types of skills, the relative importance of these skills is still debated. Understanding what skills are required and how specific skills are rewarded by the market is key for designing policies to improve labor productivity and reduce inequality in the long run.

In assessing the determinants and roles of cognitive and noncognitive skills, most research directly uses measured test scores as a proxy for cognitive and noncognitive skills. However, observable tests such as math or IQ tests are not perfect measures of abilities since they suffer from measurement errors. Furthermore, dependence across skills, schooling and labor market outcomes gives rise to the problems of reverse causality and endogeneity.

Vietnam is an interesting case study, not only because panel data is available but also from a policy point of view. Vietnam has achieved significant success in the education sector and its results in students' acquisition of cognitive skills are especially impressive. Vietnam was ranked 12th out of 76 participating countries in the Program for International Student Assessment (PISA) in 2012, above the OECD average as well as above the United States, Australia, and the United Kingdom. However, labor productivity in Vietnam is among the lowest in the Asia-Pacific region (GSO, 2016). One possible explanation for the paradox of very high levels of cognitive skills and educational achievement but very low productivity is low noncognitive skills and the lack of the right skills to match the labor market. Despite impressive achievements in cognitive skills, most of employers faced a shortage of workers with the required skills and unavailability of skilled applicants (Bodewig et al., 2014). Failure in equipping students with adequate noncognitive skills is one prominent weakness of the Vietnamese formal education and training system. The Vietnamese labor force's scores for noncognitive skills are low compared to developing countries (Roseth et al., 2016) and the formal education sector in Vietnam has provided the workforce with low 'soft' skills (Bodewig et al., 2014). In the context of the increased demand for high-skilled workers and the decreased demand for low-skilled workers as a result of rapid technological development and economic integration, strong economic growth will require Vietnam to hinge on the higher productivity of labor, which is substantially influenced by skill shortages and mismatches. While the labour force is equipped with high cognitive but low noncognitive skills, there is a shortage of workers with high noncognitive skills. The shortage of the right skills including non-cognitive skills and mismatches of skills can cause lower productivity. The evidence for the relative importance of skills in determining labor outcomes is thus crucial for suggesting which type of skills should be enhanced more intensively to improve labor outcomes, inclusive of earning capacity or labor productivity.

This study contributes to the limited evidence about the effects of different skills, especially noncognitive skills, on labor market outcomes in developing countries that identify unobserved heterogeneity, correct problems of measurement errors, reverse causality and endogeneity. First, I use a structural latent factor approach to identify true or latent skills as a source of unobserved heterogeneity and their distributions instead of employing (noisy) proxy observed skill measures to capture more accurately multiple cognitive and noncognitive skill dimensions and correct any measurement errors in observed skills. Second, I explicitly embed a model of endogenous education decisions and subsequent earnings into a latent factor model. This strategy solves the endogeneity and reverse causality problems of skills, schooling and earnings by excluding education variables from earnings equations, introducing latent skills and using longitudinal data with skills and outcomes observed at different times. Finally, the model allows schooling decisions to be endogenous to examine the effects of skills on college completion decisions and isolate the effects of these skills on earnings into components explained by schooling and productivity. This approach is crucial to understand the effects of skills and college education on labor outcomes.

1.2 Literature

Panel data are needed to assess the impacts of skills (cognitive and noncognitive skills) on schooling and labor market outcomes. Studies using cross-section data are subject to a risk of reverse causality between skills, schooling and labor market outcomes because they are observed simultaneously, and schooling and work experience may greatly influence skills. However, longitudinal data that measures both cognitive and noncognitive skills during childhood and follows those children into adulthood is rare. Most studies on skills' impacts on schooling and labor market outcomes are on developed countries, in particular, the United States; there is little such evidence in developing countries.

Studies have shown that both cognitive skills and noncognitive skills affect schooling decisions (Almlund et al., 2011) and labor market outcomes (Hanushek and Woessmann, 2008; Almlund et al., 2011; Hanushek, 2009, among others) and cognitive skills have relatively more significant effects than noncognitive skills.

Cunha et al. (2010) show that cognitive and noncognitive skills account for 16 and 12 percent of the variance in educational attainment respectively in the US. Mathematics, reading, and attention skills strongly influence educational success, while noncognitive skills have a limited impact on educational outcomes in the United Kingdom, the United States, and Canada (Duncan et al., 2007).

A large body of evidence has shown that higher cognitive skills measured by test scores such as mathematics, reading, and vocabulary were associated with higher incomes (Murnane et al., 2000; Cawley et al., 2001; Green and Riddell, 2003; Hanushek and Woessmann, 2008;

Heckman et al., 2006; Hanushek and Zhang, 2009; Hanushek et al., 2015). For example, studies in the US find that a one standard deviation increase in the 12th-grade math test score increases annual earnings by 10−15 percent (Murnane et al., 2000; Lazear, 2003). A one standard deviation increase in literacy scores increases earnings by 9.3 percent in a 13-country sample (Hanushek and Zhang, 2009).

Evidence on the relationship between noncognitive abilities, schooling and economic outcomes is much scarcer. A newly growing literature shows that noncognitive competencies have equally important effects as cognitive abilities on schooling and labor market outcomes (Heckman et al., 2006; Cunha and Heckman, 2008; Lindqvist and Vestman, 2011; Almlund et al., 2011). A review of 13 studies by Lindqvist and Vestman (2011) indicates that a one standard deviation increase in noncognitive abilities would increase wages by 4 to 8%. Heckman et al. (2006) show that noncognitive skills are as equally important as cognitive skills in explaining labor wages in the US and increasing noncognitive skills by one standard deviation would increase wages by 11.2%. Using the same data, Heckman et al. (2011) show that skills strongly impact educational attainment and influence earnings through their effects on education, but given years of schooling, noncognitive skills have little direct effects on wages.

Numerous studies directly use test scores as a proxy for cognitive and noncognitive skills (Long et al., 2015; Nordman et al., 2015; Sahn and Villa, 2015; Krishnan and Krutikova, 2013; Díaz et al., 2012 among others). However, ability is multidimensional; it depends not only on skills, but also on other factors and the dimensions of the skill set measured in the survey. Test scores are difficult to measure precisely and are noisy proxies for underlying cognitive and noncognitive abilities; using test scores as a proxy for skills suffers from measurement errors. Furthermore, these studies suffer from the problem of endogeneity and reverse causality. Endogeneity may arise when education is included in earnings equations and reverse causality arises when using cross-section data with skills and outcomes observed simultaneously. Most of the existing evidence address the association rather than causality in making inference on the effects of skills on labor market outcomes. In particular, in this approach, wages depend on schooling choices, observed cognitive and noncognitive test scores and other controls. However, schooling depends on these scores, so schooling is endogenous in wage estimations. Moreover, higher wages could affect schooling choices, and schooling at the time of tests also affects cognitive and noncognitive test scores. This causes a problem of reverse causality between schooling and wages and between schooling and observed test scores; and test scores are endogenous. Omitting schooling variables from wage equations can solve the endogeneity problem of schooling. However, this only allows us to estimate the net effects of skills, but does not fully capture the indirect effects of skills on wages via schooling and test scores are still endogenous.

Heckman et al. (2006), Cunha et al. (2010), and Heckman et al. (2011) develop and use structural measurement frameworks to address measurement errors in measuring skills and the endogeneity of observed skills. Murnane et al. (2001) and Drago (2011) use skills measured before individuals enter the job market to address reverse causality, but they have not addressed measurement errors in test scores and have not taken into account the endogeneity of schooling or investments in estimating the effects of skills.

Evidence on the impacts of cognitive and noncognitive skills for developing countries is rare because surveys measuring both cognitive and noncognitive skills during childhood and following those children into adulthood were unavailable and data sets are mainly cross-sectional data on adults' cognitive skills and noncognitive skills which are primarily related to the job. New data from developing countries allows the exploration of whether skills are as important in labor markets as in developed countries and allow for a causal identification strategy. Studies in developing countries show that both types of skills predict schooling and wages. While most of these studies (Cunningham et al., 2016; Nordman et al., 2015; Sahn and Villa, 2015; Acosta et al., 2015; Krishnan and Krutikova, 2013; Díaz et al., 2012) give the first sets of evidence on how these skills influence schooling decisions and labor market outcomes, they suffer from either the problem of reverse causality or endogeneity bias. Cunningham et al. (2016) and Acosta et al. (2015) deal with measurement errors in measuring skills. However, these two papers use cross-sectional data to examine the effects of cognitive and noncognitive skills of adults on contemporaneous labor market outcomes and they examine the direct impacts of skills. Endogeneity and reverse causality issues between skills and labor market outcomes may still remain in their cross-sectional data studies. Evidence in both developed and developing countries shows that skills affect labor outcomes not only directly but also indirectly through their effects on schooling (Heckman et al., 2006, Heckman et al., 2011; Glewwe et al., 2017). Furthermore, they use skills measured when subjects were adults instead of adolescent skills as in this study.

1.3 Data and Definition

The Vietnam Young Lives survey follows 2000 children in the Younger Cohort and 1,000 children in the Older Cohort. There are five rounds of the survey. The first round was conducted in 2002, at the age of 1 for the Younger Cohort and 8 for the Older Cohort, followed every 4 years until age 15 and 22 for the Younger and Older Cohorts respectively.

The Young Lives survey provides a rich set of data on diversified aspects of children, their families and communities. The survey includes the child, household and commune questionnaires and collects comprehensive information on individual, family, caregiver and parent characteristics and resources, their preferences and feelings as well as schools and communities. This study uses the Young Lives survey data for the Older Cohort that measures both cognitive and noncognitive skills during childhood and the survey follows those children into adulthood with high-quality information on their schooling and labor market outcomes. The data set also provides rich information on children's surrounding environment. This rich available information over time enables us to study skill formation and the impacts of skills on adult outcomes and it allows us to solve the problems of reverse causality and endogenously between skills, schooling and labor market outcomes. In this study, I use skills measured at the age of 15 and assess their impact on the decision to complete college and market outcomes at the age of 22.

Although the Young Lives survey sampling is not designed to be nationally representative of the population, it covers the diversity of children in the country in terms of a wide variety of attributes and experiences. The diversity of children allows us to analyze causal relations and the changing dynamics of childhood welfare over time. As a longitudinal survey, Young Lives is intended to show changes for individuals over time and the impact of earlier circumstances on children's later outcomes. This survey uses a sentinel-site sampling design comprising 20 purposely selected sites chosen to represent diversity, but with a pro-poor bias (Nguyen, 2008). At the site level, children were selected randomly in 2001 such that the data are representative of the birth cohort at each site. My analysis will be conducted for those who completed all cognitive and noncognitive tests and those with complete education and income data at age 22.

As with any longitudinal survey, sample attrition is always an issue. The Young Lives survey is concerned to minimize attrition. The attrition rate for the Young Lives survey for the Older Cohort in Vietnam is 9% since the start of the survey and it is relatively low compared to the other study countries and other longitudinal surveys.[1] Given that I examine cognitive and non-cognitive skills and early earnings, the panel sample is restricted to include those individuals who have complete skills, schooling and earnings data and the final panel consists of 757 observations.[2]

1.3.1 Cognitive Skills

Cognitive skills, also called cognition, cognitive abilities or intelligence, can be simply defined as knowledge and one's ability to acquire new knowledge (Glewwe et al., 2017).

[1] The cumulative attrition rates are 1%, 2.4%, 11.3 % and 9% in Round 2, 3, 4 and 5 respectively.

[2] Similar studies using data from the US, Canada and other countries drop many observations because of the sample restrictions. For example, Prada and Urzúa (2017) end up with 1,022 out of 12,686 observations from an original sample; Heckman et al. (2006) end up with 4680 out of 12,686 observations and Kottelenberg and Lehrer (2019) end up with 1,607 out of 29,687 individuals.

VandenBos (2007), in the American Psychological Association Dictionary of Psychology, defines cognitive skills as “all forms of knowing and awareness, such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving”. Cognitive skills are normally measured by cognitive test scores.

The cognitive development in the Young Lives survey is measured by the test scores in mathematics (Math test), reading comprehension (Cloze test), Peabody Picture Vocabulary Test (PPVT) and Language test (Vietnamese). The Math test and Peabody Picture Vocabulary Test were administered to the Older cohort from Round 1 to Round 3. The Cloze test was added in Round 3; and in Round 4, the PPVT test was replaced by the Language test (in Vietnamese).

Math test: The Math test was administered in Rounds 2 and 3. It includes 29 items on addition, subtraction, multiplication, division, problem-solving, measurement, data interpretation, and basic geometry.

PPVT: The PPVT is a widely-used test of receptive vocabulary. It uses a stimulus word and accompanying pictures to test receptive vocabulary. It has been extensively used to demonstrate the correlation between PPVT scores and cognitive and intellectual ability (Walker et al., 2005). The 204−item PPVT−III was used in Vietnam. Young Lives researchers in each country followed a standard process for adaptation and standardization of the PPVT.

Cloze: The Cloze test was developed to measure verbal skills and reading comprehension. The test includes 24 items that increase in difficulty. Each item consists of a sentence or short paragraph that lack one or more words; children were asked to identify a word that completed the meaning of the sentence or paragraph. A thorough analysis of psychometric characteristics was examined to establish the reliability and validity of all these tests (Crookston et al., 2014).

1.3.2 Noncognitive Skills

Noncognitive skills, also called soft skills, social-emotional skills, noncognitive competencies, noncognitive abilities, or personality traits, can be defined as patterns of feelings, thoughts and behaviors (Borghans et al., 2008; Thiel and Thomsen, 2013).

This study uses three composite indicators designed to access dimensions of self-esteem, self-efficacy and self-respect and inclusion to measure noncognitive skills.

Self-esteem: Self-esteem measures aspects related to pride and it builds on the Rosenberg scale (Rosenberg, 1965). It is related to a person's overall evaluation of their worth. The statements used to measure self-esteem are adapted from the Rosenberg Self-Esteem Scale and focus on different dimensions of the child, such as housing, clothing, work and school.

Self-efficacy: the self-efficacy scale measures aspects related to agency and builds on the Rotter scale (Rotter, 1966). It is related to a person's sense of agency or mastery over his life. The statements used to measure self-efficacy focus on different domains of the child, such as school, work and time use.

Self-respect and inclusion: focus on the social component of self-esteem (Dercon and Krishnan, 2009). The statements used to measure self-respect revolve around the concepts of pride and sense of inclusion.

The measures of self-esteem, self-efficacy and self-respect and inclusion are set on a five-point Likert scale ranging from "strongly disagree" to "strongly agree". Children were read statements and asked whether they strongly disagreed, disagreed, more or less, agreed or strongly agreed with the statements. Negative statements are recoded to reflect positive statements. The self-esteem index includes six items/statements, the self-efficacy index consists of five items and the self-respect and inclusion contain nine items. Each item is standardized with mean 0 and variance 1 and the three composite indices of the noncognitive skills - self-esteem, self-efficacy and self-respect and inclusion - are the average of standardized items used to construct each index. The aim is to place all measurements on the same scale and approximate a measure associated with values of the psychosocial competencies (Dercon and Krishnan, 2009).

The statements used to construct self-esteem, self-efficacy and self-respect and inclusion indices were drawn from the educational psychology literature, they were adapted and extensively tested during piloting to apply for children across different cultures (Dercon and Sánchez, 2013). Self-esteem and self-efficacy are the most popular noncognitive skill measures used in empirical studies (Glewwe et al., 2017). Self-respect and inclusion are related to the self-esteem measure but focus on social and psychosocial aspects of inclusion. Of personality traits, these indices have been found to strongly predict educational achievements and adult social and economic outcomes (Almlund et al., 2011) and they have been used in numerous studies in Vietnam (Dercon and Krishnan, 2009; Dercon and Sánchez, 2013; Sánchez, 2013; Sánchez and Singh, 2018; Singh, 2019).

Table 1.1: Scores Used to Measure Cognitive and Noncognitive Skills

Skills	**No. of items**
Cognitive Skills	
1. Peabody Picture Vocabulary Test (PPVT)	204
2. Mathematics Test (Math Test)	29
3. Cloze Test	24
Noncognitive skills:*	
1. Self-esteem Scale	6
2. Self-efficacy Scale	5
3. Self-respect and Inclusion	9

* The items/statements used to construct composite noncognitive measures (Self-esteem scale, Self-efficacy Scale, Self-respect and Inclusion) are detailed in Appendix A.

An exploratory factor analysis of individual skills is conducted to find whether there are factors that represent cognitive skills and noncognitive skills. The factor analysis results for cognitive and noncognitive skills are provided in Appendix B. The outputs from the factor analysis based on both Kaiser's eigenvalue rule shown in Appendix Tables B.1 and B.3 and scree tests displayed in Figures B.1 and B.2 indicate that there is one factor that should be extracted from the measures of cognitive skills and one factor should be extracted from the measures of noncognitive skills.

1.3.3 Data on Education and Earnings

Young Lives in Vietnam collects detailed data on each child's educational outcomes, including whether the child attended kindergarten, the age when a child started primary school, the highest grade completed, the highest certificate/diploma obtained, current enrolment and detailed educational history. Table 1.2 presents descriptive statistics of the data. The panel sample includes 757 young people aged from 21 to 23 and is balanced between girls and boys, with 51.7% and 48.3% of females and males respectively. They have 0.613 siblings on average. People living in urban areas account for 16.9% of the sample; this reflects the pro-poor sampling approach designed by Young Lives in Vietnam (Nguyen, 2008). Girls score higher than boys in terms of both cognitive and noncognitive abilities. The average hourly earnings are 17.184 Vietnamese Dong, of which boys earn more than girls.[3] Of the total sample, 25.4% completed and obtained a college or university degree. They have 1.833 years of work experience.

1.4 Model

This study is built on the general framework developed by the Roy model (Roy, 1951). It models self-selection into college and potential earnings. Individuals make choices so as to maximize the potential labor outcomes based on their comparative advantages of latent talents that affects their college choices, but may not be directly applied to their job. Individuals choose a college degree based on their expected income and their own abilities.

This model follows Heckman et al. (2006), Carneiro et al. (2003), Cunha et al. (2010). The

[3] The official exchange rate in 2016 is 21,935 Vietnamese Dong per U.S. dollar. Source: https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=VN. Accessed 25 December 2022.

Table 1.2: Descriptive Statistics

	Full	Female	Male
PPVT	-0.091	-0.018	-0.170
	(0.947)	(0.909)	(0.982)
Math	-0.142	0.038	-0.333
	(0.948)	(0.921)	(0.940)
Cloze	-0.083	0.068	-0.247
	(1.014)	(0.927)	(1.078)
Self-Esteem	-0.023	-0.004	-0.043
	(0.639)	(0.677)	(0.597)
Self-Efficacy	-0.030	-0.001	-0.060
	(0.518)	(0.540)	(0.493)
Self-respect and inclusion	-0.021	0.021	-0.066
	(0.564)	(0.551)	(0.575)
Female	0.517	1.000	0.000
	(0.500)	(0.000)	(0.000)
Age (in years)	22.279	22.281	22.277
	(0.336)	(0.344)	(0.328)
Urban	0.169	0.174	0.164
	(0.375)	(0.380)	(0.371)
Number of siblings aged 0-18	0.613	0.708	0.511
	(0.801)	(0.824)	(0.765)
Child's educational aspiration	0.604	0.701	0.500
	(0.489)	(0.459)	(0.501)
Wealth index	0.609	0.624	0.593
	(0.177)	(0.172)	(0.181)
Parental educational level	2.392	2.494	2.284
	(1.069)	(1.079)	(1.050)
College completion	0.254	0.327	0.175
	(0.435)	(0.470)	(0.380)
Monthly earnings (1,000 VND)	3518.837	3271.102	3783.494
	(3242.325)	(3451.621)	(2984.715)
Hourly earnings (1,000 VND)	17.184	16.198	18.236
	(15.802)	(16.528)	(14.939)
Work experience (in Years)	1.833	1.724	1.949
	(1.871)	(1.819)	(1.921)
Observations	757	391	366

Note: Standard deviations in parentheses.

model deals with main problems in estimating the effects of skills on education and income: test scores are just proxies for true abilities and the endogeneity and reverse causality of schooling, skills and income exist in earnings equations. I estimate the model in one step. The observed measures of skills when the children were in Round 3, at the age of 15, are used to estimate the unobserved abilities - the two latent skills by a measurement system. I use the factor approach to identify these factors and their distribution rather than directly use noisy proxy variables or test scores as measures of abilities, as most of the literature does. The underlying cognitive and noncognitive skills are latent rather than observable; they are unobserved to the econometricians and are, in turn, relevant determinants of outcomes, choices and scores. Since the underlying cognitive and noncognitive factors are unobserved, I integrate over the distributions of the two latent factors and examine the effects of skills on the decision whether to complete a college degree or not that the child makes after age 15 and separate the effects of latent skills on labor market outcomes at age 22 into components explained by schooling and skills.

1.4.1 Measurement System

The main challenge in estimating the parameters of this model is that ability is not directly measured. It is challenging to measure ability precisely because of its multidimensional nature. Observed measures of skills or test scores should be considered only as noisy and imperfect proxies for ability, they are based on a noisy signal of one's underlying ability, and thus they suffer from measurement errors. A factor model approach allows for extracting these unobserved skills from a large set of observed data.

Cognitive skills are governed and identified by the latent factor associated with three test scores: Mathematics (T_{math}), Cloze (T_{cloze}), and PPVT (T_{ppvt}) and noncognitive skills are governed and identified by the latent factor associated with self-esteem (T_{ses}), self-efficacy

(T_{sef}), and respect and inclusion (T_{ser}), in the following form:

$$T_{ij} = \alpha_j + \beta_j \theta_i^C + u_{ij} \tag{1.1}$$

for j = {1, 2, 3}= {math, cloze, ppvt}.

$$T_{ik} = \alpha_k + \beta_k \theta_i^{NC} + u_{ik} \tag{1.2}$$

for k = {1, 2, 3} = {ses, sef, ser}.

Where T_{ij} and T_{ik} are 3x1 vectors of the test scores or observed measures j and k of individual i found in the data associated with latent cognitive and noncognitive skills, θ_i^C and θ_i^{NC}, respectively. α_j and α_k are the constants. β_j and β_k are vectors of the factor loadings of the latent skills. u_{ij} and u_{ik} are error terms, which are independent of the associated factors $u_{ij} \perp \theta_i^C$, $u_{ik} \perp \theta_i^{NC}$ and they are mutually independent with an associated distribution $f_{u_h}(.)$ for h = {j, k} = {math, cloze, ppvt, ses, sef, ser}. This independence means that all the correlation in observed measures is captured by latent unobserved factors.

Specifically, the measurement system takes the following form:

$$\begin{aligned}
T_{i,ppvt} &= \alpha_{ppvt} + \beta_{ppvt}\theta_i^C + u_{i,ppvt} \\
T_{i,math} &= \alpha_{math} + \beta_{math}\theta_i^C + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta_{cloze}\theta_i^C + u_{i,cloze} \\
T_{i,ses} &= \alpha_{ses} + \beta_{ses}\theta_i^{NC} + u_{i,ses} \\
T_{i,sef} &= \alpha_{sef} + \beta_{sef}\theta_i^{NC} + u_{i,sef} \\
T_{i,ser} &= \alpha_{ser} + \beta_{ser}\theta_i^{NC} + u_{i,ser}
\end{aligned} \tag{1.3}$$

This structure assumes that the two factors are identified by two different sets of scores. Specifically, only the latent cognitive factor is allowed to affect the individual cognitive skill scores and the latent noncognitive factor is allowed to affect the individual noncognitive skill scores or an increase in the latent cognitive factor would increase the mathematics (T_{math}), Cloze (T_{cloze}), and PPVT (T_{ppvt}) scores and any increase in the latent noncognitive factor would increase self-esteem (T_{ses}), self-efficacy (T_{sef}), and respect and inclusion (T_{ser}). That is, each measure is allocated to a dedicated factor. An alternative setting of the factors where the cognitive scores depend on both the cognitive and noncognitive factors is also considered and the results are quite similar (see Appendix E).

The distributions of the error terms u_{ih}, $f_{u_h}(.)$, are assumed to follow normal distributions with mean zero and variance $\sigma^2_{u_h}$ and let $\theta_i = \{\theta_i^C,\ \theta_i^{NC}\}$, then

$$f(T_{ih}|\theta_i) = \frac{1}{\sqrt{2\sigma^2_{u_h}\pi}} exp\left(-\frac{(T_{ih} - \alpha_h - \beta_h\theta_i)^2}{2\sigma^2_{u_h}}\right) \tag{1.4}$$

The probability of observing measures conditional on θ_i is therefore:

$$f(T_i|\theta_i) = \prod_{h=1}^{h} f(T_{ih}|\theta_i) \tag{1.5}$$

Identification of the factors requires a number of available test scores or skill indexes such that $L >= 2k + 1$, where L is the number of scores and k is the number of factors (Cunha et al., 2010; Carneiro et al., 2003). This condition of identification in this case is satisfied since there are three test scores for the cognitive factor and three indexes for the noncognitive factor. Identification also requires normalizations, I normalize one of the loadings for each factor to one and the remaining coefficients are explained in proportion to the normalized coefficients. Specifically, I set $\beta_{ppvt} = 1$ and $\beta_{ses} = 1$, thus the cognitive skill, θ^C, takes the metrics of PPVT; the noncognitive skill, θ^{NC}, takes the metrics of self-esteem. The locations of the factors are identified by setting one of the constants for each factor to zero. I set $\alpha_{ppvt} = 0$ and $\alpha_{ses} = 0$. By making these normalizations and following the identification strategy of Cunha et al. (2010), the distribution of θ for each latent skill, $F(\theta^C)$ and $F(\theta^{NC})$, and the parameters of interest are identified. I approximate the distributions of the factors by a mixture of normals, as detailed in subsection 1.4.4 below.

1.4.2 College Decision

I now model the effects of skills at age 15 on a subsequent educational decision on whether to complete a college degree or not. Let D^* denote the net latent utility of completing a college education:

$$D_i^* = \alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC} + u_{iD} \tag{1.6}$$

Where X_{iD} is a vector of observed individual and household characteristics affecting the choice; θ^C and θ^{NC} are the unobserved abilities; u^{iD} is an error term; α_D is a vector of the coefficients associated with X_{iD}; β_D^C and β_D^{NC} indicate the effect of the corresponding factors on the decision to complete university.

D is a binary indicator that equals one if the individual completes a college degree and zero otherwise. The choice can be written as:

$$\begin{aligned} &D_i = \mathbb{1}[D_i^* > 0] \\ &or D_i = \mathbb{1}[\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC} + u_{iD} > 0] \end{aligned} \tag{1.7}$$

u_{iD} is assumed to be independent across the individual and household characteristics, factors and all the other errors in the model and logistically distributed. Conditional on the unobservable factor, the probability of observing D_i is:

$$Pr(D_i | X_{iD}, \theta_i) = \frac{(exp(\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC}))^{D_i}}{1 + exp(\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC})} \tag{1.8}$$

1.4.3 Labor Earnings

The model of labor earnings is given by:

$$Y_{iD} = \alpha_{Y_D} X_{iY_D} + \beta_{Y_D}^C \theta_i^C + \beta_{Y_D}^{NC} \theta_i^{NC} + u_{iY_D} \tag{1.9}$$

Where Y_{iD} is hourly earnings for individual i measured at age 22, $D = \{0, 1\}$ corresponding to the specific college decision above. X_{iY_D} is a vector of all other observable controls that impact earnings; θ_i^C and θ_i^{NC} are the unobserved abilities; u_{iY_D} are error terms and follow a normal distribution with mean zero and variance $\sigma^2_{u_{Y_D}}$. The probability density function of Y_{iD} is:

$$f(Y_{iD}|X_{iY_D}, \theta_i) = \frac{1}{\sqrt{2\sigma^2_{u_{Y_D}}\pi}} exp\left(-\frac{(Y_{iD} - \alpha_{Y_D}X_{iY_D} - \beta^C_{Y_D}\theta^C_i - \beta^{NC}_{Y_D}\theta^{NC}_i)^2}{2\sigma^2_{u_{Y_D}}}\right) \tag{1.10}$$

In this model, the latent factors, θ^C and θ^{NC}, reflect unobserved heterogeneity and they are the source of dependence among observed skill measures, schooling decisions, and earnings. Controlling for these latent factors solves the problem of endogeneity arising from the endogeneity of skills and schooling and the reverse causality among skills, schooling and earnings (Heckman et al., 2006). Using latent cognitive and noncognitive skills also solves the measurement error problem. Furthermore, using skills and outcomes observed at different times deters the reverse causality among skills, schooling and labor market outcomes; skill measures are not affected by college degrees.

1.4.4 Estimation

Equations 1.3, 1.7 and 1.9 constitute the following structural model by which the college decision and wage equation are estimated jointly with the measurement system:

$$
\begin{aligned}
&T_{i,math} = \alpha_{math} + \beta^{C}_{math}\theta^{C}_{i} + u_{i,math}\\
&T_{i,cloze} = \alpha_{cloze} + \beta^{C}_{cloze}\theta^{C}_{i} + u_{i,cloze}\\
&T_{i,ppvt} = \alpha_{ppvt} + \beta^{C}_{ppvt}\theta^{C}_{i} + u_{i,ppvt}\\
&T_{i,ser} = \alpha_{ser} + \beta^{NC}_{ser}\theta^{NC}_{i} + u_{i,ser}\\
&T_{i,sef} = \alpha_{sef} + \beta^{NC}_{ses}\theta^{NC}_{i} + u_{i,sef}\\
&T_{i,ses} = \alpha_{ses} + \beta^{NC}_{ses}\theta^{NC}_{i} + u_{i,ses}\\
&D_{i} = \mathbb{1}[\alpha_{D}X_{iD} + \beta^{C}_{D}\theta^{C}_{i} + \beta^{NC}_{D}\theta^{NC}_{i} + u_{iD} > 0]\\
&Y_{iD} = \alpha_{Y_D}X_{iY_D} + \beta^{C}_{Y_D}\theta^{C}_{i} + \beta^{NC}_{Y_D}\theta^{NC}_{i} + u_{iY_D}
\end{aligned}
\tag{1.11}
$$

The distributions of the latent factors may follow many forms and the assumption of the factor distributions is important and must be flexible enough to capture data. I approximate the factor distributions as a mixture of two normals. This assumption ensures flexibility with fewer restrictions on the distributions (Ferguson, 1983; Attanasio et al., 2017). With this assumption, the probability density function of the factor is:

$$f(\theta) = \sum_{c=1}^{2} \tau_c f(\theta|\mu_c, \Omega_c) \tag{1.12}$$

Where μ_c, Ω_c and τ_c are the mean, covariance and the mixture probability of the two normals.

Let Ψ be all the parameters of the model, $\Psi = \{\alpha, \beta, \sigma, \tau_c, \mu_c, \Omega_c\}$, $\theta = \{\theta^C, \theta^{NC}\}$ be the vectors of the cognitive and noncognitive factors, $X = \{X_{i_D}, X_{iY_D}\}$. Thus, from the density and probability functions 1.5, 1.8, 1.10 and 1.12, the full model likelihood function can be derived as:

$$\begin{aligned}
L(\Psi) &= \prod_{i=1}^{N} \iint [f(T_i|\theta^C, \theta^{NC}) \times f(Y_{i,D=1}|X_{iY_{D=1}}, \theta^C, \theta^{NC})^D \times \\
& \quad f(Y_{i,D=0}|X_{iY_{D=0}}, \theta^C, \theta^{NC})^{1-D} \times Pr(D_i|X_D, \theta^C, \theta^{NC})] dF(\theta^C) dF(\theta^{NC}) \\
&= \prod_{i=1}^{N} \iint f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta^C, \theta^{NC}) dF(\theta^C) dF(\theta^{NC}) \\
&= \prod_{i=1}^{N} \int f(T_i, D_i, Y_i|X_{i_D} X_{iY_D}, \theta) dF(\theta) \\
&= \prod_{i=1}^{N} \int f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta
\end{aligned} \tag{1.13}$$

The full model log-likelihood function is

$$\mathcal{L}(\Psi) = \sum_{i=1}^{N} ln \int f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta \tag{1.14}$$

Given the unobservable nature of the factors, the likelihood function is integrated over the distributions of these unobservable factors. I estimate the log-likelihood function 1.14 using maximum likelihood estimation (MLE). I take a one-step estimation procedure using the minorization-maximization algorithm that is presented in Appendix C.

1.5 Results

1.5.1 Measurement System

The estimation results from the measurement system described in Equation set 1.3, α_j, α_k and β_j, β_k, are presented in Table 1.3. The measurement system examines the importance of the given latent skills, θ^C and θ^{NC}, in the six tests. The factor loadings of cognitive and noncognitive skills (β_j, β_k) on respective test scores are all significantly positive, meaning that both latent skills are positively associated with test scores as expected. Latent cognitive ability is more highly associated with the mathematics and reading comprehension (Cloze) test scores, while latent noncognitive ability more highly relates to an individual's self-esteem and self-respect and inclusion indexes. Specifically, a one standard deviation increase in cognitive ability is associated with a 0.664, 0.748 and 0.739 standard deviation increase in the PPVT, Math and Cloze scores respectively and a one standard deviation in noncognitive ability is associated with a 0.389, 0.271 and 0.492 standard deviation increase in individual self-esteem, self-efficacy and self-respect and inclusion indexes respectively.

Table 1.3: Measurement System

	PPVT	Math	Cloze	Self-Esteem	Self-Efficacy	Self Respect and Inclusion
Panel A: Estimated parameters						
Constant	0	-0.037**	0.009	0	-0.014	0.007
		(0.019)	(0.021)		(0.011)	(0.014)
Cognitive	1	1.126***	1.113***	-	-	-
		(0.053)	(0.055)			
Noncognitive	-	-	-	1	0.696***	1.263***
					(0.039)	(0.069)
Panel B: Average Marginal Effects of Factors (AME)[a]						
Cognitive AME	0.664***	0.748***	0.739***	-	-	-
	(0.025)	(0.019)	(0.027)			
Noncognitive AME	-	-	-	0.389***	0.271***	0.492***
				(0.015)	(0.013)	(0.017)
Average value	-0.107	-0.158	-0.110	-0.022	-0.030	-0.021
Panel C: Variance Decomposition						
Signal	0.485***	0.609***	0.513***	0.370***	0.274***	0.762***
	(0.026)	(0.022)	(0.024)	(0.022)	(0.023)	(0.039)
Noise	0.515***	0.391***	0.487***	0.630***	0.726***	0.238***
	(0.026)	(0.022)	(0.024)	(0.022)	(0.023)	(0.039)
N	738	747	739	757	757	757

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

[a] Average marginal effects of a one standard deviation increase of each factor, holding other variables fixed.

To assess the information content contained in each measure from the factors and measurement errors, I calculate the contribution of each factor and measurement error in explaining the variance of the observed measures.

$$P_h^{\theta^k} = \frac{(\beta_h)^2 var(\theta_i^k)}{(\beta_h)^2 var(\theta_i^k) + var(u_{ih})} \tag{1.15}$$

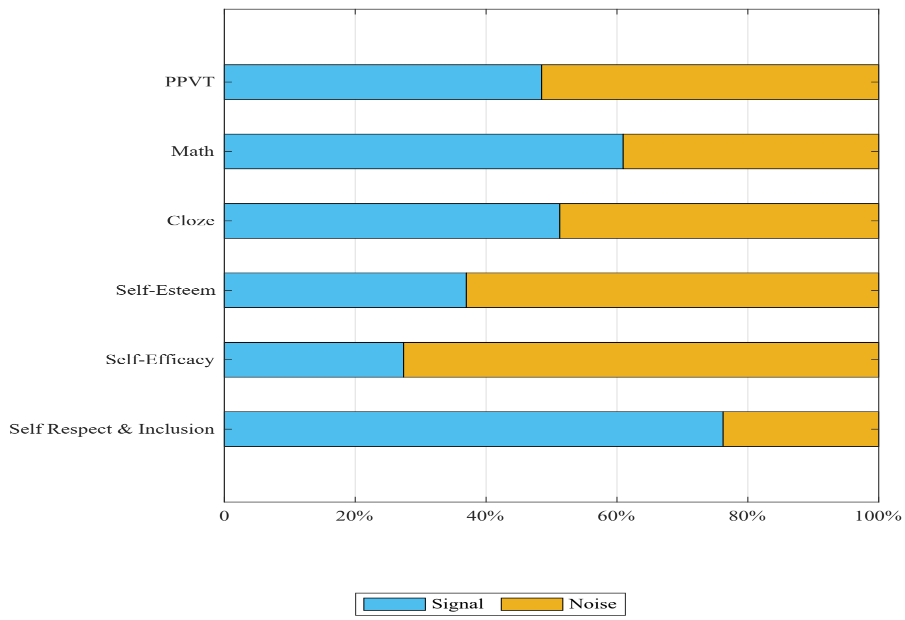
$$P_h^{u_h^k} = \frac{var(u_{ih})}{(\beta_h)^2 var(\theta_i^k) + var(u_{ih})} \tag{1.16}$$

Where $P_h^{\theta^k}$ is the proportion of the variance of the hth observed measures explained by the latent factor k or signal and $P_h^{u_h^k}$ is the variance of the measure explained by the measurement error or noise. $P_h^{u_h^k}$ is the proportion of the hth measure variance that remains unexplained.

Table 1.3 Panel C and Figure 1.1 present the fraction of the variance of each measure explained by each factor (signal) and by the measurement error (noise). It is clear that the measures for each factor contain a substantial amount of information. The cognitive skill factor accounts for an important proportion of the variance of the cognitive measures - from 48.5% to 60.9%. The related measures on noncognitive skills are also informative. From 27.4% to 76.2% of the variance of the noncognitive measures are explained by signal. Although the factors explain an important proportion of the variance of the observed measures, these proportions are far from 100%, 23.8 - 72.6% of the variance of the observed measures remains unexplained and is attributed to measurement errors. This indicates that we could have serious measurement error problems if we use observed measures on their own and demonstrates the importance of the latent factor approach in measuring skills.

The estimated distributions of latent cognitive and noncognitive skills ($f(\theta^C)$, $f(\theta^{NC})$) and the skill distributions by college completion, displayed in Figures 1.2 and 1.3, show that distributions of latent skills are non-normal. These results highlight the importance of

Figure 1.1: Signals and Noises for the Measures



assuming a flexible distribution function for the skill distributions.[4]

Figure 1.3 displays the cognitive and noncognitive skill distributions by college completion. Individuals who completed college education seem to have higher cognitive and noncognitive skills; individuals with a college degree have distributions of both skills lying to the left compared to those without a college education. Although there is a difference in the distributions between those with and without a college education, they show a substantial overlap. Therefore, I will also explore the effects of the variation of skills on the outcomes.

[4] The factor means, standard deviations and correlation of the estimated factors are presented in Appendix D.

Figure 1.2: Estimated Cognitive Skill Distribution

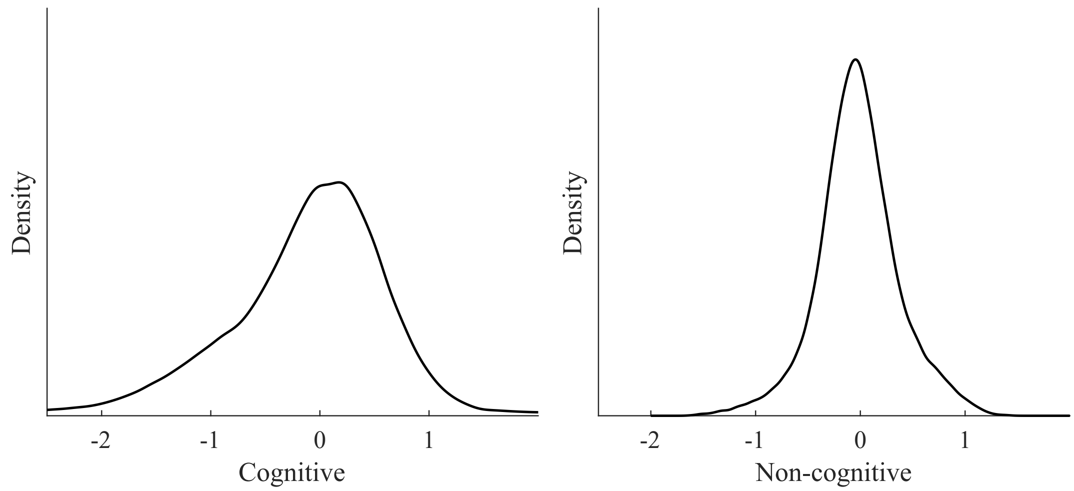
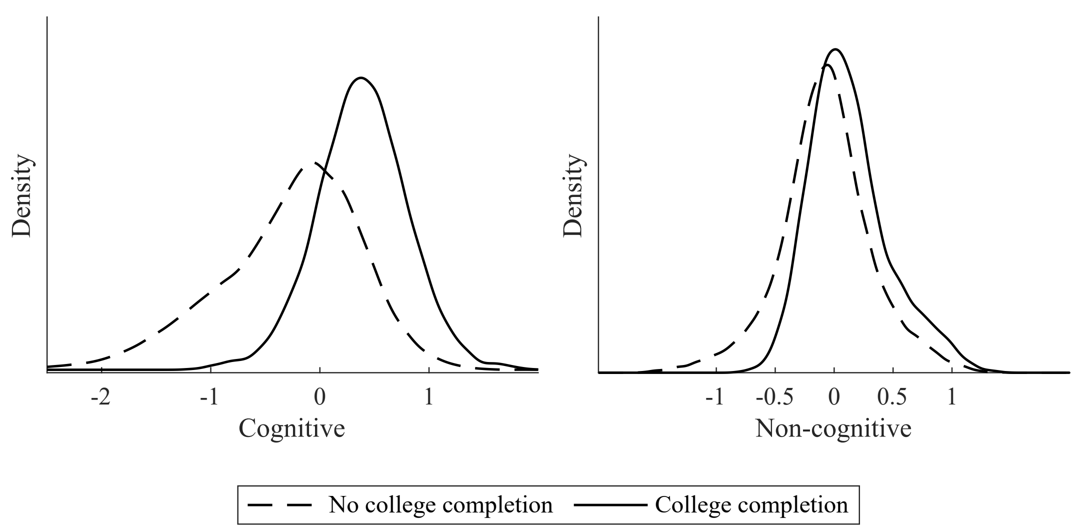


Figure 1.3: Estimated Skill Distribution by Educational Level



1.5.2 Effects of Skills on College Decision

Table 1.4 shows the effects of latent cognitive and noncognitive abilities on the decision to complete college. Both latent cognitive and noncognitive abilities are important determinants of completing college. The probability of completing college increases dramatically with cognitive skills and the effect of noncognitive skills is smaller. Specifically, increasing cognitive abilities by a one standard deviation would increase the likelihood of completing college by 17.2% percent and a one standard deviation increase in noncognitive abilities would lead to an increase in the likelihood of completing college by 2.6% percent.

Table 1.4 also shows the effects of other controlled variables on completing higher education. Girls are more likely than boys to complete college, while a child's living areas (rural/urban) do not affect the probability of completing college. Parental educational levels are important in determining college completion, while wealth does not seem to influence college completion. A youth with few siblings is more likely to obtain a college degree. The results also suggest that the child's aspiration for college education significantly determines educational attainment.

1.5.3 Effects of Skills on Earnings

Table 1.5 reports the estimates of the parameters, α_{Y_D}, $\beta^{C}_{Y_D}$ and $\beta^{NC}_{Y_D}$, from Equation 1.9 for earnings conditional on college completion. $D = 1$ and $D = 0$ indicate those who have obtained and have not obtained a college degree respectively. These results show the importance of cognitive and noncognitive skills after conditioning on college completion.

Noncognitive ability does improve income among college graduates. For this group, a one standard deviation increase in latent noncognitive ability will increase earnings by 4.107VND, which represent an substantial increase of about 17.6% over the average earnings. However, noncognitive ability does not provide any additional rewards for higher earnings among those not completing college. Cognitive ability is significantly associated with higher earnings for

Table 1.4: Probability of College Completion as a Function of Skills

Variables	Coefficients	Average Marginal Effects
Cognitive	1.924***	0.172***
	(0.214)	(0.016)
Noncognitive	0.537***	0.026***
	(0.197)	(0.010)
Female	0.662***	0.082***
	(0.119)	(0.015)
Urban	0.148	0.018
	(0.120)	(0.015)
Number of siblings aged 0-18	-0.753***	-0.084***
	(0.111)	(0.010)
Wealth index	0.077	0.009
	(0.609)	(0.074)
Parental educational level	0.462***	0.059***
	(0.061)	(0.008)
Child educational aspiration	1.237***	0.146***
	(0.182)	(0.019)
Constant	-3.579***	-
	(0.415)	
Baseline probability	0.254	-
N	757	-

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

non-college graduates, but has no statistically significant effects on earnings for those with a college education. A one standard deviation increase in cognitive ability would increase earnings for non-college graduates by 14.7%.

These results may suggest that the effect of cognitive skills on earnings is indirect and operate mainly through educational decisions. It may be due to the fact that I examine earnings early in the career and employees are relatively new to employers. Employers had little information and opportunity to distinguish and reward higher skills, and educational levels are a meaningful signal for judgements and advancement in pay. The results also once again indicate that females earn much less than males for both groups of non-college and college graduates and individuals from urban areas earn more than those from rural areas.

Table 1.5: Effects of Skills on Earnings

Variables	Hourly earnings	
	D = 0	D = 1
Cognitive skills	3.338***	2.717
	(0.530)	(2.456)
Noncognitive skills	1.565	10.551***
	(1.034)	(3.081)
Cognitive AME[a]	2.217***	1.805
	(0.338)	(1.627)
Noncognitive AME[a]	0.609	4.107***
	(0.402)	(1.140)
Female	-3.822***	-2.572
	(0.524)	(1.568)
Urban	3.045***	7.526***
	(0.754)	(1.576)
Experience	2.304***	8.604***
	(0.438)	(2.418)
Experience squared	-0.309***	-1.724**
	(0.056)	(0.698)
Constant	14.726***	13.885***
	(0.662)	(1.676)
Average value	15.116	23.282

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
[a] Average Marginal Effects of Factors.

One important advantage of structural models is the ability to simulate counterfactual outcomes (Heckman et al., 2011; Carneiro et al., 2003). To understand the effect of having a college degree, I calculate the average treatment effect (ATE) of a college degree, the treatment effect on the treated (TOT) and the treatment effect on the untreated (TOU) as follows:

$$
\begin{aligned}
ATE &= E[Y_1 - Y_0|X,\theta] = E[Y_1|X,\theta) - (Y_0|X,\theta)] \\
TOT &= E[Y_1 - Y_0|X,\theta,D=1] = E[Y_1|X,\theta,D=1) - (Y_0|X,\theta,D=1)] \\
TOU &= E[Y_0 - Y_1|X,\theta,D=0] = E[Y_0|X,\theta,D=0) - (Y_1|X,\theta,D=0)]
\end{aligned} \tag{1.17}
$$

Table 1.6 shows the difference between the means of earnings conditioning on the decision to complete a college degree and the respective counterfactual earnings. ATE and TOT are positive and TOU is negative. The results suggest that, on average, young people with a college degree would have higher earnings. Even people with their given background and latent skills who decided not to have a college degree would have higher earnings if they had a college degree. In particular, on average, individuals would increase their hourly earnings by 3,082 (equivalent to 17.9% relative to the mean earnings) if they decided to have a college degree. Conditioning on completing a college degree, the mean of hourly earnings is 6,536 (equivalent to 38% relative to the mean earnings) higher than the mean of hourly earnings that they would have earned if they had decided not to have a college degree. In contrast, the mean of hourly earnings conditional on not completing a college degree is 1,909 (equivalent to 11.1% relative to the mean earnings) lower than the means of counterfactual hourly earnings that they would have earned if they had decided to complete a college degree.

Table 1.6: Treatment Effects

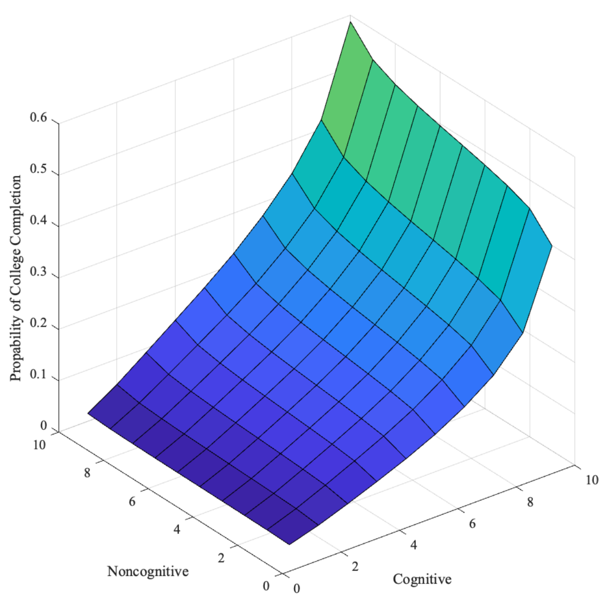
	Estimates
$E[Y_1\|X,\theta) - (Y_0\|X,\theta)]$	3.082*
	(2.200)
$E[Y_1\|X,\theta,D=1) - (Y_0\|X,\theta,D=1)]$	6.536***
	(0.721)
$E[Y_0\|X,\theta,D=0) - (Y_1\|X,\theta,D=0)]$	-1.909
	(2.893)

Notes: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Figures 1.4, 1.5 and 1.6 graphically present how the college decision and earnings vary across deciles of cognitive and noncognitive abilities. In these figures, I present each outcome as a function of deciles of the skill distribution and display the mean value of these outcomes by deciles of the skills.

Figure 1.4 shows the probability of college completion by each decile of the cognitive and noncognitive skill distribution. Both types of skills show strong effects on the probability of completing a college degree. A steeper gradient for cognitive ability shows that its effect on college completion is more important than noncognitive ability. The probability of graduating from a college increases dramatically with cognitive skills while the effect of noncognitive skills is stronger for a higher level of cognitive skills.

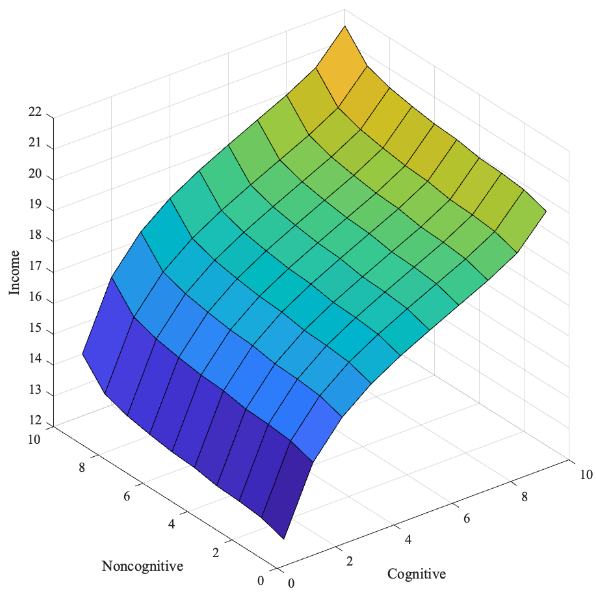
Figure 1.4: Probability of College Completion by Deciles of the Skills



Note: z-axis is the probability of college completion within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively.

Figure 1.5 displays the effects of skills on earnings by deciles of the skill distribution. The effect of cognitive skills is again stronger than noncognitive skills. This result can be explained by the fact that the skills not only have direct effects on earnings, but also have indirect effects on earnings through schooling that generates effects on earnings, while the effect of the cognitive skills on schooling is more important than noncognitive skills.

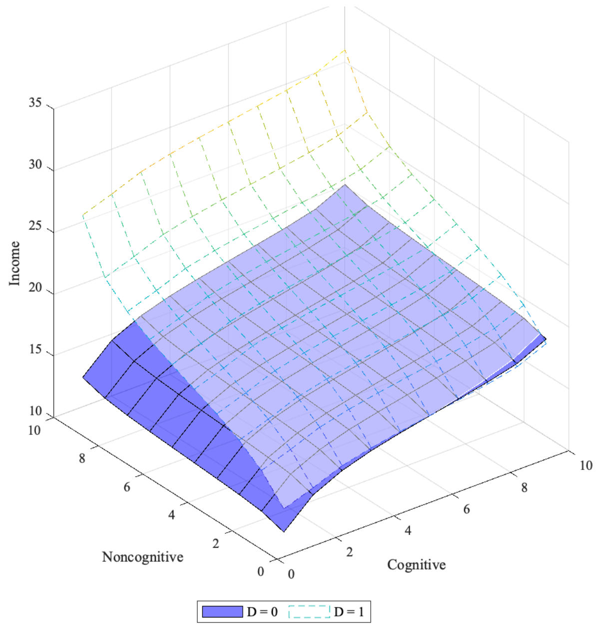
Figure 1.5: Earnings by Deciles of the Skills



Note: z-axis is the mean earnings within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively.

Figure 1.6 shows the effects of skills on earnings by college completion across deciles of skills. Earnings increase across deciles of cognitive and noncognitive abilities from about 10,000 to 20,000VND and from 12,000 to 31,000VND for those without and with a college degree respectively. Noncognitive skills play a significant role in earnings for those with a college degree, while those without a college degree need a certain level of cognitive skills to get higher earnings.

Figure 1.6: Earnings by College Completion by Deciles of the Skills



Note: z-axis is the mean earnings within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively. D = 1: college completion, D = 0 otherwise.

The results share commonalities with the literature. First, both cognitive and noncognitive abilities affect schooling and labor market earnings (Almlund et al., 2011; Hanushek and Woessmann, 2008; Almlund et al., 2011; Hanushek, 2009; Heckman et al., 2006). Second, the effects of these abilities on earnings are mediated by levels of schooling (Heckman et al., 2006; Heckman et al., 2011). Third, our results about the effects of skills on schooling are consistent with Cunha et al. (2010) and Duncan et al. (2007) in the sense that both cognitive and noncognitive skills have effects on schooling, while cognitive skills have stronger effects.

Despite these consistencies, there are certain differences in terms of effects and magnitudes of effects of skills on education decisions and labor market earnings between my findings and the literature. These deferences arise from different reasons. First, the differences in markets, policies and institutions can result in different labor outcomes. Second, different questionnaires, measures to capture the underlying cognitive and noncognitive skills and different methodologies may influence the conclusions about the effects. My findings indicate that noncognitive skills are highly valued, while cognitive skills are not rewarded once students graduate from college. This result is in contrast with Heckman et al. (2006) for the US who found that noncognitive traits have little value, while cognitive skills have a strong effect on earnings for 4-year-college graduates and both skills have strong effects in the 2-year-college market. This reflects the fact in Vietnam that the education system equipped the workforce with a low level of 'soft' skills and there is a high demand for these skills (Bodewig et al., 2014).

Appendix E shows the results for an alternative specification for the measurement system that allows for correlated cognitive and noncognitive factors where each of the cognitive measures depends on both the cognitive and noncognitive factors and the noncognitive measures are a function of the noncognitive factor only. I impose the same normalizations on the scales and locations of the factors. The results in Appendix Tables E.2 and E.3 show few differences in the effects of skills on college completion and earnings between the two specifications. Although noncognitive abilities play a smaller role, but the difference is insignificant.

1.6 Conclusion

This study uses high-quality data from the Vietnam Young Lives survey, Older Cohorts and the latent factor approach with a two-dimensional latent factor structure to examine the roles of both cognitive and noncognitive abilities in explaining schooling decisions and subsequent earnings in Vietnam. The results suggest that both cognitive and noncognitive skills play a role in determining earnings. The analysis shows that the effects of skills on earnings operate not only indirectly through the educational channel but also directly in the labor market. Because of the nature of endogenous schooling decisions, the dynamics in decision making is crucial in investigating the effects of skills on earnings. Among college graduates, noncognitive skills not only directly influence earnings in the labor market but also have indirect effects through educational choices. The results suggest that it is equally important to improve noncognitive skills as cognitive skills.

There is strong evidence showing that human capital is shaped early in the life cycle and skills beget skills in a complementary and dynamic fashion. Child development at an early age has direct long-lasting effects on social and economic outcomes for individuals and society. Policies should give equal attention to improving different dimensions of noncognitive skills in early childhood as with cognitive skills. This is especially true for Vietnam, which achieves impressive results in cognitive skills, but soft skills and labor productivity are relatively low. Skills are affected by a combination of inputs, including individual abilities, family investments, and home, school, and community environments. Therefore, policies should consider a combination of factors, and investments in childhood development are a cost-effective strategy for improving productivity, promoting economic growth, and reducing inequality.

Appendices

Appendix A: Description of Variable Construction

Table A.1: Description of Variable Construction

Variables	Description
Cognitive Skills	
PPVT score	The PPVT is a test of receptive vocabulary. It uses a stimulus word and accompanying pictures to test receptive vocabulary. The PPVT-III with 204 items is used in the Young Lives survey in Vietnam. PPVT scores are standardized scores in The PPVT.
Math score	The mathematics test (Math test) include 29 items on addition, subtraction, multiplication, division, problem-solving, mea- surement, data interpretation, and basic geometry. Math scores are standardized scores in math test.
Cloze test scores	The Cloze test is developed to measure verbal skills and reading comprehension. The test include 24 items that increase in difficulty. Cloze test scores are standardized scores in Cloze test.

Continued on next page

Table A.1: Description of Variable Construction *Continued*

Variables	Description
Noncognitive Skills	
Self-esteem	The self-esteem scale is constructed as the average of the following standardized items/statements (five-point Likert scales). 1. 'I am proud of my clothes'; 2. 'I feel my clothing is right for all occasions'; 3. 'I am proud of my shoes or of having shoes'; 4. 'I am proud because I have the right books, pencils or other equipment for school'; 5. 'I am proud that I have the correct uniform'; 6. 'I am proud of the work I have to do'.
Self-efficacy	The Self-efficacy index is the average of the following standardized items (five-point Likert scales): 1. 'If we try hard we can improve my situation in life'; 2. 'Other people in my family make all the decisions about how we spend my time'; 3. 'I like to make plans for my future studies and work'; 4. 'If we study hard we will be rewarded with a better job in the future'; 5. 'I have no choice about the work I do - I must do this sort of work'.

Continued on next page

Table A.1: Description of Variable Construction *Continued*

Variables	**Description**
Self-respect and Inclusion	This index is the average of the following standardized items (five-point Likert scales): 1. 'When I am at the shops/market I am usually treated by others with fairness and respect'; 2. 'Adults in my community treat me as well as they treat other children of my age'; 3. 'The other children in my class treat me with respect'; 4. 'Other pupils in my class tease me at school'; 5. 'My friends will stand by me during difficult times'; 6. 'I feel I belong at my school'; 7. 'My friends look up to me as a leader'; 8. 'I have people I look up to' 9. 'I have opportunities to develop job skills'.
Other Variables	
Hourly earnings	Hourly earnings from all paid activities by child in the past 12 months.
College completion	Binary variable equal to one if the individual completes a college degree and zero otherwise.
Wealth Index	The wealth index is a composite measure of living standards, it is the average of the three sub-indexes: consumer durable, housing quality and access to service indexes. It takes values from 0 to 1, a higher value reflect a wealthier household.

Continued on next page

Table A.1: Description of Variable Construction *Continued*

Variables	**Description**
Parental educational level	The highest level of education of Parent: 1 = less than primary; 2 = primary; 3 = Lower secondary; 4 = Upper secondary; 5 = post-seconday.

Appendix B: Exploratory Factor Analysis

This appendix provides details of the factor analysis to find whether there are factors that represent cognitive skills and noncognitive skills and how many factors retained.

Kaiser's eigenvalue rule: The Kaiser's criterion consists of retaining only factors with eigenvalues greater than 1 (Kaiser, 1960). The intuition behind this rule is that a factor should extract more variance than contained in a single variable, otherwise it should be dropped.

Scree plot: The scree plot was introduced by Cattell (1966). It is a visual tool used to help determine the number of important factors based on the analyst's inspection of a plot of the eigenvalues associated with the data. The number of factors should be equal to the number of eigenvalues before which the smooth decrease of eigenvalues levels off to the right of the plot.

Cognitive skills:

Table B.1: Factor Analysis/Correlation - Cognitive Skills (Principal Component Factors)

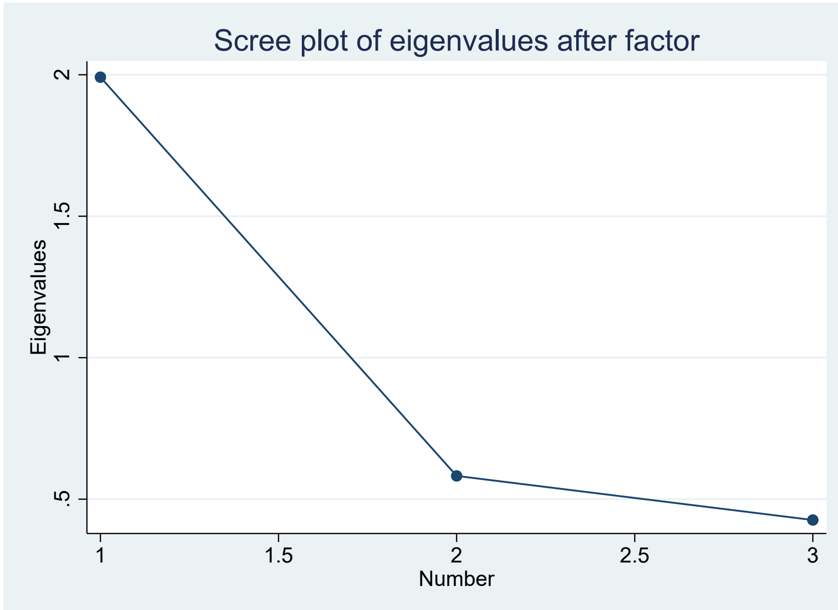
	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.99155	1.40944	0.66385	0.66385
Factor2	0.58211	0.15577	0.19404	0.85789
Factor3	0.42634	.	0.14211	1.00000

LR test: independent vs. saturated: chi2(3) = 504.68.
Prob>chi2 = 0.0000.
Retained factors = 1, 718 observations.

Table B.2: Factor Loadings (Pattern Matrix) and Unique Variances - Cognitive Skills

	Factor1	Uniqueness
PPVT test	.7873078	.3801465
Math test	.8538751	.2708974
Cloze test	.8016167	.3574107

Figure B.1: Scree Plot - Cognitive Skills



Noncognitive Skills:

Table B.3: Factor Analysis/Correlation - Noncognitive Skills (Principal Component Factors)

	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.88172	1.20759	0.62724	0.62724
Factor2	0.67413	0.22998	0.22471	0.85195
Factor3	0.44415	.	0.14805	1.00000

LR test: independent vs. saturated: chi2(3) = 433.27.

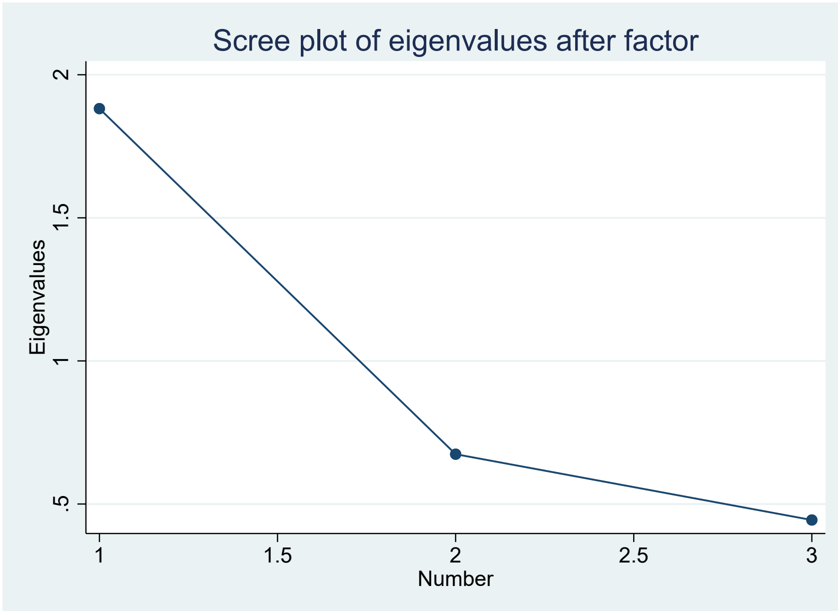
Prob>chi2 = 0.0000.

Retained factors = 1, 757 observations.

Table B.4: Factor Loadings (Pattern Matrix) and Unique Variances - Noncognitive Skills

	Factor1	Uniqueness
Self-esteem	.7866501	.3811816
Self-efficacy	.7349131	.4599027
Self Respect and Inclusion	.8501783	.2771969

Figure B.2: Scree Plot - Noncognitive Skills



Appendix C: Model Estimation Procedure

The full log-likelihood function I want to estimate is Equation 1.14:

$$\mathcal{L}(\Psi) = \sum_{i=1}^{N} ln\left(\int f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta\right) \tag{C.1}$$

Where Ψ are all the parameters of the model that I want to estimate, $\Psi = \{\alpha, \beta, \sigma, \tau_c, \mu_c, \Omega_c\}$.

I maximize the log-likelihood function C.1 using the minorization-maximization algorithm developed in James (2017) and Aucejo and James (2021).

Given an initial value of parameters, Ψ^0, the log-likelihood function $\mathcal{L}(\Psi)$ can be bounded below by a quadratic function:

$$Q(\Psi|\Psi^0) = \sum_{i=1}^{n} \int ln(f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta)) h(\theta | T_i, D_i, Y_i, \Psi^0) d\theta \tag{C.2}$$

where

$$h(\theta|T_i, D_i, Y_i, \Psi^0) = \frac{f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta)}{\displaystyle\int f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta') f(\theta') d\theta'} \tag{C.3}$$

Given the integral in the surrogate function $Q(\Psi|\Psi^0)$, it must be simulated by drawing R values of θ from $f(\theta|\Psi^0)$ and approximating $h(\theta|B_i, P_i, \Psi^0)$ by the weight:

$$w_{ir}^0 = \frac{f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0)}{\displaystyle\sum_{r=1}^{R} f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0)} \tag{C.4}$$

The lower bound function is now:

$$Q(\Psi|\Psi^0) = \sum_{i=1}^{n} \sum_{r=1}^{R} w_{ir}^0 ln(f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0) f(\theta_{ir}^0)) \tag{C.5}$$

Maximizing this function gives a new set of parameters, Ψ^1, that guarantee $\mathcal{L}(\Psi^1) >$

$\mathcal{L}(\Psi^0)$. Replacing Ψ^1 with Ψ^0 and iterating this process until the parameters converge. Let m denote the mth iteration of the algorithm. The parameter updates at the mth iteration are found by:

$$\hat{\tau}_{ir} = w_{ir}^m \frac{\tau_c^m normpdf(\theta_{ir}^m, \mu_c^m, \Omega_c^m)}{\sum_{c'=1}^{C} \tau_{c'}^m normpdf(\theta_{ir}^m, \mu_{c'}^m, \Omega_{c'}^m)} \tag{C.6}$$

$$\begin{aligned} \tau_c^{m+1} &= \frac{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}}{n} \\ \mu_c^{m+1} &= \frac{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}\theta_{ir}^m}{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}} \\ \Omega_c^{m+1} &= \frac{\sum_{i=1}^{n}\sum_{i=1}^{R} \hat{\tau}_{ir}(\theta_{ir}^m)(\theta_{ir}^m)'}{\sum_{i=1}^{n}\sum_{i=1}^{R} \hat{\tau}_{ir}} - (\mu_c^{m+1})(\mu_c^{m+1})' \end{aligned} \tag{C.7}$$

Since θ are treated as observed variables, the updated parameters $\{\alpha, \beta, \sigma\}$ can be estimated by standard OLS and logit models for the continuous and binary dependent variables respectively with the weights. In particular, for simplicity, let y_i be dependent variables including the observed measures, college choice and income and x_i be independent variables, including observed covariates and unobserved factors. Equation system 1.11, which I want to estimate, take the form $y_i = x_i{}'\beta + u_i$.

If y_i is continuous, then

$$\begin{aligned} &\beta^{m+1} = (XX)^{-1} * XY \\ &Where\ XX = \sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})(x_{ir})' \\ &and\ XY = \textstyle\sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})(y_i) \end{aligned} \tag{C.8}$$

If y_i is binary, then

$$\begin{aligned}
&\beta^{m+1} = \beta^m - B^{-1} * XY \\
&Where\ B = -\tfrac{1}{4}\sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^{m}(x_{ir})(x_{ir})' \\
&and\ XY = \sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^{m}(x_{ir})({y_i}' - p_{ir}^{m\prime}) \\
&with\ p_{ir}^{m} = \frac{(exp({x_{ir}}'\beta))^{D_i}}{1 + exp({x_{ir}}'\beta)}
\end{aligned} \tag{C.9}$$

Appendix D: Factor Distribution Moments

Table D.1: Factor Means, Standard Deviation and Correlation

	Cognitive skills	Noncognitive skills
Factore means	-0.107	-0.022
	(0.020)	(0.011)
Factor standard deviation	0.664	0.389
	(0.025)	(0.015)
Factor correlation:		
Cocnitiveskill	1	–
Noncognitiveskill	0.290	1
	(0.028)	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Table D.2: Mixture Component Means

	Cognitive skills	Noncognitive skills	Type share
Type 1	-0.351	-0.011	0.516
	(0.098)	(0.034)	(0.071)
Type 2	0.152	-0.034	0.484
	(0.041)	(0.034)	(0.071)

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Appendix E: An Alternative Specification for the Factors

An alternative setting to the factor loadings is triangular as follows:

$$T_{ij} = \alpha_j + \beta_j^C \theta_i^C + \beta_j^{NC} \theta_i^{NC} + u_{ij} \tag{E.1}$$

for j = {1, 2, 3}= {math, cloze, ppvt}.

$$T_{ik} = \alpha_k + \beta_k \theta_i^{NC} + u_{ik} \tag{E.2}$$

for k = {1, 2, 3}= {ses, sef, ser}.

Specifically, the measurement system takes the following form:

$$\begin{aligned}
T_{i,ppvt} &= \alpha_{ppvt} + \beta_{ppvt}^C * \theta_i^C + \beta_{ppvt}^{NC} \theta_i^{NC} + u_{i,ppvt} \\
T_{i,math} &= \alpha_{math} + \beta_{math}^C \theta_i^C + \beta_{math}^{NC} \theta_i^{NC} + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta_{cloze}^C \theta_i^C + \beta_{cloze}^{NC} \theta_i^{NC} + u_{i,cloze} \\
T_{i,ses} &= \alpha_{ses} + \beta_{ses}^{NC} * \theta_i^{NC} + u_{i,ses} \\
T_{i,sef} &= \alpha_{sef} + \beta_{sef}^{NC} \theta_i^{NC} + u_{i,sef} \\
T_{i,ser} &= \alpha_{ser} + \beta_{ser}^{NC} \theta_i^{NC} + u_{i,ser}
\end{aligned} \tag{E.3}$$

The factor loadings are as follows:

$$\left[\beta_T^C, \beta_T^{NC}\right] = \begin{bmatrix} \beta_{math}^C & \beta_{math}^{NC} \\ \beta_{cloze}^C & \beta_{cloze}^{NC} \\ \beta_{ppvt}^C & \beta_{ppvt}^{NC} \\ \beta_{ser}^C & \beta_{ser}^{NC} \\ \beta_{sef}^C & \beta_{sef}^{NC} \\ \beta_{ses}^C & \beta_{ses}^{NC} \end{bmatrix} = \begin{bmatrix} \beta_{math}^C & \beta_{math}^{NC} \\ \beta_{cloze}^C & \beta_{cloze}^{NC} \\ 1 & \beta_{ppvt}^{NC} \\ 0 & \beta_{ser}^{NC} \\ 0 & \beta_{sef}^{NC} \\ 0 & 1 \end{bmatrix} \tag{E.4}$$

Where both the cognitive and noncognitive factors load onto or affect the observed cognitive measures and only the noncognitive factor load onto the noncognitive measures.

All of the equations and the analysis on the schooling decision and earnings outcome are the same as in the main text.

Tables E.1 , E.2 and E.3 show the estimates of this alternative setting. The coefficients of controls, loadings and latent factors are not much different from the main specification.

Table E.1: Measurement System - Correlated Factors

	PPVT	Math	Cloze	Self-Esteem	Self-Efficacy	Self Respect and Inclusion
Panel A: Estimated parameters						
Constant	0	-0.038**	0.010	0	-0.014	0.008
		(0.018)	(0.022)		(0.011)	(0.014)
Cognitive	1	1.140***	1.131***	0	0	0
		(0.058)	(0.060)			
Noncognitive	0	-0.063	-0.021	1	0.707***	1.280***
		(0.069)	(0.076)		(0.040)	(0.074)
Panel B: Average Marginal Effects of Factors (AME)[a]						
Cognitive AME	0.661***	0.754***	0.747***	0	0	0
	(0.025)	(0.022)	(0.028)			
Noncognitive AME	0	-0.024	-0.008	0.385***	0.272***	0.493***
		(0.026)	(0.029)	(0.016)	(0.013)	(0.016)
Average value	-0.106	-0.157	-0.109	-0.023	-0.030	-0.021
Panel C: Variance Decomposition						
Signal	0.479***	0.608***	0.521***	0.363***	0.276***	0.765***
	(0.026)	(0.023)	(0.022)	(0.024)	(0.023)	(0.041)
Noise	0.521***	0.392***	0.479***	0.637***	0.724***	0.235***
	(0.026)	(0.023)	(0.022)	(0.024)	(0.023)	(0.041)
N	738	747	739	757	757	757

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

[a] Average Marginal Effects of Factors.

Table E.2: Probability of College Completion as a Function of Skills - Correlated Factors

Variables	Coefficients	Average Marginal Effects
Cognitive	1.949***	0.174***
	(0.221)	(0.017)
Noncognitive	0.473**	0.023**
	(0.196)	(0.010)
Female	0.658***	0.082***
	(0.119)	(0.015)
Urban	0.154	0.019
	(0.121)	(0.015)
Number of siblings aged 0-18	-0.748***	-0.084***
	(0.111)	(0.010)
Wealth index	0.080	0.010
	(0.608)	(0.074)
Parental educational level	0.463***	0.060***
	(0.061)	(0.008)
Child's educational aspiration	1.220***	0.145***
	(0.180)	(0.019)
Constant	-3.580***	-
	(0.418)	
Baseline probability	0.254	-
N	*757*	-

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table E.3: Effects of Skills on Earnings - Correlated Factors

Variables	Hourly earnings	
	D = 0	D = 1
Cognitive	3.299***	3.131
	(0.538)	(2.410)
Noncognitive	1.425	10.470***
	(1.022)	(2.982)
Cognitive AME	2.180***	2.069
	(0.339)	(1.590)
Noncognitive AME	0.549	4.030***
	(0.397)	(1.097)
Female	-3.818***	-2.554
	(0.524)	(1.570)
Urban	3.079***	7.578***
	(0.750)	(1.571)
Experience	2.300***	8.527***
	(0.441)	(2.411)
Experience squared	-0.308***	-1.713**
	(0.056)	(0.692)
Constant	14.695***	13.767***
	(0.663)	(1.673)
Average value	15.114	23.287

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

[a] Average Marginal Effects of Factors.

Chapter 2

Parental Investment and Child Development

2.1 Introduction

Developing human capital can offer a way for children to take advantage of new opportunities to improve their lives and contribute to sustainable economic growth and development. There is strong evidence showing that human capital is formed early in life and child development at early ages has long-lasting effects on adult social and economic outcomes (Knudsen, 2004; Knudsen et al., 2006; Cunha et al., 2006; Heckman et al., 2006; Urzúa, 2008; O'Neill, 1990). However, in developing countries, children face various risk factors and developmental deficits in every aspect of human capital development including cognitive, noncognitive skills and health that deter their development. Evidence shows that policies and interventions are effective in early childhood and for disadvantaged children (Knudsen et al., 2006; Doyle et al., 2009; Cunha et al., 2006; Cunha and Heckman, 2007; Heckman, 2008; Engle et al., 2007). There is also an increasing consensus that human capital is multidimensional with various components, including cognitive skills, noncognitive skills and health, and there are important dynamic complementarities and interactions among different components and factors. These dynamic complementarities and interactions, together with the fact that skills are malleable, give rise to potential early interventions and policies that can improve child development and thereby improve individual productivity. The effectiveness of such interventions and policies requires an understanding of the evolution of human capital throughout childhood: how its various components, including cognitive skills, noncognitive skills and health are formed and interacted, the importance of investments and the role of family background in driving child development and growth. However our understanding of these mechanisms, roles and interactions is relatively limited to date.

In this study, I use high-quality data from the Vietnam Young Lives survey to estimate a dynamic production function model for the various dimensions of human capital with endogenous parental investments to examine dynamic complementarities and interactions among different inputs and factors in forming child human capital. I examine the process by which current stocks of cognitive skills, noncognitive skills and health depend on past cognitive and noncognitive skills, past health, parental cognitive and noncognitive skills, and parental investments. I use a maximum likelihood approach to estimate the joint distribution of the latent factors and dynamic CES production functions of human capital.

This research contributes to the existing literature in several ways. First, as far as I know, this is the first attempt to analyze the determinants and interactions of three important dimensions of human capital: cognitive skills, noncognitive skills and health. This research examines the dynamic production of these three key components of human capital that are likely to be fundamental determinants of children's productivity and future development. Second, it provides rare evidence in developing country settings. It utilizes high-quality longitudinal data from the Young Lives survey to identify the determinants and interactions of cognitive skills, noncognitive skills and health over two critical childhood development stages, aged 12 and 15. Third, the research uses a latent factor approach to identify the latent, unobserved factors instead of (noisy) proxy variables to correct measurement error problems, capture multiple skill dimensions more accurately and explore the endogeneity of investments.

2.2 Literature Review

Although research on skill foundation has been growing recently with a number of significant contributions (Cunha and Heckman, 2008; Cunha et al., 2010), the literature on skill foundation is still scarce. First, this is because it requires special longitudinal surveys that follow children throughout different periods of their life. The second reason arises from difficulties in directly measuring skills. Skill measures in survey data indirectly reflect true or latent cognitive and noncognitive skills.

The recent literature shows several important features of human capital formation. First, evidence shows that the development of human capital is dynamic, and the components of human capital - cognitive, noncognitive skills and health - are malleable and influenced by many external factors (Cunha and Heckman, 2007; Cunha et al., 2006; Attanasio, 2015). They are dynamically self-productive, i.e., the current stock of one human capital component begets the future stock of this own component and cross-productive, i.e., the current stock of one human capital component augments the development of another human capital component in the future. Cunha and Heckman (2008) show evidence of the self-productivity of skills, i.e., skills accumulated in one period foster the development of skills in future periods. Cunha and Heckman (2008) also find evidence of the cross-productivity of skills. In particular, their estimates show strong cross-productivity effects of noncognitive skills on cognitive skills, but the reverse seems weak. A second important feature of human capital development is dynamic complementarity, i.e. the productivity of investments in subsequent periods depends on skills, health and investments in previous periods. There is some evidence showing that parental investments play a key role in children's skill development (Doyle et al., 2009; Cunha and Heckman, 2008; Cunha and Heckman, 2007; Coneus et al., 2012). Evidence about skills' self-productivity, cross-productivity and dynamic complementarity show the importance of parental investments in children's early life (Cunha and Heckman, 2007).

While a series of research for developing countries show mixed results of self-productivity and cross-productivity, they find strong evidence of dynamic complementarity of human capital (Helmers and Patnam, 2011; Attanasio et al., 2017; Attanasio et al., 2020; Sánchez, 2017). For example, using the Young Lives data from India, Helmers and Patnam (2011) show that cognitive skills strongly affect both cognitive (self-productivity) and noncognitive skills (cross-productivity), but they find no evidence of self-productivity for noncognitive skills and cross-productivity effects from noncognitive to cognitive skills. This result differs from the results from Sánchez (2017), which favor the cross-productivity of noncognitive skills on cognitive skills. Attanasio et al. (2020)'s research on Mexico provides strong evidence of self-productivity, and they also find evidence of the cross-productivity effect from cognitive skills on noncognitive skills, but not vice versa. Consistent with Cunha and Heckman (2008), research in developing countries shows the importance of parental investments and dynamic complementarity in developing cognitive and noncognitive skills. Studying human capital development at early ages in developing countries is particularly important to boost policies, interventions and investments in children to minimize the loss of human potential given the evidence that they are exposed to various risk factors and face developmental deficits in every aspect of human capital development (Engle et al., 2007).

2.3 Model

2.3.1 Dynamics of Skill Formation

Literature has shown that human capital constituents (cognitive skills, noncognitive skills and health) are dynamic and influenced by many external factors. My framework builds on the dynamic factor models of Cunha et al. (2010), Cunha and Heckman (2007), Attanasio et al. (2017), Agostinelli and Wiswall (2016) and Aucejo and James (2021) to incorporate a variety of factors in the process of human capital production. In this framework, past skills and health produce future period skills and health; and investments can promote skill development and health and vice versa, the past stocks of skills and health can affect the next period's stock of skills and health indirectly by inducing investments in them, and the stocks of parental skills can affect their child's development. The current stocks of cognitive skills, noncognitive skills and health are determined by the past cognitive skills, noncognitive skills, health, parental investments and the parental stocks of cognitive skills and noncognitive skills. In particular, the stocks of cognitive and noncognitive skills and health of child i at time t (denoted by $\Theta^C_{i,t}$, $\Theta^{NC}_{i,t}$ and $\Theta^H_{i,t}$ respectively) are a function of the child's stock of cognitive skills, noncognitive skills and health at time t-1 (($\Theta^C_{i,t-1}$, $\Theta^{NC}_{i,t-1}$ and $\Theta^H_{i,t-1}$), the parental stocks of cognitive skills (P^C_i) and noncognitive skills (P^{NC}_i) and the investments made by the parent I_i.

$$\Theta^k_{i,t} = f(\Theta^C_{i,t-1}, \Theta^{NC}_{i,t-1}, \Theta^H_{i,t-1}, I_{i,t}, P^C_i, P^{NC}_i, X_{i,t}, A^k_t, v_{i,t}, \varepsilon^k_{i,t}) \tag{2.1}$$

$$k \in \{C, NC, H\}$$

I use a Constant Elasticity of Substitution (CES) production function. The CES production function has been recently used and it has been the most flexible functional form used in the human capital production literature. The CES functional form allows for a great level of flexibility in exploring substitutability between various inputs in the production function.

$$\begin{aligned}\Theta^k_{i,t} = \quad & \left[\gamma^k_{1t}(\Theta^C_{i,t-1})^{\rho^{tk}} + \gamma^k_{2t}(\Theta^{NC}_{i,t-1})^{\rho^{tk}} + \gamma^k_{3t}(\Theta^H_{i,t-1})^{\rho^{tk}} + \gamma^k_{4t}(I_{i,t})^{\rho^{tk}}\right. \\ & \left.\gamma^k_{5t}(P^C_i)^{\rho^{tk}} + \gamma^k_{6t}(P^{NC}_i)^{\rho^{tk}}\right]^{1/\rho^{tk}} e^{X'_{i,t}\delta^k_t + A^k_t + \mu^k v_{i,t} + \varepsilon^k_{i,t}}\end{aligned} \tag{2.2}$$

$$k \in \{C, NC, H\}$$

Where C, NC and H stand for cognitive skills, noncognitive skills and health respectively, $\gamma^k_{1t} + \gamma^k_{2t} + \gamma^k_{3t} + \gamma^k_{4t} + \gamma^k_{5t} + \gamma^k_{6t} = 1$.

In addition to the five different inputs mentioned above, I also include other components that contribute to the accumulation of human capital. $X_{i,t}$ are observable variables which include child background characteristics (gender of the child, the number of siblings in the family, residential region). $\varepsilon^k_{i,t}$ are normally distributed unobserved shocks. The term A^k_t represents total factor productivity (TFP). The production functions include the residual of a investment function, $v_{i,t}$, as a control function to control for endogenous investments that will be discussed in Sections 2.3.3 and 2.4 about parental investments and estimation.

2.3.2 Measurement System

The Young Lives survey contains rich data with multiple variables for human capital production functions. As discussed in Chapter 1, it is not efficient and feasible to use all of the available measures as separate variables in the production function. Furthermore, using measures observed in the data as a proxy for skills, investments and health suffers from measurement errors since all these measures provide imperfect proxies of latent skills, investments and health. These measures should only be considered as noisy, error-ridden proxies for latent unobserved skills, investment and health factors.

I use a latent factor model to extract the latent factors of interest from a large set of measures observed in the data and remove the measurement errors. The basic idea behind the factor approach is that one can relate measures observed in the data to unobserved, latent, underlying factors.

In this model, I estimate latent factors measuring child cognitive, noncognitive skills and health at time t ($\Theta^C_{i,t}$, $\Theta^{NC}_{i,t}$, $\Theta^H_{i,t}$), child cognitive, noncognitive skills and health at time t - 1 ($\Theta^C_{i,t-1}$, $\Theta^{NC}_{i,t-1}$, $\Theta^H_{i,t-1}$), parental cognitive and noncognitive skills (P^C_i, P^{NC}_i) and parental investments ($I_{i,t}$). Since the latent factors ($\Theta^C_{i,t}$, $\Theta^{NC}_{i,t}$, $\Theta^H_{i,t}$, $\Theta^C_{i,t-1}$, $\Theta^{NC}_{i,t-1}$, $\Theta^H_{i,t-1}$, P^C_i, P^{NC}_i, and $I_{i,t}$) are not directly measured, I use the factor model approach to extract these unobserved variables from a large set of observed data.

Since I estimate the log of the production function in Equation 2.2 and as required by the model that the factors are positive, I define the natural log of the factors as $\theta^k_{i,t} = ln(\Theta^k_{i,t})$, $\theta^k_{i,t-1} = ln(\Theta^k_{i,t-1})$, $\mathcal{P}^C_i = ln(P^C_i)$, $\mathcal{P}^{NC}_i = ln(P^{NC}_i)$ and $\mathcal{I}_{i,t} = ln(I_{i,t})$ so that latent factors only take positive values. With this definition, I assume that the observed measures proxy the natural log of the factors (Cunha et al., 2010; Agostinelli and Wiswall, 2016; Attanasio et al., 2017; Aucejo and James, 2021).

Let $T^k_{i,j,\tau}$ be the jth measure relating to latent factor k for individual i at time τ ($\tau \in \{t-1, t\}$. There are two types of the observed measures, continuous and binary measures. The continuous measures are described by:

$$T^k_{i,j,\tau} = \alpha^k_{j,\tau} + \beta^k_{j,\tau}\theta^k_{i,\tau} + u^k_{i,j,\tau} \tag{2.3}$$

Where $\alpha^k_{j,\tau}$ are the intercepts, $\beta^k_{j,\tau}$ is the factor loadings on factor k for measure j at time τ. $u^k_{i,j,\tau}$ are measurement errors which are assumed to be normally distributed with mean zero and variance $\sigma^2_{u_{j,\tau}}$, independent of the latent factors and mutually independent. $u^k_{i,j,\tau}$ reflect that the observed measures are imperfect proxies of the latent factors.

The binary measures are described by:

$$T^k_{i,j,\tau} = \mathbb{1}[\alpha^k_{j,\tau} + \beta^k_{j,\tau}\theta^k_{i,\tau} + u^k_{i,j,\tau} > 0] \tag{2.4}$$

Where $\mathbb{1}$ is an indicator function that equals one if $T^{*k}_{i,j,\tau} = \alpha^k_{j,\tau} + \beta^k_{j,\tau}\theta^k_{i,\tau} + u^k_{i,j,\tau} > 0$. $u^k_{i,j,\tau}$ are assumed to be logistically distributed, independent of the latent factors and mutually independent.

This equation maps the jth measure to latent, unobserved factor k. The assumption is that the observed measures are imperfect, error-ridden proxies for the underlying factors.

Specifically, the system of equations for continuous measures can be written as follows:

For child's skills and health:

$$\begin{aligned} T^C_{i,j,\tau} &= \alpha^C_{j,\tau} + \beta^C_{j,\tau}\theta^C_{i,\tau} + u^C_{i,j,\tau} \\ T^{NC}_{i,j,\tau} &= \alpha^{NC}_{j,\tau} + \beta^{NC}_{j,\tau}\theta^{NC}_{i,\tau} + u^{NC}_{i,j,\tau} \\ T^H_{i,j,\tau} &= \alpha^H_{j,\tau} + \beta^H_{j,\tau}\theta^H_{i,\tau} + u^H_{i,j,\tau} \end{aligned} \tag{2.5}$$

Parental skills and parental investments follow the same structure.

For parental skills:

$$\begin{aligned} T_{i,j}^{PC} &= \alpha_j^{PC} + \beta_j^{PC}\mathcal{P}_i^{C} + u_{i,j}^{PC} \\ T_{i,j}^{PNC} &= \alpha_j^{PNC} + \beta_j^{PNC}\mathcal{P}_i^{NC} + u_{i,j}^{PNC} \end{aligned} \tag{2.6}$$

For parental investments:

$$T_{i,j,\tau}^{I} = \alpha_{j,\tau}^{I} + \beta_{j,\tau}^{I}\mathcal{I}_{i,\tau} + u_{i,j,\tau}^{I} \tag{2.7}$$

The system of equations for binary measures can be specified as follows:

$$\begin{aligned} T_{i,j,\tau}^{C} &= \mathbb{1}[\alpha_{j,\tau}^{C} + \beta_{j,\tau}^{C}\theta_{i,\tau}^{C} + u_{i,j,\tau}^{C} > 0] \\ T_{i,j,\tau}^{NC} &= \mathbb{1}[\alpha_{j,\tau}^{NC} + \beta_{j,\tau}^{NC}\theta_{i,\tau}^{NC} + u_{i,j,\tau}^{NC} > 0] \\ T_{i,j,\tau}^{H} &= \mathbb{1}[\alpha_{j,\tau}^{H} + \beta_{j,\tau}^{H}\theta_{i,\tau}^{H} + u_{i,j,\tau}^{H} > 0] \\ T_{i,j}^{PC} &= \mathbb{1}[\alpha_j^{PC} + \beta_j^{PC}\mathcal{P}_i^{C} + u_{i,j}^{PC} > 0] \\ T_{i,j}^{PNC} &= \mathbb{1}[\alpha_j^{PNC} + \beta_j^{PNC}\mathcal{P}_i^{NC} + u_{i,j}^{PNC} > 0] \\ T_{i,j,\tau}^{I} &= \mathbb{1}[\alpha_{j,\tau}^{I} + \beta_{j,\tau}^{I}\mathcal{I}_{i,\tau} + u_{i,j,\tau}^{I} > 0] \end{aligned} \tag{2.8}$$

Since the underlying/latent factors are unobserved, for identification, we need to normalize one of the loadings for each factor to one and one of the intercepts for each factor to zero. Since a variety of measures that may change from age to age are used, each factor is normalized on the same measure at every age to make the comparisons over time consistent. Child cognitive skills are always normalized on the Peabody Picture Vocabulary Test (PPVT), child noncognitive skills are normalized on self-esteem and child health is normalized on height. Parental cognitive skills are normalized on the mother's years of education, Parental noncognitive skills are normalized on parental self-esteem and parental investments are normalized on expenditure on the Young Lives child. Another condition to identify factors is that the number of observable measures $L >= 2k + 1$, where L is the number of measures and k is the number of factors (Cunha et al., 2010). This condition is satisfied since there are 27 measures for 9 factors in the model.

The distributions of the log factors $(f(\theta), f(\mathcal{P}), f(\mathcal{I}))$ are assumed to be jointly distributed as a mixture of two normals. The assumption of the mixture of normal distributions of the factors is essential in this model. First, Fewer restrictions are imposed on the distributions and distributions are flexible enough to capture data. Second, the production function functions are non-linear, so the distributions need to be general and flexible enough to be consistent with the model (Attanasio et al., 2017).

2.3.3 Parental Investments

Parental investments reflect parents' choices and they depend on parents' objectives, resources and how effective the investments in their children are. Investments are endogenously determined by parental resources, expectations regarding the returns to investments in their children, and the parent's levels of cognitive and noncognitive skills. Parents make investment choices taking into account the child's stocks of cognitive and noncognitive skills and health since returns to investments may depend on their child's stocks of human capital, in particular, if the child's stocks of human capital and investments are complementary. Investments may depend on the parent's levels of cognitive and noncognitive skills because parents with higher levels of human capital may be better aware of the value of investments and may have higher lifetime resources.

Parental investments are an input in the production function and reflect parents' choices considering the evolution of the child's human capital. Parents react to their child's human capital when they choose their investments in their children. Therefore, parental investments could be endogenous. Parental investments could be correlated with unobserved shocks or omitted inputs that are relevant for child human capital accumulation.

To deal with the endogenous nature of parental investments, I use the household wealth index, economic shocks and regional prices as instruments and use a control function approach inspired by Attanasio et al. (2017). These instruments are valid provided that wealth index, economic shocks and regional prices affect cognitive and noncognitive skills and health only through their impacts on parental investments.

In particular, the parental investment function is specified as follows:

$$\begin{aligned} lnI_{i,t} = \quad & \alpha_{1,t} + \alpha_{2,t} ln\Theta^{C}_{i,t-1} + \alpha_{3,t} ln\Theta^{NC}_{i,t-1} + \alpha_{4,t} ln\Theta^{H}_{i,t-1} + \alpha_{5,t} lnP^{C}_{i} + \alpha_{6,t} lnP^{NC}_{i} \\ & + \alpha_{7,t} X_{i,t} + \alpha_{8,t} Z_{i,t} + v_{i,t} \end{aligned} \tag{2.9}$$

Where $X_{i,t}$ includes child gender, urban/rural residence and the number of siblings. $Z_{i,t}$ a vector of the instrumental variables that determine the parental investment choices and are not included in the production function. $Z_{i,t}$ are the log of wealth index reflecting parental resources, household economic shocks (Shock in input prices[1], drought, flood, crop failure and illness of household members) and the log of regional prices. $v_{i,t}$ is an error term.

Because data on family income are not available in the survey, the wealth index is used to proxy for parental resources. Parental resources and prices included in the model reflect budget constraints. This model can be considered as an approximation to a dynamic model of household choice and parental investments with liquidity constraints in which parents make investment choices to maximize a welfare function with arguments of human capital and consumption, subject to a budget constraint and the production functions (Del Boca et al., 2013; Attanasio et al., 2017).

2.4 Estimation

The model estimation consists of two steps. In the first step, I estimate the measurement system to recover the parameters $\beta^{C}_{j,\tau}$, $\beta^{NC}_{j,\tau}$, $\beta^{H}_{j,\tau}$, β^{PC}_{j}, β^{PNC}_{j}, $\beta^{I}_{j,t}$, α^{C}_{j}, α^{NC}_{k}, α^{PC}_{j}, α^{PNC}_{j}, $\alpha^{I}_{j,t}$ and the latent factor distributions by the estimation approach described in Chapter 1.

[1] Shock in input prices refers to a large increase in the prices of inputs such as fertilizers, plant seeds or machinery and equipment for agricultural production.

In the second step, I use the estimated parameters of the factor distributions from the first step to take individual-specific draws and use these draws as observable data to estimate investment and production functions.

The parental investment function takes the form of Equation 2.9:

$$\begin{aligned} lnI_{i,t} = \quad & \alpha_{1,t} + \alpha_{2,t} ln\Theta^{C}_{i,t-1} + \alpha_{3,t} ln\Theta^{NC}_{i,t-1} + \alpha_{4,t} ln\Theta^{H}_{i,t-1} + \alpha_{5,t} lnP^{C}_{i} + \alpha_{6,t} lnP^{NC}_{i} \\ & + \alpha_{7,t} X_{i,t} + \alpha_{8,t} Z_{i,t} + v_{i,t} \end{aligned} \tag{2.10}$$

The $v_{i,t}$ is the residual of the investment function as a control function. In this specification, household wealth index, economic shocks and regional prices are included in the investment function but not in the production functions as follows:

$$ln\Theta^{k}_{i,t} = ln(g(\Theta^{C}_{i,t-1}, \Theta^{NC}_{i,t-1}, \Theta^{H}_{i,t-1}, I_{i,t}, P^{C}_{i}, P^{NC}_{i})) + X'_{i,t}\delta^{k}_{t} + A^{k}_{t} + \mu^{k} v_{i,t} + \varepsilon^{k}_{i,t} \tag{2.11}$$

Where $g(.)$ is the CES production function indicated earlier.

2.5 Data and Variables

I use the data for the Older Cohort from the Young Lives survey in Vietnam that follows 1,000 children from the age of 8 to age 22. It provides a rich data set on individual, family and community characteristics, health and cognitive and noncognitive skills. This research uses data from Round 2 (at age 12) and Round 3 (at age 15). This is because I want to investigate the impact of investments on child skills and health during adolescence and these two rounds contain cognitive, noncognitive skill and health measures needed for the research.

The household survey contains information on an extensive set of socio-economic and demographic characteristics, alongside a wealth of information around parenting, parental characteristics, and maternal skills, including mothers' years of education, verbal ability, IQ, depressive symptoms, and knowledge of child development. Table 2.1 presents descriptive statistics on the general characteristics of the sample. The sample includes 961 children.

There is a balance between boys and girls. Around 80% of the child live in rural areas.[2] On average, the number of children in the household is around 1.

Table 2.1: Key Descriptive Statistics

	Age 12, Round 2	Age 15, Round 3
Female	0.504	0.504
	(0.500)	(0.500)
Urban	0.198	0.199
	(0.398)	(0.399)
Number of siblings aged 0-18	1.347	0.965
	(0.982)	(0.947)
Wealth index	0.537	0.624
	(0.178)	(0.181)
Observations	961	961

Note: Standard deviations in parentheses.

Table 2.2 shows how the jth observed measure is mapped to the kth latent/underlying factor. The first loading of each factor is normalised to unity and thus the scale of the latent factors is defined by these measures. The next subsections 2.5.1, 2.5.2 and 2.5.3 describe the variables and corresponding latent factors used in the measurement system.

2.5.1 Children's Measures: Cognitive Skills, Noncognitive Skills and Health

Children's cognitive skill indicators for Round 3 are measured by the test scores in Peabody Picture Vocabulary Test (PPVT), mathematics test (math test), and reading comprehension test (Cloze test), and children's cognitive skill measures for Round 2 are measured by the PPVT score, math test score, and children's reading and writing levels.

Math test: The math test was administered in Rounds 2 and 3. It includes 29 items

[2] Young Lives in Vietnam uses a pro-poor sampling strategy but represents the diversity of children in the country (Nguyen, 2008). This sampling design allows for studying the human capital development of young people in relatively low-resource settings.

Table 2.2: Observed Variables in the Young Lives Surveys and Corresponding Latent Factors

Latent factors		**Observed variables**
Child's cognitive skills - Round 2	θ_2^C	1. PPVT test 2. Math Test 3. Reading level 4. Writing level
Child's cognitive skills - Round 3	θ_3^C	1. PPVT test 2. Math Test 3. Cloze
Child's noncognitive skills - Round 2 and Round 3	θ_2^{NC}, θ_3^{NC}	1. Self-esteem score 2. Self-efficacy score 3. Self-respect and inclusion score
Child's health - Round 2 and Round 3	θ_2^H, θ_3^H	1. Child height for age z-score 2. Child weight 3. How is child health?
Parental cognitive skills	$\mathcal{P}^C$	1. Mother's years of education 2. Father's years of education
Parental noncognitive skills	$\mathcal{P}^{NC}$	1. Self-esteem score 2. Self-efficacy score 3. self-respect and inclusion score
Parental Investments	$\mathcal{I}_3$	1. Expenditure on the Young Lives child 2. Number of hours studying outside school as a proxy for the time that parents dedicate to the child 3. Quality of relationship between child and parents

on addition, subtraction, multiplication, division, problem-solving, measurement, data interpretation, and basic geometry.

PPVT: The PPVT is a widely-used test of receptive vocabulary. It uses a stimulus word and a set of accompanying pictures to test receptive vocabulary. It has been used extensively to demonstrate the correlation between PPVT scores and cognitive and intellectual ability (Walker et al., 2005). The 204−item PPVT−III was used in Vietnam. Young Lives researchers in each country followed a standard process for adaptation and standardization of the PPVT.

Cloze: The Cloze test was developed to measure verbal skills and reading comprehension. The test includes 24 items that increase in difficulty. Each item consists of a sentence or short paragraph that lack one or more words, children were asked to identify a word that completed the meaning of the sentence or paragraph. A thorough analysis of psychometric characteristics was examined to establish the reliability and validity of all these tests (Crookston et al., 2014).

I use three indicators designed to access dimensions of self-esteem, self-efficacy and self-respect and inclusion to measure children's noncognitive skills.[3]

Self-esteem: Self-esteem measures aspects related to pride and it builds on the Rosenberg scale (Rosenberg, 1965).

Self-efficacy: The self-efficacy scale measures aspects related to agency and it builds on the Rotter scale (Rotter, 1966).

Self-respect and inclusion: focuses on the social component of self-esteem (Dercon and Krishnan, 2009). The statements used to measure self-respect revolve around the concepts of pride and the sense of inclusion.

Self-esteem and self-efficacy have been extensively studied and widely used. They have been validated and proved reliable in psychological and economic literature. Self-esteem and self-efficacy are the most common noncognitive skill variables used in empirical studies

[3] The items/statements used to assess these scales are described in Appendix A.

(Glewwe et al., 2017). Self-respect and inclusion are related to the self-esteem measure but focus on the social and psychosocial aspects of inclusion. The single measures of self-esteem, self-efficacy and self-respect and inclusion are set on a Likert scale ranging from "strongly disagree" to "strongly agree". Children were read statements and asked whether they strongly disagreed, disagreed, more or less, agreed or strongly agreed with the statements. Negative statements are recoded to reflect positive statements.

Child health measures used in this study include height for age z-score, weight and self-rated health status. The rationale for using these measures is because height for age z-score may reflect the information of longer-term health and nutrition status and it is calculated according to the World Health Organization (WHO) standards, while weight likely captures short-term health status. Height for age, weight and self-rated health are used to capture nutrition and health status in several studies using the Young Lives survey data (Helmers and Patnam, 2011; Attanasio et al., 2017; Sánchez, 2017).

2.5.2 Parental Cognitive Skills and Noncognitive Skills

In the models, parental cognitive skills and noncognitive skills are used to control for parental background. I use maternal education and paternal education to measure parental cognitive skills. Parental noncognitive skills are measured by three indicators designed to access dimensions of self-esteem, self-efficacy and self-respect and inclusion.[4] Parental cognitive and noncognitive skills are measured and treated as fixed in Round 2. The single measures of self-esteem, self-efficacy and self-respect and inclusion are set on a four-point Likert scale ranging from "strongly disagree" to "strongly agree". Parents were read statements and asked whether they strongly disagreed, disagreed, agreed or strongly agreed with the statements. Negative statements are recoded to reflect positive statements.

[4] The construction of these indicators is described in Appendix A.

Table 2.3: Key Descriptive Statistics - Child Measures

	Age 12, Round 2	Age 15, Round 3
Cognitive skill measures:		
PPVT test	137.900	166.471
	(25.336)	(27.702)
Math test	7.465	17.810
	(1.857)	(7.607)
Cloze test	-	17.934
		(4.979)
Reading level	0.972	-
	(0.166)	
Writing level	0.942	-
	(0.233)	
Noncognitive skill measures:		
Self-Esteem, raw score	3.446	3.850
	(0.443)	(0.569)
Self-Efficacy, raw score	3.394	4.168
	(0.367)	(0.504)
Self-respect and inclusion, raw score	3.538	3.778
	(0.397)	(0.421)
Health measures:		
Height-for-age z-score	-1.457	-1.420
	(1.081)	(0.909)
Child's weight (kg)	32.990	44.268
	(6.570)	(7.127)
How is child health?	0.715	0.503
	(0.452)	(0.500)
Observations	961	961

Note: Standard deviations in parentheses.

Table 2.4: Key Descriptive Statistics - Parental Skill Measures

Cognitive skill measures:	
Mother's years of education	6.014
	(3.707)
Father's years of education	6.829
	(3.680)
Noncognitive skill measures:	
Self-Esteem, raw score	3.536
	(0.445)
Self-Efficacy, raw score	3.587
	(0.548)
Self-respect and inclusion, raw score	3.653
	(0.445)
Observations	961

Note: Standard deviations in parentheses.

2.5.3 Parental Investments

I use three variables that measure parental resources devoted to the child in terms of money, time and the quality of the relationship between the parent and the child to extract the latent factor of parental investments. The first variable measures material investments that include expenditure on education, clothing, shoes, and books specifically devoted to the child. The second variable is the average number of hours per day the child studied outside school as a proxy for the time that parents dedicated to the child. The last variable measures the quality of the relationship between the child and the parent.[5]

[5] This scale is constructed using the items listed in Appendix A.

Table 2.5: Descriptive Statistics: Parental Investments

	Round 2, age 12	Round 3, age 15
Expenditure on the Young Lives child	851.094	2559.822
	(985.913)	(3005.632)
Study hours outside school	2.890	3.066
	(1.597)	(2.117)
Quality of relationship, raw score	3.367	2.663
	(0.405)	(0.420)
Observations	961	961

Note: Standard deviations in parentheses.

2.6 Results

2.6.1 Measurement System

The measurement system relies on observed measures to identify the latent factors and it is used to assess how factors load on each of the measures.

To assess the information content contained in each measure from factors and measurement errors, I calculate the contribution of latent factors and measurement errors in explaining the variance of the observed measures.

$$P_j^{\theta_\tau^k} = \frac{(\beta_{j,t}^k)^2 var(\theta_{i,\tau}^k)}{(\beta_{j,\tau}^k)^2 var(\theta_{i,\tau}^k) + var(u_{i,j,\tau}^k)} \tag{2.12}$$

$$P_j^{u_{j,\tau}^k} = \frac{var(u_{i,j,\tau}^k)}{(\beta_{j,\tau}^k)^2 var(\theta_{i,\tau}^k) + var(u_{i,j,\tau}^k)} \tag{2.13}$$

Where $P_j^{\theta_\tau^k}$ is the proportion of the variance of the jth observed measure explained by latent factor k at time τ or signal and $P_j^{u_{j,\tau}^k}$ is the variance of the measure explained by the measurement error or noise.

Table 2.6 shows the measures assigned to each factor and the estimates of factor loadings onto the log of the factors. It also reports the fraction of the variance of each measure explained by each factor (signal) and by the measurement error (noise). I find that, for the most part, the measures proposed in Section 5 contain a substantial amount of information for each factor. The cognitive skill factors in both rounds account for an important fraction of the variance of each observed measure - from 45.5% to 69.8%. The measures on noncognitive skills are also very informative, from 15.9% to 68.7% of the variance of the noncognitive measures are accounted for by signal. Similarly, the latent factors of parental cognitive and noncognitive skills and parental investments explain an important proportion of the variance of the related observed measures, from 16.1% to 99.5% with the exceptions of parental self-efficacy and the parental quality of relationship with the child which are close to zero. The factor for health explains an considerable share of the variance of the health indicators with the signals exceeding 50%. Although the factors explain an essential proportion of the variance of the observed measures, these proportions are far from 100%. This shows that the observed measures capture the latent/true factors with significant measurement errors and demonstrates the importance of the latent factors in assessing human capital accumulation.

Table 2.6: Measurement System[6]

Latent factors and measures	Data type	Loading	AME	Signal	Noise
Child's cognitive skills - Round 3, age 15					
1. PPVT test	Continuous	1	0.765	0.572	0.428
2. Math test	Continuous	1.071	0.819	0.649	0.351
3. Cloze	Continuous	0.899	0.687	0.455	0.545
Child's cognitive skills - Round 2, age 12					
1. PPVT test	Continuous	1	0.834	0.698	0.302
2. Math Test	Continuous	1.020	0.851	0.603	0.397
3. Reading	Binary	1.687	0.018	–	–
4. Writing	Binary	1.530	0.035	–	–
Child's noncognitive skills - Round 3, age 15					
1. Self-esteem score	Continuous	1	0.381	0.358	0.642
2. Self-efficacy score	Continuous	0.759	0.289	0.308	0.692
3. Self-respect and inclusion score	Continuous	1.200	0.457	0.687	0.313
Child's noncognitive skills - Round 2, age 12					
1. Self-esteem score	Continuous	1	0.435	0.561	0.439
2. Self-efficacy score	Continuous	0.539	0.234	0.159	0.841
3. Self-respect and inclusion score	Continuous	0.773	0.336	0.327	0.673
Child's health - Round 3, age 15					
1. Height for age z-score	Continuous	1	0.623	0.391	0.609

Continued on next page

Table 2.6 – *Continued from previous page*

Lalent factors and Measures	Data type	Loading	AME	Signal	Noise
2. Weight	Continuous	1.513	0.942	0.891	0.109
3. How is child health?	Binary	0.636	0.094	–	–
Child health - Round 2, age 12					
1. Height for age z-score	Continuous	1	0.758	0.577	0.423
2. Weight	Continuous	1.258	0.954	0.925	0.075
3. How is child health?	Binary	0.886	0.111	–	–
Parental cognitive skills					
Mother's years of education	Continuous	1	0.818	0.674	0.326
Father's years of education	Continuous	0.986	0.806	0.658	0.342
Parental noncognitive skills					
1. Self-esteem score	Continuous	1	0.471	0.522	0.478
2. Self-efficacy score	Continuous	0.141	0.066	0.022	0.978
3. Self-respect and inclusion score	Continuous	0.757	0.356	0.266	0.734
Parental Investments					
1. Expenditure on the Young Lives child	Continuous	1	1.001	0.995	0.005
2. Number of hours studying outside school	Continuous	0.399	0.400	0.161	0.839
3. Quality of relationship	Continuous	0.067	0.067	0.009	0.991

[6] The estimated factor distribution moments, including factor means, standard deviations, correlations and mixture components are shown in Appendix B.

2.6.2 Determinants of Parental Investments in Children

Parental investments reflect parental choices and decisions. The parental investment choices depend on their belief in child development, resources and preferences. The investment equation is a function of the child's cognitive and concognitive skills and health in the previous period, parental cognitive and noncognitive abilities and household characteristics.

Table 2.7 reports the estimates of the parental investment equation. Parental perception reflected by their cognitive and noncognitive abilities and their resources reflected by the wealth index have large effects on parental investment choices in their children. A 10% increase of a standard deviation in parental cognitive ability would increase parental investments in the child by 4.02% of a standard deviation, while a 10% increase of a standard deviation in parental noncognitive skills would increase investments by 1.72% of a standard deviation. This shows that better parental cognitive and noncognitive skills lead to better intergenerational skill outcomes through the investment channel. The wealth index has a positive, large and significant effect on investments; increasing the wealth index by a 10% would increase parental investments in children by 1.77 % of a standard deviation. Child health has a positive and significant effect on parental investments; a 10% increase of a standard deviation in child health would increase investments by 2.18% of a standard deviation.

Surprisingly, the child's cognitive and noncognitive skills do not have any impact on parental investments. It reflects that parents may not be aware of their child's skills and have too low expectations of returns to investments in skills. Attanasio et al. (2019) show that parents have distorted views about the child development process.

Parents in urban areas invest more in their children than parents in rural areas. Investments in female children are higher than investments in male counterparts. The number of siblings does not affect investments significantly.

The results also show that an increase in prices, except the price of notebook would increase investments, while shocks have negative effects on investments, except shock because of storms. In the sence of investments, the coefficients on prices should be negative. However, the goods considered are necessities in a relatively low-resource setting. Furthermore, I exploit the special (regional) variation in prices at the community level which is assumed to be not driven by demand differences and in a complex model with some alternative investment goods being complements to the goods used to proxy investment in the measurement system. Therefore, the estimated coefficients could be positive.

Table 2.7: Estimates of Parental Investment Function

Variables	Parental investments
Child's cognitive skills, age 12	-0.024
	(0.030)
Child's noncognitive skills, age 12	-0.102
	(0.076)
Child's health, age 12	0.218***
	(0.032)
Parental cognitive skills	0.402***
	(0.057)
Parental noncognitive skills	0.172***
	(0.066)
Wealth index	0.177***
	(0.045)
Female	0.080***
	(0.029)
Urban	0.530***
	(0.049)
Number of siblings aged 0-18	-0.032***
	(0.012)
Shock in input prices	-0.075**
	(0.034)
Drought	0.058**
	(0.029)
Flood	-0.060**
	(0.025)
Crops failure	-0.079***
	(0.022)
Illness of child's father	-0.006
	(0.047)
Illness of child's mother	-0.084**
	(0.041)
Illness of other household members	-0.114***
	(0.041)
Storm	0.118**
	(0.050)

Continued on the next page

Table 2.7 – *Continued from the previous page*

Variables	Parental investments
Price of notebook	0.038
	(0.024)
Price of clothes	0.250***
	(0.049)
Price of food	0.673***
	(0.136)
Price of medicine	0.069***
	(0.024)
Constant	-4.373***
	(0.654)
Observations	*961*

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

2.6.3 Production Functions

Table 2.8 presents the production function estimates for cognitive skills, noncognitive skills and health. To characterize more cohesively the size and significance of the overall effects of each input in the production functions and assess the sensitivity of our major inputs, I analyze their marginal effects and recreate the dynamic process by using the estimated parameters.[7] First, I calculate the marginal effect to access the overall effects of each main input (Table 2.9). Second, I explore the role of self-productivity and cross-productivity in producing skills and health. Third, I consider the dynamic complementarity between skills, health and parental investments. My results show several essential features of human capital accumulation.

[7] The marginal effects are derived in Appendix C.

Table 2.8: Estimates of Production Functions

	Cognitive skills at age 15	Noncognitive skills at age 15	Health at age 15
	(1)	(2)	(3)
Child's cognitive skills at age 12	0.613***	0.001	0.064**
	(0.044)	(0.039)	(0.029)
Child's noncognitive skills at age 12	0.145***	0.339***	0.130**
	(0.055)	(0.077)	(0.053)
Child's health at age 12	0.042*	0.065	0.670***
	(0.025)	(0.053)	(0.073)
Parental Investments	0.261***	0.505***	0.346**
	(0.075)	(0.167)	(0.170)
Parental cognitive skills	0.017	-0.005	-0.184
	(0.041)	(0.122)	(0.193)
Parental noncognitive skills	-0.079	0.096	-0.026
	(0.050)	(0.081)	(0.065)
A_t	-0.010	0.211***	0.205***
	(0.024)	(0.031)	(0.024)
Control Function	-0.126*	-0.132	-0.420**
	(0.068)	(0.262)	(0.179)
Female	0.062***	-0.036	-0.238***
	(0.016)	(0.026)	(0.027)
Urban	-0.021	-0.328*	-0.421***
	(0.059)	(0.194)	(0.133)
Number of siblings aged 0-18	0.010	0.038***	0.007
	(0.009)	(0.011)	(0.009)
Complementarity(ρ)	-0.168	-1.851**	0.034
	(0.117)	(0.757)	(0.327)
Elasticity of substitution	0.856***	0.351*	1.036***
	(0.141)	(0.211)	(0.195)
Observations		*961*	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9: Marginal Effects

	Cognitive skills at age 15	Noncognitive skills at age 15	Health at age 15
	(1)	(2)	(3)
Child's cognitive skills at age 12	0.614***	0.002	0.064**
	(0.043)	(0.039)	(0.029)
Child's noncognitive skills at age 12	0.145***	0.316***	0.130**
	(0.055)	(0.076)	(0.052)
Child's health at age 12	0.042*	0.085	0.670***
	(0.025)	(0.062)	(0.065)
Parental investments	0.261***	0.513***	0.346**
	(0.075)	(0.162)	(0.169)
Parental cognitive skills	0.017	-0.008	-0.184
	(0.042)	(0.123)	(0.194)
Parental noncognitive skills	-0.079	0.093	-0.026
	(0.050)	(0.078)	(0.067)
Observations		*961*	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

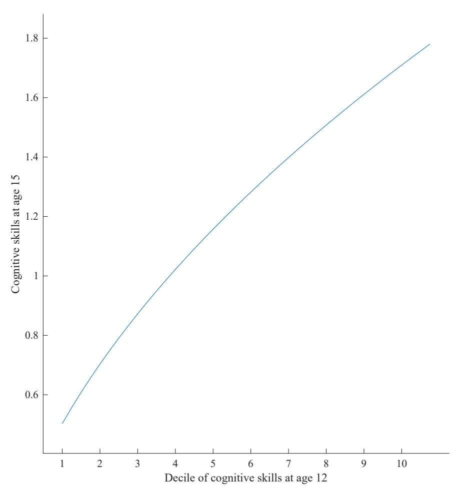
2.6.3.1 Cognitive Skills

Table 2.8, Column 1 presents the estimates of the production function for cognitive skills and Table 2.9, Column 1 shows the marginal effects of main inputs. The coefficient on the investment residuals (control function) is significant and negative. It implies that investments are endogenous in the production of cognitive skills. The negative sign of this coefficient suggests that parents tend to increase their investments to compensate for an adverse shock that is unobserved but perceived by the parents and causes a decline in the child's cognitive skills. Ignoring this effect could lead to an underestimate of the impact of investments. Appendix D, Table D.1 and D.2 show that when investments are taken as exogenous, the coefficient on the investments is much lower, while the coefficients on other inputs are not dramatically affected. This result shows a compensatory role of parents to shocks to the child. The results show several other features of cognitive skill accumulation.

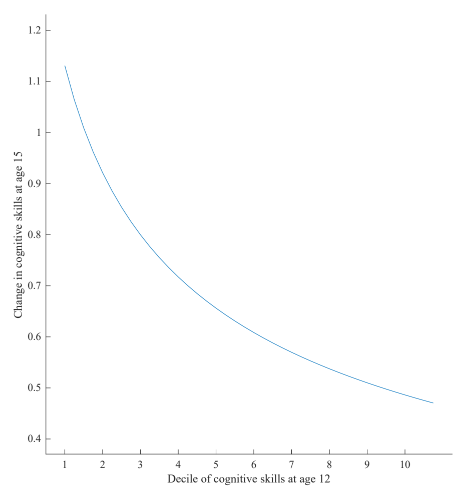
First, cognitive skills show a very strong self-productivity effect. That is, past cognitive skills have a strong and positive effect on current cognitive skills. Increasing cognitive skills at age 12 by 10% of a standard deviation would increase cognitive skills at age 15 by 6.13% of a standard deviation. Figure 2.1a illustrates the level of cognitive skills in the current period for each decile of past levels of cognitive skills, keeping all other inputs at their mean values. It shows that high cognitive skills produce high future cognitive skills. Figure 2.1b displays the self-productivity of cognition ($\partial\Theta_t^C/\partial\Theta_{t-1}^C$) for each decile of the levels of cognitive skills in the last period. These figures show that marginal increments of past cognitive skills are very productive ($\partial\Theta_t^C/\partial\Theta_{t-1}^C$ is high and positive for the entire space). It also demonstrates that the productivity of cognitive skills is higher for lower cognitive skill levels.

Figure 2.1: Cognitive Skills: Self-productivity

(a) Θ_t^C as a Function of Θ_{t-1}^C



(b) Cognitive Skills: Self-Productivity ($\partial\Theta_t^C/\partial\Theta_{t-1}^C$)



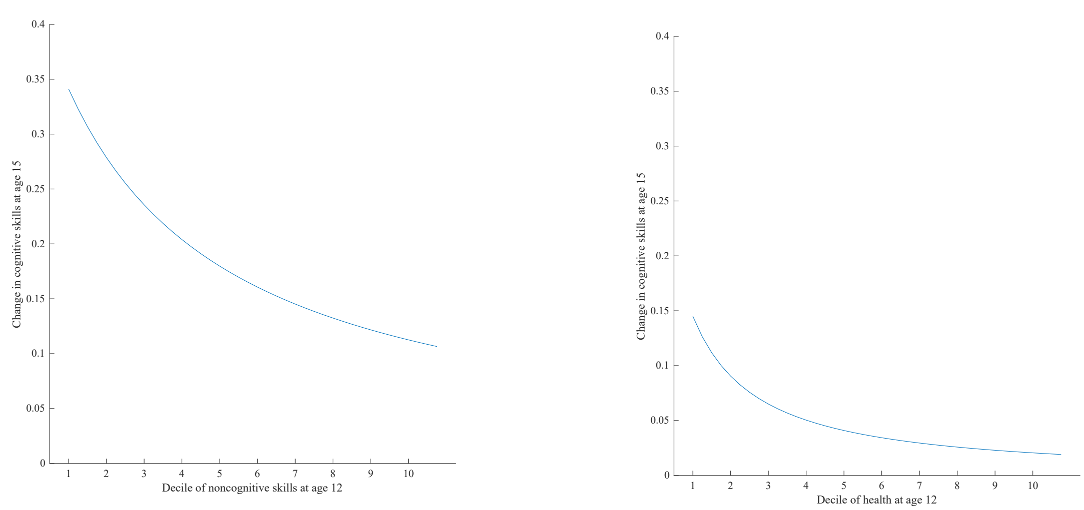
Second, there is cross-productivity from noncognitive skills and health to cognitive skills. That is, noncognitive skills and health in the last period foster the development of current cognitive skills. The cross-effect of noncognitive skills on cognition is strong and the cross-effect of health on cognition is small compared to that of noncognitive skills. A 10% increase of a standard deviation in the stocks of noncognitive skills and health in the last period would increase the current level of cognitive skills by 1.45% and 0.42% of a standard deviation respectively. Figure 2.2 shows the extent to which noncognitive skills and health affect the development of cognitive skills. $\Theta_t^C/\partial\Theta_{t-1}^{NC}$ and $\Theta_t^C/\partial\Theta_{t-1}^{H}$ are positive for the entire space and the cross-effect of noncognitive skills is stronger than that of health. The largest impacts are for children with the lowest level of noncognitive skills and health. These results demonstrate the importance of noncognitive skills and good health in developing cognitive skills.

These results are consistent with previous studies (Cunha et al., 2010; Helmers and Patnam, 2011; Attanasio et al., 2017; Sánchez, 2017). Cunha et al. (2010) and Helmers and Patnam (2011) find that self-productivity of cognition and cross-productivity from noncognitive skills to cognitive skills play an important role in the formation of skills with larger level effects of self-productivity of cognition. The result of cross-productivity effects of health on cognition is aligned with that of Attanasio et al. (2017) and Sánchez (2017) which indicates that health is important for future cognitive skill development.

Figure 2.2: Cognitive skills: Cross-productivity

(a) Cross-productivity from Noncognitive Skills to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^{NC}$)

(b) Cross-productivity from Health to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^{H}$)



Figures 2.3, 2.4 and 2.5 show that the self-effect of cognitive skills and cross-effects of noncognitive skills and health in producing future cognitive skills are higher for those who have higher levels of cognitive skills, noncognitive skills and health. That is, at the same

initial stock of cognition, the self-effect is higher for those with higher initial noncognitive skills and better health. Similarly, with the same initial level of noncognitive skills, the cross-effect from noncognitive skills to cognitive skills is higher for those with higher initial cognition and better health, and the cross-effect from health to cognitive skills is higher for those with higher initial stocks of cognitive and noncognitive skills. These results mean that higher stocks of a certain dimension of human capital make the marginal increments of other dimensions of human capital more productive.

Figure 2.3: Cognitive Skills: Self-productivity ($\partial\Theta_t^C/\partial\Theta_{t-1}^C$)

(a) (b)

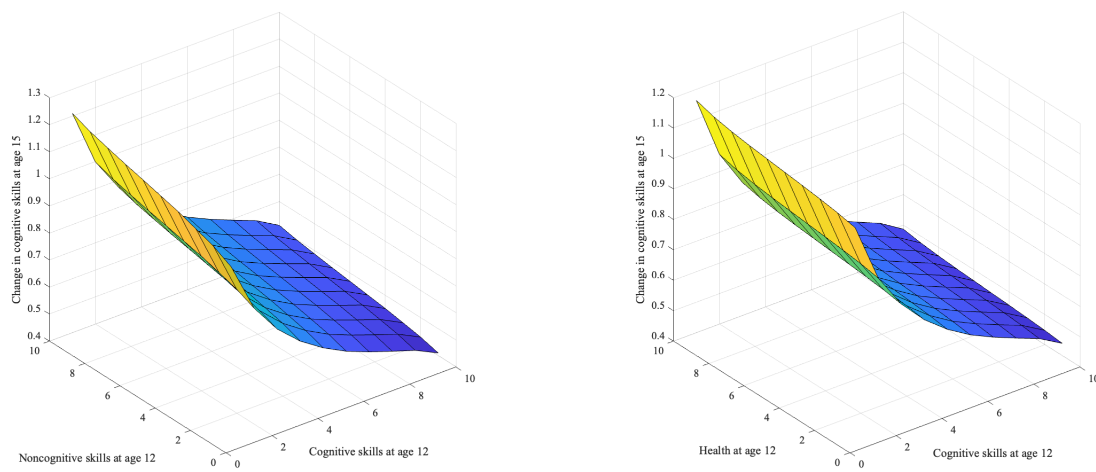


Figure 2.4: Cognitive Skills: Cross-productivity from Noncognitive Skills to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^{NC}$)

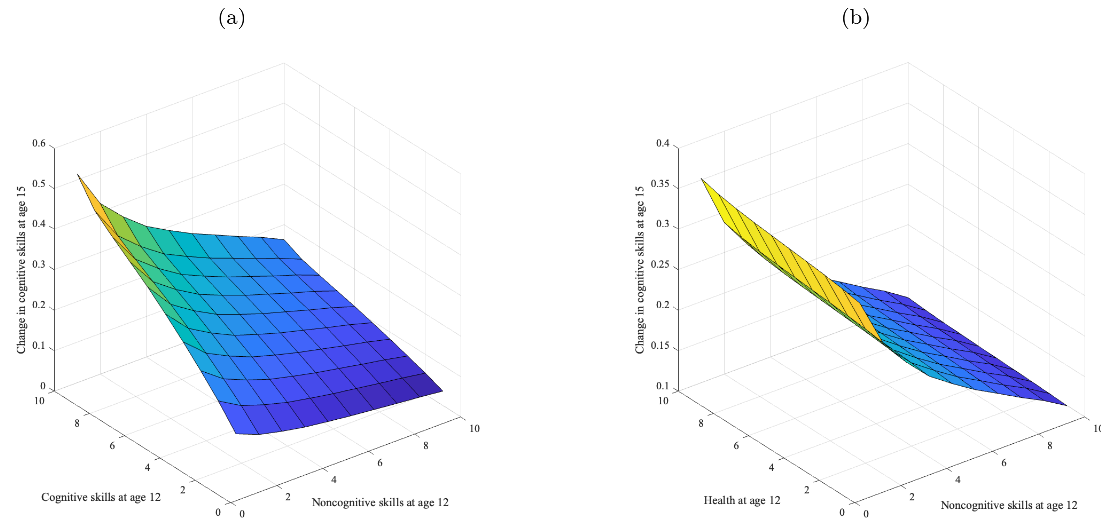
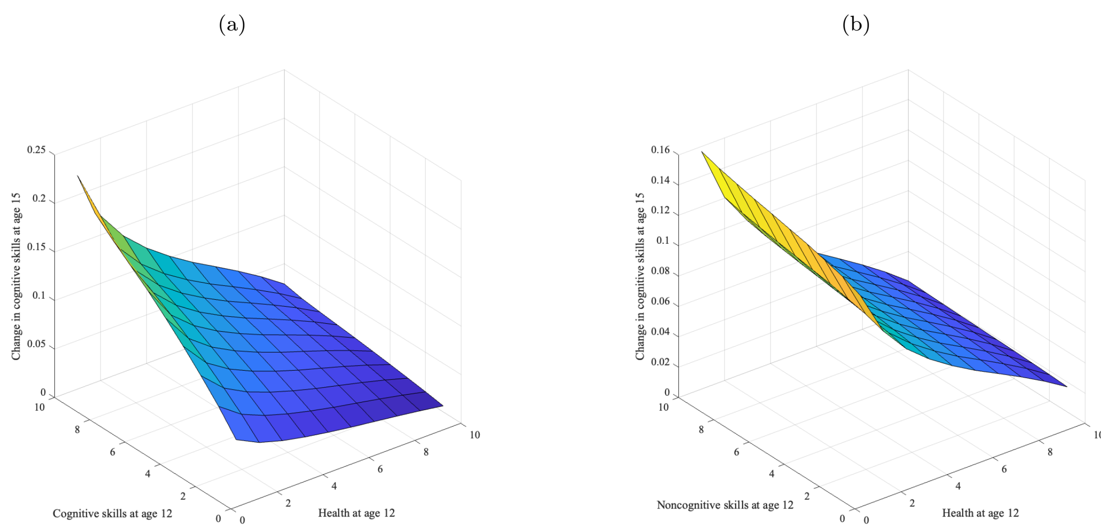


Figure 2.5: Cognitive Skills: Cross-productivity from Health to Cognitive Skills ($\partial\Theta_t^C/\partial\Theta_{t-1}^{H}$)



Third, parental cognition and con-cognition do not have significant impacts on child cognitive skills. It indicates that parental background of cognitive and noncognitive skills only plays an important role in their investment decisions in children and indirectly develops their child's cognitive skills through investments. Girls are more likely to have higher cognitive skills than boys. The residence of the child and the number of siblings do not seem to have a significant effect on cognitive skill accumulation, although they matter for parental investment decisions.

Fourth, one of the key estimates is the complementarity coefficient ($\rho = -0.168$). The elasticities of substitution between the various inputs ($\sigma = 1/(1\text{-}\rho) = 0.856$) show some degree of complementarities. This result is aligned with the existing literature (Attanasio et al., 2020; Cunha et al., 2010). I also reject the hypothesis that $\rho = 1$, which indicates that the production function is linear and the inputs are additively separable. These results imply that it is not easy to compensate and remediate low levels or deficits of skills and health from the previous periods.

Finally, the key result, which largely motivates this study, is the role of parental investments. First, parental investments have a very strong effect on the child's cognitive skills. A 10% increase of a standard deviation in parental investments would increase cognitive skills by 2.61% of a standard deviation respectively. Second, to understand the extent to which parental investment can affect the accumulation of cognitive skills of the child, I explore the dynamic complementarity between cognitive skills and investments, a concept introduced in Cunha and Heckman (2007) to imply that past cognitive skills and past investments in those skills increase the productivity of current investments ($\partial^2\Theta_t^C/\partial\Theta_t^I\partial\Theta_{t-1}^C > 0$). Figure 2.6 shows a strong dynamic complementarity between cognitive skills and investments, meaning that returns to investments are higher for children with better initial cognitive skills or higher past cognitive skills make investments more productive. Figure 2.7 shows compounded complementarity effects under the effects of noncognitive skills and health, higher noncognitive skills and health make investments even more productive. These results is in line with the existing literature (Attanasio et al., 2020; Attanasio et al., 2017; Cunha and Heckman, 2007; Cunha et al., 2010).

Figure 2.6: Complementarity between Investments and Cognitive Skills

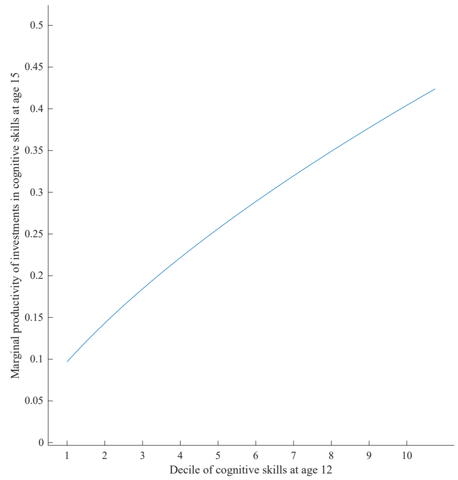
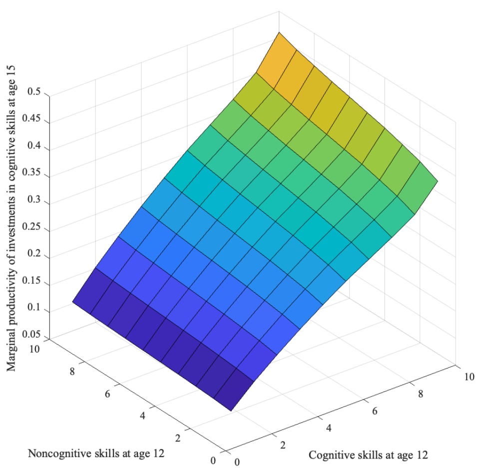
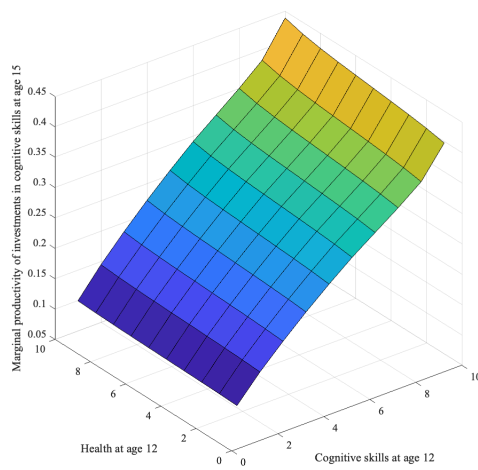


Figure 2.7: Complementarity between Investments and Cognitive Skills

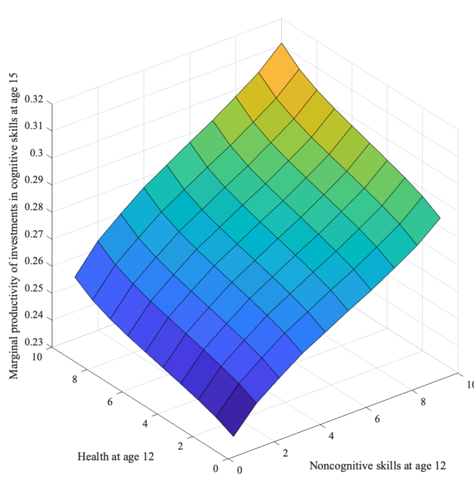
(a)



(b)



(c)



The self-effects, cross-effects and dynamic complementarity together become a dynamic multiplier effect mechanism of cognitive skill accumulation whereby skills, health and investments produce cognitive skills. These multiplier effects would lead to different rates of cognitive skill growth for children with different initial skills and health. Furthermore, parental skills and resources are positively associated with parental investments, which implies that children with better family backgrounds get more investments and they use these investments more effectively. These effects could lead to substantial increases in inequality in producing skills and finally lead to social inequality. Dynamic complementarity is crucial since it could be a source of inequality and it shows the role of investments. These results also indicate the importance of interventions by boosting investments at early ages that can alter child development path, especially for disadvantaged children. A lack of parental investments can seriously hinder the development of a child.

2.6.3.2 Noncognitive Skills

The analysis of noncognitive skills follows the same methodology as the one for cognitive skills, presented in the previous section. I find no evidence of endogenous investments in noncognitive skill production. The results are shown in Table 2.8, Column 2 and Table 2.9, Column 2 confirming the evidence of self-productivity of noncognitive skills in which noncognitive skills in an earlier period produce better noncognitive skills in the next period. A 10% increase of a standard deviation in noncognitive skills at age 12 would increase noncognitive skills at age 15 by 3.16% of a standard deviation. Figure 2.8 shows that noncognitive skills are productive in inducing better noncognitive outcomes ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{NC} > 0$) and the self-effects of noncognitive skills are higher for children with higher initial levels of cognition and health (Figure 2.9).

Figure 2.8: Noncognitive skills: Self-productivity

(a) Θ_t^{NC} as a Function of Θ_{t-1}^{NC}

(b) Noncognitive Skills: Self-productivity $(\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{NC})$

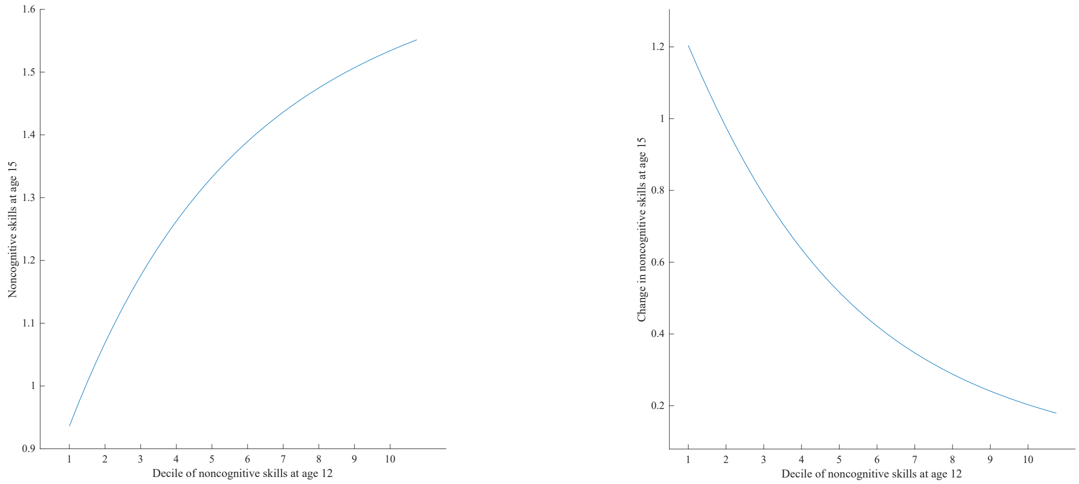
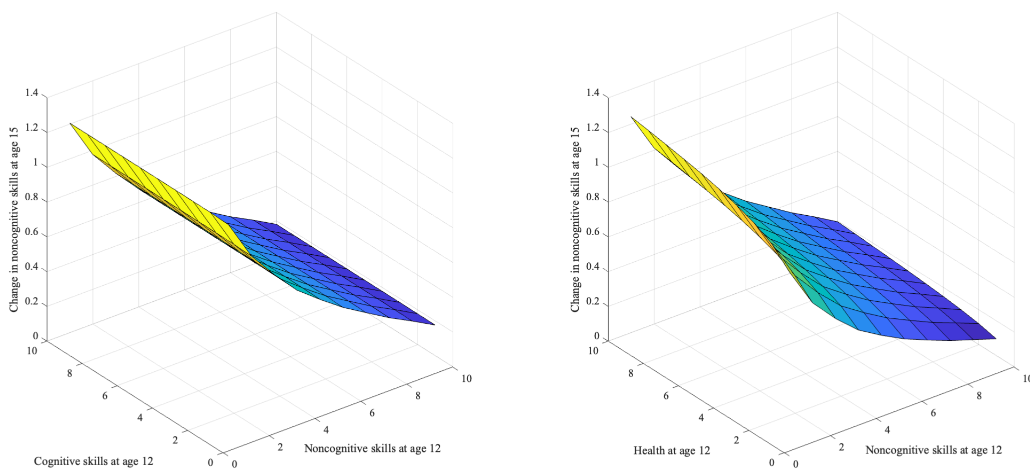


Figure 2.9: Noncognitive Skills: Self-productivity $(\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{NC})$

(a)

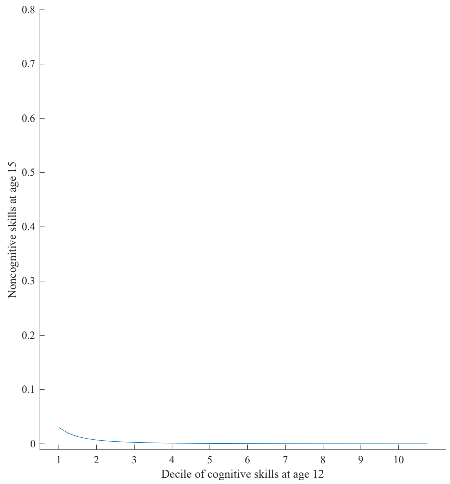
(b)



The results do not show the existence of cross-productivity effects of cognitive skills and health on noncognitive skills. The initial levels of cognitive skills and health do not contribute to producing noncognitive skills. The cross-effects are similar among children with different initial levels of noncognitive skills except for the lowest deciles of cognition and health (Figures 2.10, 2.11 and 2.12).

Figure 2.10: Noncognitive Skills: Cross-productivity from Cognitive Skills and Health

(a) Cross-productivity from Cognitive Skills to Noncognitive skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{C}$)



(b) Cross-productivity from Health to Noncognitive Skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{H}$)

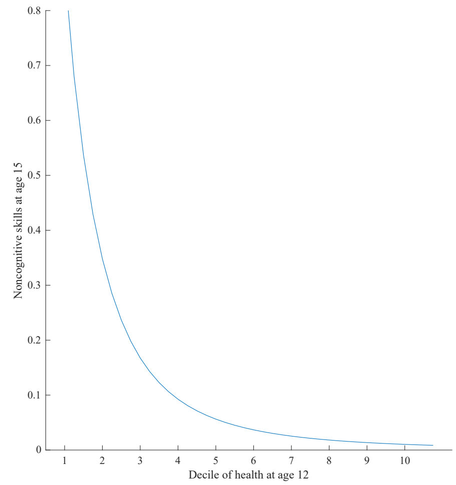
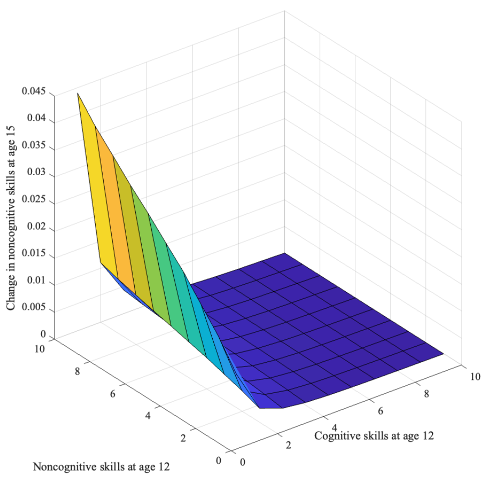


Figure 2.11: Noncognitive Skills: Cross-productivity from Cognitive Skills to Noncognitive Skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{C}$)

(a)



(b)

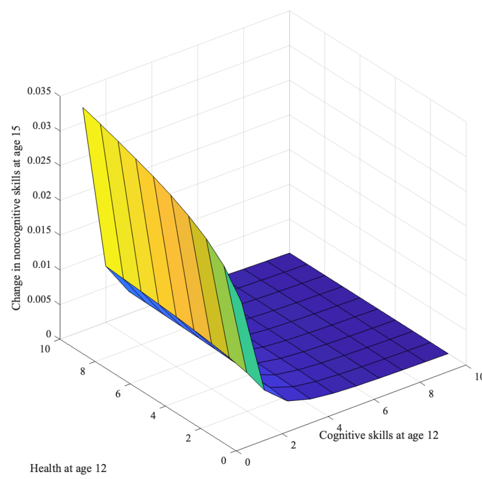
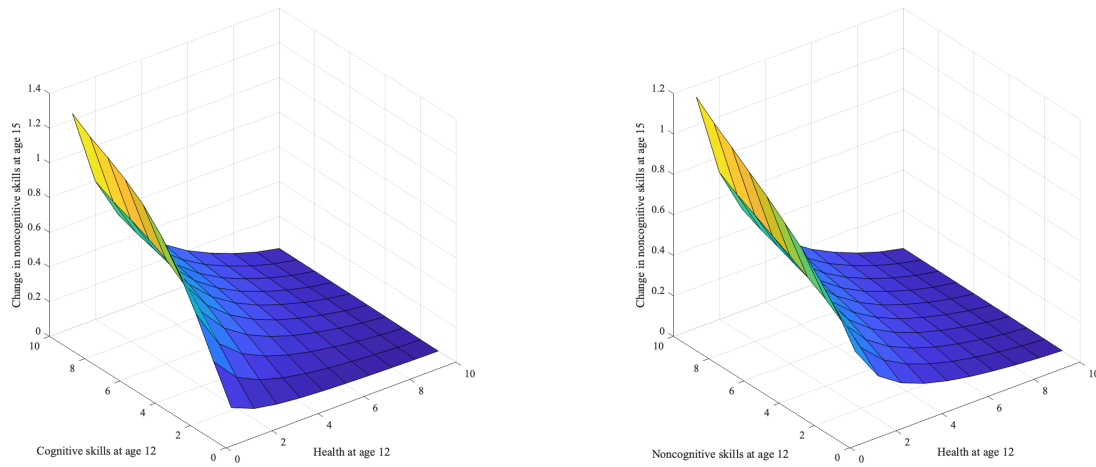


Figure 2.12: Noncognitive Skills: Cross-productivity from Health to Noncognitive Skills ($\partial\Theta_t^{NC}/\partial\Theta_{t-1}^{H}$)

(a) (b)

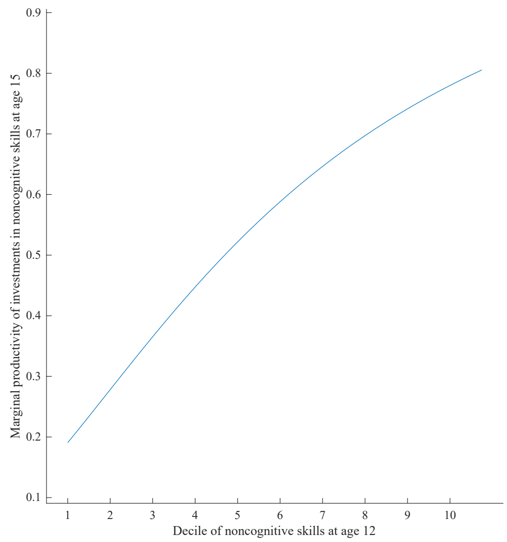


There is no evidence that parental cognitive and noncognitive skills have a direct effect on the child's noncognitive skills.

The complementarity coefficient (ρ = -1.851) and the elasticities of substitution (σ = 1/(1-ρ) = 0.351) show is relatively small compared to those of cognitive skills. This suggests the complementarity among inputs is stronger for the noncognitive production function and it is more difficult to remedy deficits of noncognitive skills by investments.

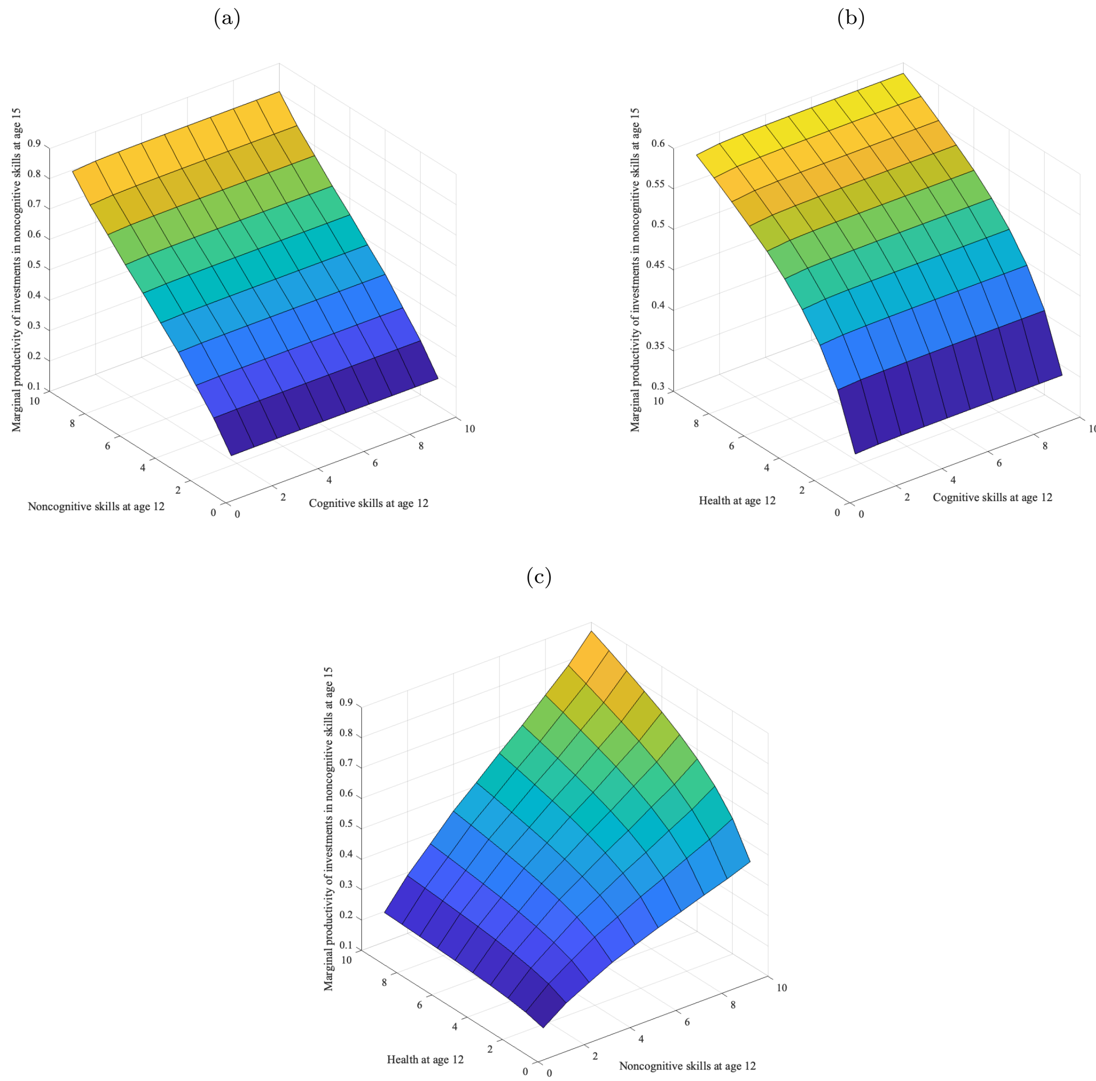
Parental investments have a very strong effect on a child's noncognitive skills. A 10% increase of a standard deviation in parental investments would increase noncognitive skills by 5.13% of a standard deviation. The effect of investments on noncognitive skills is larger than that on cognitive skills and health, which will be discussed in Section 2.6.3.3. Figures 2.13 and 2.14 show a strong dynamic complementarity between noncognitive skills and investments and the marginal effects of investments are higher among children with higher initial noncognitive skills. It is noted that child cognitive skills and health in the previous period do not have significant impacts on the current stock of noncognitive skills as discussed above.

Figure 2.13: Complementarity between Investments and Noncognitive Skills



My results share certain commonalities with the study of Cunha et al. (2010) for the US in the sense that noncognitive skills are self-productive, cognition is not cross-productive for noncognitive skill formation and investments are an important factor for noncognitive skill accumulation. However, the result about cross-productivity effects contrasts with that of Attanasio et al. (2020) and Sánchez (2017), which find that current cognitive skills foster future noncognitive skill accumulation. My result about self-productivity is consistent with that of Attanasio et al. (2020), Attanasio et al. (2017), Cunha et al. (2010)) and Sánchez (2017).

Figure 2.14: Complementarity between Investments and Noncognitive Skills



2.6.3.3 Child Health

Now turn to the production of health. Table 2.8, Column 3 and Table 2.9, Column 3 present the estimates of the production process of health and the marginal effects of main inputs. As with the cognitive skill production function, the coefficient on the investment residuals is significant and negative. This implies that investments are endogenous in the production function of health and the negative sign suggests that parents tend to compensate for adverse shocks to their children by increasing investments.

The estimated results show strong evidence of self-productivity and the existence of cross-productivity from cognitive and noncognitive skills to health in the production process of health. A 10% increase of a standard deviation in health in the previous stage would increase health in the current stage by 6.7% of a standard deviation. Increasing child cognitive and noncognitive skills in the previous period by 10% of a standard deviation would increase health in the current period by 0.64% and 1.3% of a standard deviation respectively.

Figures 2.15 and 2.16 show the self-productivity of health and Figures 2.17, 2.18 and 2.19 show cross-productivity from cognitive skills and noncognitive skills to health. Figures 2.18 and 2.19 show the cross-effects of cognitive and noncognitive skills on health are higher for those with high initial levels of health and this productivity does not change much among deciles of cognitive and noncognitive skills.

Figure 2.15: Health: Self-productivity

(a) Θ_t^H as a Function of Θ_{t-1}^H (b) Health: Self-Productivity ($\partial\Theta_t^H/\partial\Theta_{t-1}^H$)

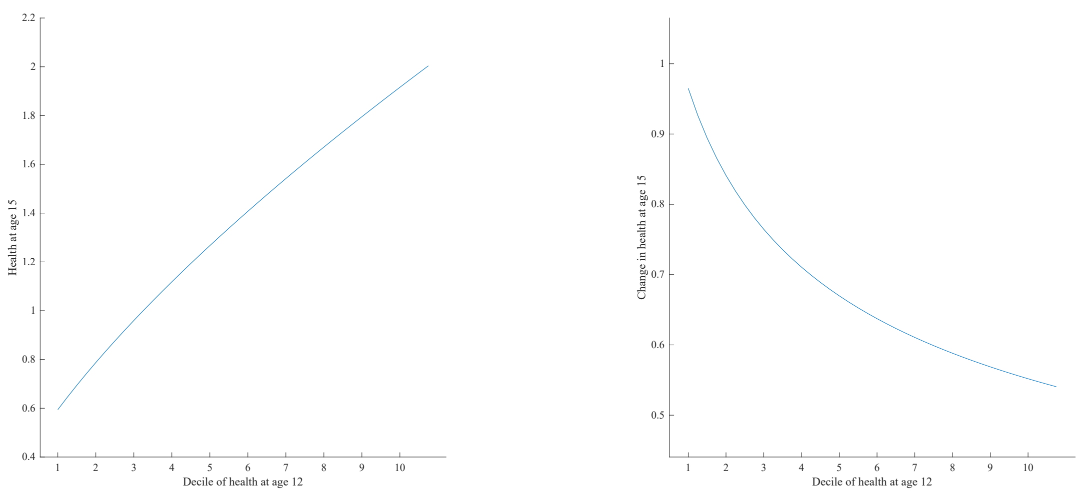
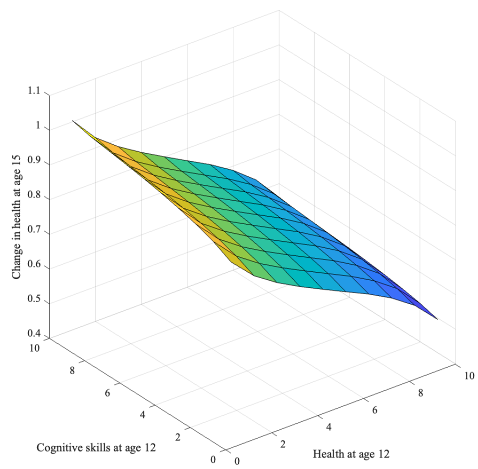


Figure 2.16: Health: Self-productivity ($\partial\Theta_t^H/\partial\Theta_{t-1}^H$)

(a)



(b)

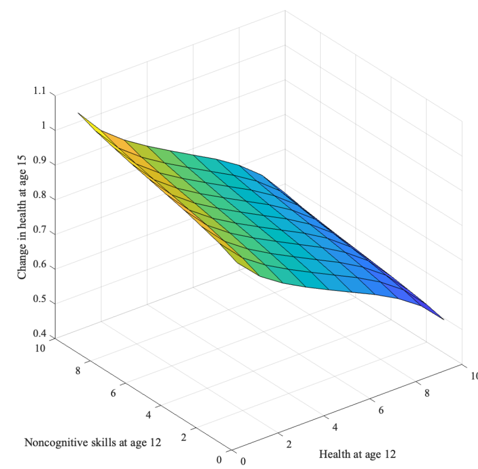
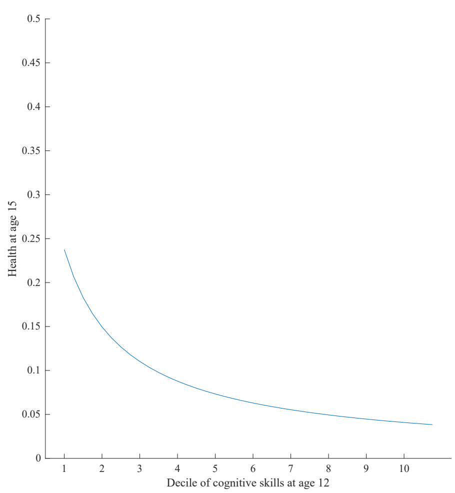


Figure 2.17: Health: Cross-productivity

(a) Cross-productivity from Cognitive Skills to Health($\partial\Theta_t^H/\partial\Theta_{t-1}^C$)



(b) Cross-productivity from noncognitive skills to health ($\partial\Theta_t^H/\partial\Theta_{t-1}^{NC}$)

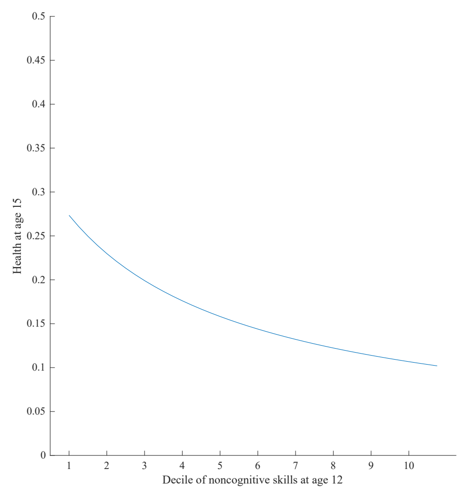


Figure 2.18: Health: Cross-productivity from Cognitive Skills to Health ($\partial\Theta_t^H/\partial\Theta_{t-1}^C$)

(a) (b)

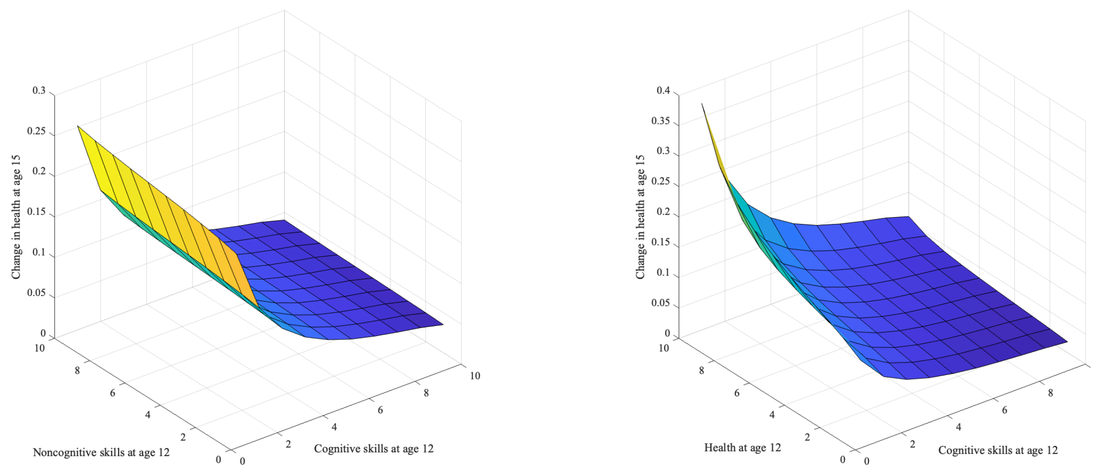
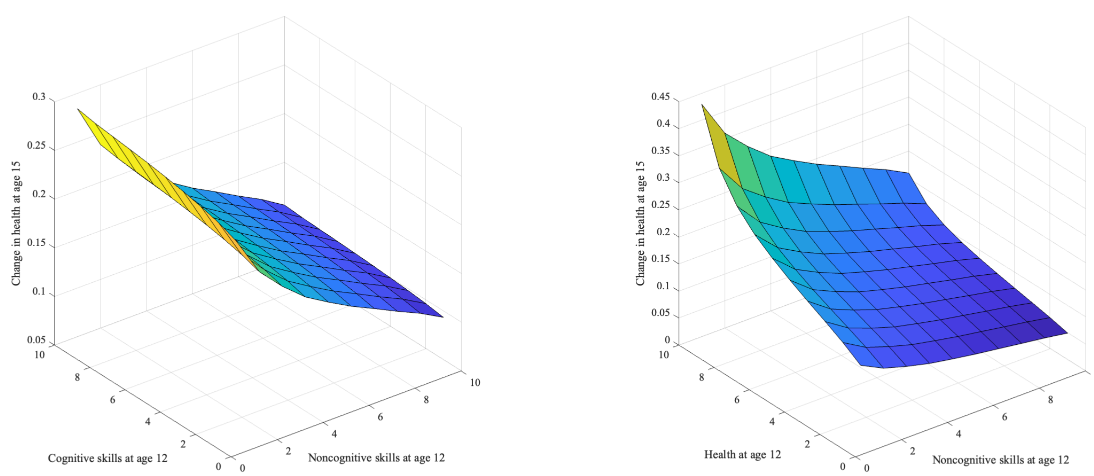


Figure 2.19: Health: Cross-productivity from Noncognitive Skills to Health ($\partial\Theta_t^H/\partial\Theta_{t-1}^{NC}$)

(a) (b)



The results also show a strong impact of investments on health and strong evidence of a dynamic complementarity between parental investments and their child's health. Table 2.9, Column 3 shows that A 10% increase of a standard deviation in parental investments would increase health by 3.46% of a standard deviation. . Figures 2.20 and 2.21 show strong complementarity between investments and child health. With the important roles of health on child development as discussed here and given that parental investments strongly affect child health, interventions that address health deficits as early as possible are critical for child development, especially in low-resource settings where health deficits at early ages are common.

While Attanasio et al. (2017) reported mixed results about self-productivity effects of cognitive skills on health, my result is aligned with them in terms of self-productivity effects and large impacts of investments.

Figure 2.20: Complementarity between Investments and Child Health

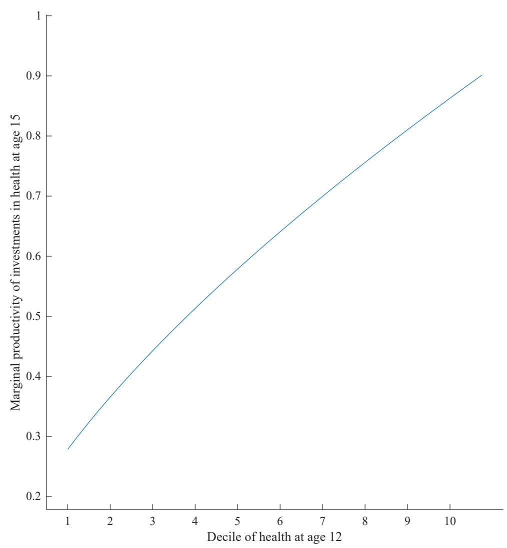
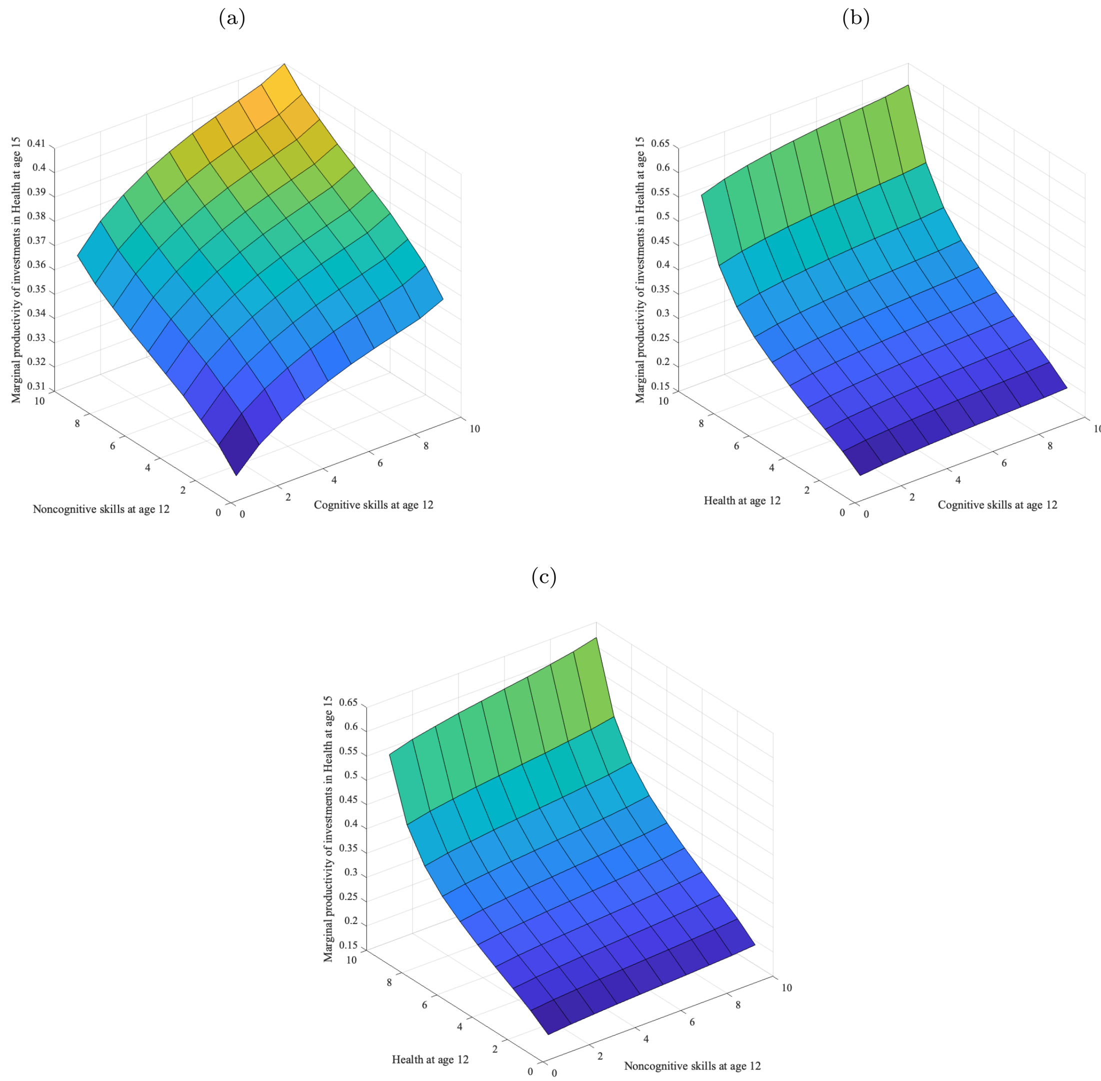


Figure 2.21: Complementarity between Investments and Child Health



Parental cognitive and noncognitive skills have no impact on health. Being a boy, living in rural area and having fewer siblings have positive impacts on health.

2.7 Conclusion

Understanding the evolution of human capital, how its various constituents (cognitive skills, noncognitive skills and health) interact in childhood and its intergenerational transmission is vital to design policies and interventions that can improve productivity and contribute to sustainable economic growth and development. This is especially true in the context of developing countries where children face various developmental risk factors and health deficits that deter their development. With rich and unique data from the Young lives survey that contains multiple dimensions and indicators of cognitive skills, noncognitive skills and health, this study uses the dynamic factor analysis approach to recover latent factors that can more precisely capture the multidimensional nature of abilities by lower dimensional factors and correct for measurement errors. This study estimates the endogenous parental investment function jointly with the dynamic model of cognitive skills, noncognitive skills and health.

The results show that self-effects are present and strong in the production of all human capital dimensions. That is, skills produce skills and health produces health. The results also confirm the existence of cross-effects, except cross-productivity from cognitive and health to noncognitive skills: the existing stock of noncognitive skills and health foster the production process of cognitive skills; high cognitive and noncognitive skills lead to better health; and cognitive skills and health are unimportant to noncognitive skill development. These results indicate that there is a high cost for the accumulation of human capital for those who start with lower skill and health levels.

Most importantly, the results confirm the vital importance of parental investments. First, investments strongly and directly affect the accumulation of skills and health. Their impacts are the largest for noncognitive skills, followed by health and cognition. Second, there is a dynamic complementarity among the inputs in human capital production. This implies that returns to investments are higher for children with better initial conditions. Furthermore, it also implies that higher initial stocks of skills and health make those skills and health more productive.

The self-productivity, cross-productivity and dynamic complementarity together become a dynamic multiplier effect mechanism of skill and health accumulation whereby skills, health and parental investments produce skills and health. Furthermore, parental investment decisions strongly depend on parental skills and wealth. This indicates that children with better backgrounds get more investments, and they can also use these investments more productively. These effects could lead to substantially different growth rates of human capital and substantial increases in inequality in producing skills and finally lead to social inequality. The results provide some insight as to how parents make investment choices in their children and show the role of parental investments as a source of child development and inequality. The findings also indicate the importance of interventions by boosting investments at early ages that can alter child development path, especially for disadvantaged children. A lack of parental investments can seriously hinder the development of a child.

My results provide important evidence that skills and health are produced from a combination of an individual's skills and health, parental skills and investments and other individual and family factors. Therefore, policies and interventions to develop human capital need to take into account of the complex interactions over childhood among these factors. Policies, interventions and investments in children at early ages are key to improving skills and health deficits in human development and contributing to social inequality reduction.

Appendix A: The Construction of Measures of Child's and Parental Noncognitive Skills and Quality of Relationship

Table A.1: Construction of Measures of Noncognitive Skills and Quality of Relationship

Scale/Index	**Items/Statements**
Child self-esteem - Round 3, age 15	The self-esteem scale is constructed using the following items/Statements: 1. 'I am proud of my clothes'; 2. 'I feel my clothing is right for all occasions'; 3. 'I am proud of my shoes or of having shoes'; 4. 'I am proud because I have the right books, pencils or other equipment for school'; 5. 'I am proud that I have the correct uniform'; 6. 'I am proud of the work I have to do'.
Child self-esteem - Round 2, age 12	The items are: 1. 'I am proud of my clothes'; 2. 'I am proud of my shoes or of having shoes'; 3. 'I am proud because I have the right supplies for school'; 4. 'I am proud that I have the correct uniform'; 5. 'I feel proud to show my friends where I live'. 6. 'I feel proud of the job done by the head of household'. 7. 'I am proud of my achievements at school'.

Continued on the next page

Table A.1: The Construction of Measures *Continued*

Scale	Items/Statements
Child self-efficacy - Round 3, age 15	The items are: 1. ‘If we try hard we can improve my situation in life’; 2. ‘Other people in my family make all the decisions about how we spend my time’;* 3. ‘I like to make plans for my future studies and work’; 4. ’If we study hard we will be rewarded with a better job in the future’; 5. ‘I have a choice about the work I do’.
Child self-efficacy - Round 2, age 12	The items are: 1. ‘If we try hard we can improve my situation in life’; 2. ‘Other people in my family make all the decisions about how we spend my time’;* 3. ‘I like to make plans for my future studies and work’; 4. ’If we study hard we will be rewarded with a better job in the future’;

Continued on the next page

Table A.1: The Construction of Measures *Continued*

Scale	Items/Statements
Child self-respect and Inclusion - Round 3, age 15	1. 'When I am at the shops/market I am usually treated by others with fairness and respect'; 2. 'Adults in my community treat me as well as they treat other children of my age'; 3. 'The other children in my class treat me with respect'; 4. 'Other pupils in my class tease me at school'; 5. 'My friends will stand by me during difficult times'; 6. 'I feel I belong at my school'; 7. 'My friends look up to me as a leader'; 8. 'I have people I look up to' 9. 'I have opportunities to develop job skills'.
Child self-respect and Inclusion - Round 2, age 12	1. 'At the shops I am treated with fairness'; 2. 'Adults in my street treat me worse than other children of my age';[*] 3. 'The other children in my class treat me with respect'; 4. 'Other pupils in my class tease me at school';[a] 5. 'My teachers treat me worse than other children';[*]
Parental self-esteem - measured at Round 2	1. ' I feel proud to show my friends or other visitors where I live'; 2. 'I am ashamed of my clothes';[*] 3. 'I feel proud of the job done by the household head'; 4. 'The job I do makes me feel proud'; 5. 'I feel proud of my children;

Continued on the next page

Table A.1: The Construction of Measures *Continued*

Scale	**Items/Statements**
Parental self-efficacy - measured at Round 2	The items are: 1. ‘If we try hard we can improve my situation in life’; 2. ‘I like to make plans for my future studies and work’; 3. ‘I have no choice about which school to send my child to’;* 4. ’If my child gets sick I can do little to help him/her get better’;* 5. ’I can do little to help my child do well in school no matter how hard I try’;*
Parental self-respect and Inclusion - measured at Round 2	1. ‘At the shops I am treated with fairness and respect’; 2. ‘Other people in the street look down on me and my family’;* 3. ‘My children’s teachers are unfriendly or rude to me’;*

Continued on the next page

Table A.1: The Construction of Measures *Continued*

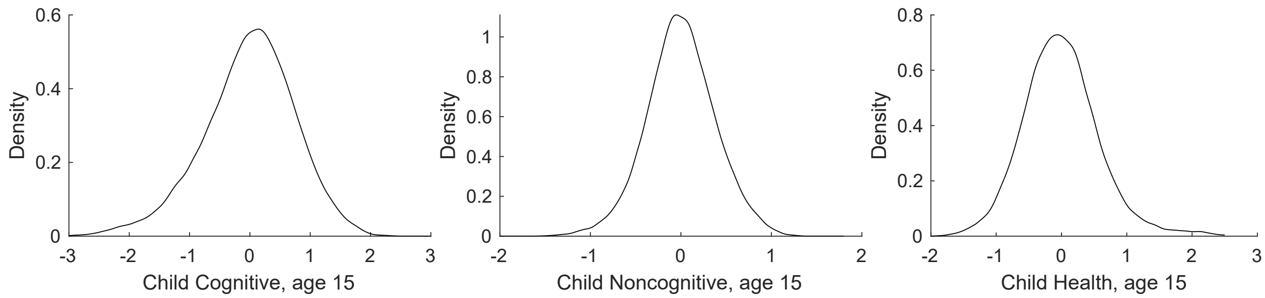
Scale	Items/Statements
Quality of relationship - Round 3	1. 'I always obey my parents' ; 2. 'My parents rarely talk to me about the things that matter to me';* 3. 'I always feel loved by my parents'; 4. ' My parents never support me in the things I want to do';* 5. 'I usually feel able to speak my views and feelings with my parents'; 6. Most of the time my parents treat me fairly when I do something wrong'; 7. 'Compared to my sisters fewer things are provided for me';* 8. 'I receive lots of time and attention from my parents'; 9. 'Compared to my brothers fewer things are provided for me';* 10. 'Compared to my brothers I have less freedom to leave the house when I want';* 11. 'Compared to my sisters I have less freedom to leave the house when I want';* 12. 'My parents treat me worse than other children in my family';*

* The item is recoded to reflect a positive statement

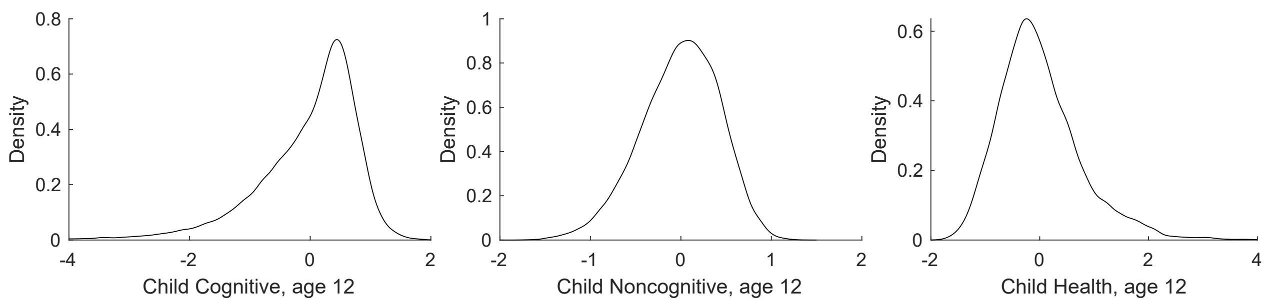
Appendix B: Factor Moments

Figure B.1: Factor Distributions

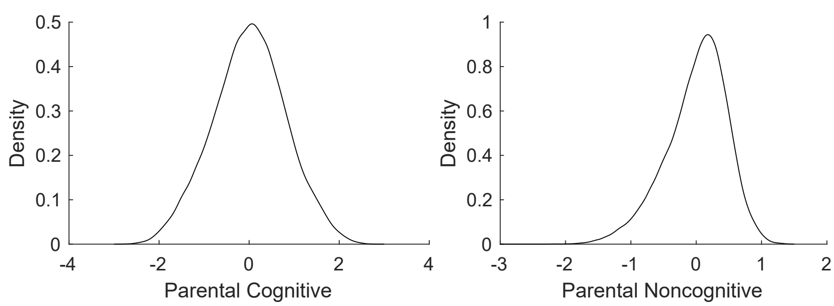
(a)



(b)



(c)



(d)

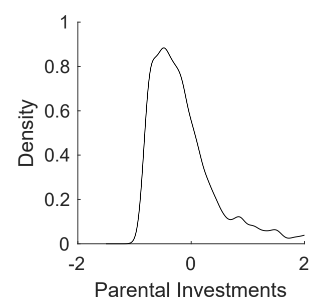


Table B.1: Factor Means, Standard Deviations and Correlations

	Child Cognitive skill, age 15	Child Noncognitive skill, age 15	Child Health, age 15	Parental Investment	Child Cognitive skill, age 12	Child Noncognitive skill, age 12	Child Health, age 12	Parental Cognitive skill	Parental Noncognitive skill
Factore means	-0.016	0.001	0.010	0.010	-0.033	-0.008	-0.001	0.007	0.001
	(0.019)	(0.010)	(0.015)	(0.017)	(0.017)	(0.009)	(0.018)	(0.018)	(0.009)
Factor standard deviations	0.765	0.381	0.623	1.001	0.834	0.435	0.758	0.818	0.471
	(0.021)	(0.015)	(0.021)	(0.050)	(0.035)	(0.013)	(0.017)	(0.015)	(0.021)
Factor correlation:									
Child's cognitive skill at age 15	1	–	–	–	–	–	–	–	–
Child's noncognitive skill at age 15	0.288	1	–	–	–	–	–	–	–
	(0.022)								
Child's health at age 15	0.275	0.115	1	–	–	–	–	–	–
	(0.024)	(0.019)							
Parental investment	0.471	0.198	0.231	1	–	–	–	–	–
	(0.018)	(0.017)	(0.020)						
Child's cognitive skill at age 12	0.830	0.293	0.265	0.383	1	–	–	–	–
	(0.014)	(0.021)	(0.022)	(0.013)					
Child's noncognitive skill at age 12	0.271	0.246	0.122	0.152	0.237	1	–	–	–
	(0.026)	(0.028)	(0.024)	(0.019)	(0.027)				
Child's health at age 12	0.367	0.164	0.835	0.381	0.295	0.146	1	–	–
	(0.021)	(0.018)	(0.016)	(0.019)	(0.019)	(0.022)			
Parent's cognitive skill	0.652	0.289	0.202	0.503	0.681	0.306	0.274	1	–
	(0.015)	(0.021)	(0.023)	(0.023)	(0.013)	(0.023)	(0.022)		
Parent's noncognitive skill	0.214	0.197	0.118	0.212	0.182	0.700	0.150	0.310	1
	(0.023)	(0.029)	(0.027)	(0.013)	(0.023)	(0.028)	(0.022)	(0.022)	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Table B.2: Mixture Component Means

Factor	Type 1	Type 2
Child's cognitive skill at age 15	-0.209	0.608
	(0.028)	(0.032)
Child's noncognitive skill at age 15	-0.037	0.123
	(0.012)	(0.018)
Child's health at age 15	-0.107	0.386
	(0.018)	(0.033)
Parental investments	-0.324	1.085
	(0.014)	(0.097)
Child's cognitive skill at age 12	-0.201	0.511
	(0.029)	(0.016)
Child's noncognitive skill at age 12	-0.030	0.062
	(0.011)	(0.029)
Child's health at age 12	-0.223	0.714
	(0.020)	(0.055)
Parental cognitive skill	-0.181	0.611
	(0.025)	(0.048)
Parental noncognitive skill	-0.062	0.202
	(0.012)	(0.026)
Type share	0.763	0.237
	(0.019)	(0.019)

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Appendix C: Marginal Products of the CES Functions

This appendix derives the marginal products of inputs for the CES production function indicated in Equation 2.2, which is:

$$\begin{aligned}\Theta_{i,t}^{k} = \ & \left[\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + +\gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}}\right. \\ & \left.\gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}\right]^{1/\rho^{tk}} e^{X_{it}'\delta_{t}^{k} + A_{t}^{k} + \varepsilon_{it}^{k}}\end{aligned} \tag{C.1}$$

Then the marginal product of cognitive skills at time t (age 12) with the outputs being cognitive skills, noncognitive skills and health at time t (age 15) - $ln(\Theta_{i,t}^{k})/ln(\Theta_{i,t-1}^{C})$ for k $\in$ (C, N, H) are derived as follows:

$$\begin{aligned}ln(\Theta_{i,t}^{k}) = \ & \tfrac{1}{\rho^{tk}} ln\left[\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} +\right. \\ & \left.\gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}\right] + X_{it}'\delta_{t}^{k} + A_{t}^{k} + \varepsilon_{it}^{k}\end{aligned} \tag{C.2}$$

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial \Theta_{i,t-1}^{C}} = \frac{1}{\rho^{tk}}\left(\frac{1}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right)\rho^{tk}\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}-1}$$

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial \Theta_{i,t-1}^{C}/\Theta_{i,t-1}^{C}} = \frac{1}{\rho^{tk}}\left(\frac{1}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right)\rho^{tk}\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}}$$

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial ln \Theta_{i,t-1}^{C}} = \left(\frac{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}}}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right) \tag{C.3}$$

Similarly, the marginal products of child noncognitive skills, child health, parental investment and parental cognitive and noncognitive skills are:

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial ln \Theta_{i,t-1}^{NC}} = \left(\frac{\gamma_{2,t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}}}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right) \tag{C.4}$$

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial ln \Theta_{i,t-1}^{H}} = \left(\frac{\gamma_{3,t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}}}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right) \tag{C.5}$$

$$\frac{\partial ln(\Theta_{i,t}^{k})}{\partial ln \Theta_{i,t-1}^{I}} = \left(\frac{\gamma_{4,t}^{k}(\Theta_{i,t-1}^{I})^{\rho^{tk}}}{\gamma_{1,t}^{k}(\Theta_{i,t-1}^{C})^{\rho^{tk}} + \gamma_{2t}^{k}(\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^{k}(\Theta_{i,t-1}^{H})^{\rho^{tk}} + \gamma_{4t}^{k}(I_{i,t})^{\rho^{tk}} + \gamma_{5t}^{k}(P_{i}^{C})^{\rho^{tk}} + \gamma_{6t}^{k}(P_{i}^{NC})^{\rho^{tk}}}\right) \tag{C.6}$$

$$\frac{\partial ln(\Theta^k_{i,t})}{\partial ln\Theta^{PC}_{i,t-1}} = \left(\frac{\gamma^k_{5,t}(\Theta^{PC}_{i,t-1})^{\rho^{tk}}}{\gamma^k_{1,t}(\Theta^{C}_{i,t-1})^{\rho^{tk}} + \gamma^k_{2t}(\Theta^{NC}_{i,t-1})^{\rho^{tk}} + \gamma^k_{3t}(\Theta^{H}_{i,t-1})^{\rho^{tk}} + \gamma^k_{4t}(I_{i,t})^{\rho^{tk}} + \gamma^k_{5t}(P^C_i)^{\rho^{tk}} + \gamma^k_{6t}(P^{NC}_i)^{\rho^{tk}}} \right) \tag{C.7}$$

$$\frac{\partial ln(\Theta^k_{i,t})}{\partial ln\Theta^{PN}_{i,t-1}} = \left(\frac{\gamma^k_{6,t}(\Theta^{PN}_{i,t-1})^{\rho^{tk}}}{\gamma^k_{1,t}(\Theta^{C}_{i,t-1})^{\rho^{tk}} + \gamma^k_{2t}(\Theta^{NC}_{i,t-1})^{\rho^{tk}} + \gamma^k_{3t}(\Theta^{H}_{i,t-1})^{\rho^{tk}} + \gamma^k_{4t}(I_{i,t})^{\rho^{tk}} + \gamma^k_{5t}(P^C_i)^{\rho^{tk}} + \gamma^k_{6t}(P^{NC}_i)^{\rho^{tk}}} \right) \tag{C.8}$$

Appendix D: Estimates of Production Functions without Endogenous Investments

Table D.1: Estimates of Production Functions without Endogenous Investments

	Cognitive skills at age 15 (1)	Noncognitive skills at age 15 (2)	Health at age 15 (3)
Child's cognitive skills at age 12	0.619***	0.000	0.035
	(0.044)	(0.000)	(0.028)
Child's noncognitive skills at age 12	0.152***	0.425***	0.123***
	(0.055)	(0.080)	(0.043)
Child's health at age 12	0.079***	0.071***	0.801***
	(0.020)	(0.021)	(0.025)
Parental investments	0.120***	0.396***	-0.051**
	(0.031)	(0.026)	(0.021)
Parental cognitive skills	0.071**	0.007	0.012
	(0.033)	(0.006)	(0.030)
Parental noncognitive skills	-0.042	0.101	0.080*
	(0.047)	(0.074)	(0.044)
A_t	-0.033	0.223***	0.164***
	(0.021)	(0.020)	(0.021)
Female	0.071***	-0.027*	-0.212***
	(0.015)	(0.014)	(0.023)
Urban	0.071***	-0.259***	-0.129***
	(0.027)	(0.025)	(0.022)
Number of siblings aged 0-18	0.008	0.035***	0.002
	(0.008)	(0.007)	(0.006)
Complementarity(ρ)	-0.150	-2.423***	-0.355
	(0.137)	(0.259)	(0.224)
Elasticity of substitution	0.870***	0.292***	0.738***
	(0.122)	(0.023)	(0.111)
Observations		*961*	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table D.2: Marginal Effects - Production Functions without Endogenous Investments

	Cognitive skills at age 15 (1)	Noncognitive skills at age 15 (2)	Health at age 15 (3)
Child's cognitive skills at age 12	0.619***	0.001	0.036
	(0.043)	(0.001)	(0.028)
Child's noncognitive skills at age 12	0.152***	0.356***	0.125***
	(0.055)	(0.061)	(0.043)
Child's health at age 12	0.079***	0.107***	0.796***
	(0.019)	(0.019)	(0.022)
Parental investments	0.120***	0.427***	-0.052**
	(0.032)	(0.024)	(0.022)
Parental cognitive skills	0.071**	0.015*	0.012
	(0.033)	(0.009)	(0.032)
Parental noncognitive skills	-0.042	0.094	0.082*
	(0.048)	(0.063)	(0.046)
Observations		*961*	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Chapter 3

The Consequences of Bullying Victimization on Health and Psychosocial Outcomes in Young Children

3.1 Introduction

Bullying victimization is a very common experience among adolescents and a global phenomenon. Evidence shows that one in three children is a victim of peer bullying (Due and Holstein, 2008; UNESCO, 2019). Being bullied has not only intermediate negative impacts but also long-lasting effects on physical and mental health, health and social behaviors, psychological well-being and earnings (Olweus, 1997; Lereya et al., 2015; Brown et al., 2008; Takizawa et al., 2014; Smith and Brain, 2000). While most research on this phenomenon has been conducted in high-income countries, far less is known about bullying victimization and its consequences on health and psychosocial outcomes in low and middle-income countries due to data limits.

I use a rich set of data on diversified aspects of adolescents, their families, and communities from the Older cohort of the Young Lives study in Vietnam to examine the consequences of bullying victimization on health and psychosocial outcomes including alcohol consumption, self-rated health, subjective well-being or life satisfaction and distress.

My contributions are threefold. First, I examine the heterogeneous effects of bullying.

Previous studies have typically used simple observed measures of bullying. I use structural estimations of underlying, true or latent bullying rather than measured indicators. This approach aims to identify unobserved heterogeneity to more accurately capture the richness of the variation in the type and frequency of bullying and correct potential measurement errors in the self-reported bullying measures. Second, I deal with the endogeneity of bullying arising from selection in terms of unobservable and observable factors and reserve causality in the various forms of bullying by using an instrumental variable approach. I use the number of a child's friends being physically bullied and the percentage of children in the Young Lives survey clusters being physically punished by their parents as instruments for bullying and structural model. Third, this study contributes to the very limited evidence about the impacts of a variety of types of bullying on various health and psychosocial outcomes in a developing country setting because of data scarcity.

The rest of the chapter is structured as follows. Section 3.2 discusses the literature. Section 3.3 introduces data, definitions and measures used in this chapter. Section 3.4 uses conventional regression methods to produce key findings usually reported in the previous literature and discusses problems with this approach. Section 3.5 discusses the empirical strategy. Section 3.6 presents the key results, and section 3.7 concludes.

3.2 Literature Review

Bullying victimization has been paid attention to in education, psychology, and sociology. The literature examining bullying and its impacts in these fields can be found in Boulton and Underwood (1993), Ouellet-Morin et al. (2011), Ladd et al. (2017), Arseneault et al. (2010), Smith et al. (2004), among others. However, economics research examining bullying victimization is scarce, especially in developing countries due to the lack of data about bullying and there is little research on bullying addressing the endogeneity and non-random selection issues in bullying victimization.

Evidence shows that bullying victimization negatively affects children's health and psychosocial well-being (Olweus, 1996; Lereya et al., 2015; Hawker and Boulton, 2000; Nguyen et al., 2016). There are some studies examining bullying victimization and child health, risk behaviors and psychosocial problems in developing countries. Crookston et al. (2014), Lister et al. (2015b), and Lister et al. (2015a) consider independently the consequences of some experiences of bullying victimization in Peru and their findings suggest the association between victimization and health as well as risky behaviors and psychosocial problems. Crookston et al. (2014) indicate that children experiencing bullying at both ages 8 and 15 were 1.58 times more likely to smoke cigarettes, 1.57 times more likely to drink alcohol, and 2.17 times more likely to have ever had a sexual relationship compared with all other children. Malhi et al. (2014) show that victimized children likely have a lower self-concept and a higher risk of emotional difficulties than non-victimized youth.

Although the above studies are informative, they provide limited quantitative evidence about the relationships between bullying victimization and poor health, risky behaviors and psychosocial problems. First, these studies do not consider selection and endogeneity issues either through reverse causality or uncontrolled confounding variables and measurement error problems. Bullying victims are not randomly selected in some observable and unobservable factors that confound the consequences of bullying, and bullying and outcomes of interest are jointly determined (Eriksen et al., 2014; Sarzosa and Urzúa, 2021; also see the Summary Statistics in Section 3.3 below). The obvious solutions for this problem are by using instrumental variables and structural models. Second, these studies used observable measures independently that are imperfect proxies for unobservable heterogeneity and face measurement error problems in measuring bullying.

Eriksen et al. (2014) use instrumental variables to address the endogeneity issue of victimization. Using Danish data, they instrument bullying with the proportion of children in the class whose parents have a criminal conviction. They confirm the detrimental impact of being bullied on 9th-grade GPA. However, they do not deal appropriately with measurement errors and unobservable heterogeneity.

Sarzosa and Urzúa (2021) use Korean data of 14-18 year old children to examine the effects of bullying on mental and physical health and risky behaviors. They estimate a structural model to deal with the endogeneity that rises from the observable and intrinsic unobservable characteristics that determine bullying and also influence the outcomes of interest. The authors control for unobserved heterogeneity in the form of cognitive and noncognitive skills. They define children as bullying victims if they are severely teased, threatened, collectively harassed, severely beaten, or robbed, but bullying victimization is ultimately defined as a single binary variable. They show that bullying victims at age 15 increases sickness, mental health issues and stress caused by friendships by 6.5%, 7.9% and 23.5% of a standard deviation, respectively, at age 18.

I aim to extend the existing literature by combining both approaches of instrumental variables and structural modeling to address the endogeneity of bullying victimization, correct measurement errors and control for unobserved heterogeneity in the form of bullying victimization. Victimization is self-reported and subjective; it is likely to suffer from measurement errors that might cause effects to be attenuated. Therefore, I directly correct the measurement errors of bullying reports and use a latent bullying victimization factor as a source of unobserved heterogeneity (Hu and Schennach, 2008; Cunha and Heckman, 2008; Schennach, 2004; Sarzosa and Urzúa, 2021). I instrument victimization with the number of children's friends being physically bullied and the percentage of children in the Young Lives survey clusters being physically punished by their parents. Although the literature focuses more on physical victimization, evidence shows that other types of bullying behaviors, including verbal, relational and property victimization, have as adverse effects as physical bullying on physical and mental health and psychosocial well-being, if not greater (Carbone-Lopez et al., 2010; Rivers and Smith, 1994). In addition, relational or indirect victimization is less easily detected since children are less likely to tell an adult if they are relationally abused (Rivers and Smith, 1994). I consider different types of bullying and my methodology allows me to combine these types of victimization into a single and composite factor that reflects the true, latent and unobservable measure of bullying victimization.

3.3 Data, Definition and Measures

I use data from the Young Lives study - a longitudinal study of childhood poverty that tracks two cohorts of children from four developing countries: Ethiopia, India, Peru and Viet Nam. In each country, the younger cohort follows 2,000 children every 4 years from age 1 until age 15 and the older cohort tracks 1,000 children every 4 years from age 8 to 22. The detailed data on adolescent peer bully victimization are only available for the Older Cohort in Round 3 when the children are 15 years of age. Round 3 of the survey also provides detailed measures of physical and mental health, health and social behaviors and psychological well-being. Hence, this research focuses on data from the older cohort for Vietnam in Round 3 when kids are 15 years old. The focus of the research on 15-years-old children is not only because data is available, but more importantly this is a critical developmental period of children in which they are susceptible to bullying, risky behaviors and psychosocial adjustments that initiate behaviors and outcomes later in life.

3.3.1 Measures of Peer Bullying Victimization

Bullying victimization is defined as repeated and intentional exposure to hostile actions that cause harm or discomfort over time by others (Rigby, 2002; Olweus, 1993). The definition of bullying typically has three features: intentional acts of aggression, repetition and an imbalance of power (Olweus, 1993; Olweus, 1997). Aggression may be direct or indirect. Direct aggression may include physical and verbal attacks and attacks on property and indirect aggression involves actions aimed at social isolation, exclusion and manipulation and it is usually referred to as social or relational bullying (Mynard and Joseph, 2000). This research focuses on being bullied and these four forms of bullying exposure, given the data availability in the Young Lives study.

Bullying and victimization in the Young Lives study are assessed based on the 9-item self-administered questionnaires. These nine items use the 9-item Social and Health Assessment Peer Victimization Scale (Ruchkin et al., 2004), an adapted version of the longer Multidimensional Peer Victimization Scale (Mynard and Joseph, 2000). The scale has been used and validated in multiple studies in multiple countries, usually in school environments (Crookston et al., 2014; Cluver and Orkin, 2009; Karlsson et al., 2014; Stadler et al., 2010; Boyes et al., 2014). The scale was piloted before being applied to the Young Lives survey. The questionnaires asked children whether other young people had bullied them and, if so, how frequently they had experienced each bullying behavior during the past year: never, once, two or three times, or four or more times. The questions addressed exposure to 9 forms (9 items) of bullying corresponding to 4 types of bullying: physical (punched, kicked, or beaten up; hurt physically in some other way), verbal (called names or sworn at; made fun of for some reason), relational (tried to cause trouble with the youth's friends, refused to talk to youth, made youth uncomfortable by staring), and property victimization (broke or damaged property; took something without permission or stole something). 9 items and the corresponding 4 types of bullying victimization are listed in Table 3.1.

In line with the definition of bullying as repetitive actions, I exclude random, one-off incidents of victimization, and individuals are considered being bullied if they experienced each behavior of being bullied twice or more times.

Table 3.2 shows the percentage of youth who experienced different forms of bullying. Prevalence for repeated experience of bullying range from 3.9% to 21.0%. Based on the types of bullying, the experience of these types ranges from 6.9% to 27.1%. In total, 37.9% of children reported being victims of bullying over the last year. The proportion of boys suffering physical victimization is much higher than that of girls, while girls tend to be more likely to experience relational and property victimization.

Table 3.1: Measures of Bullying Victimization

9 items of victimization	
Children were asked the questions: during the last 12 months, I want to know whether other young people did the following bullying behaviors. Response options are Never, Once, 2-3 times, 4 or more times. In this study, the response options are recoded as a binary variable equal to 1 if the options are '2–3 times' or '4 or more times' or children experienced these bullying behaviors at least twice and 0 otherwise.	
1. punched, kicked or beat you up	
2. hurt you physically in some other way	
3. made fun of you for some reason	
4. called you names or swore at you	
5. refused to talk to you or made other people not talk to you	
6. tried to get you into trouble with your friends	
7. made you uncomfortable by staring at you for a long time	
8. took something without permission or stole things from you	
9. tried to break or damaged something of yours	
Overall victimization	Indicates whether a child has been victimized of any types above. It equals 1 if the child experienced any of the above victimizing behaviors twice or more and zero otherwise.
Types of bullying	
Physical victimization	Indicates whether a child has been physically victimized. It takes a value of 1 if any of items 1 and 2 is 1 and 0 otherwise.
Verbal victimization	Indicates whether a child has been verbally victimized. It takes a value of 1 if any of items 3 and 4 is 1 and 0 otherwise.
Relational victimization	Indicates whether a child has been relationally victimized. It takes a value of 1 if any of items 5,6 and 7 is 1 and 0 otherwise.
Attacks on property	Indicates whether a child has been victimized by property attacks. It takes a value of 1 if any of items 8 and 9 is 1 and 0 otherwise.

Table 3.2: Percentage of Children Experiencing Different Forms of Bullying Two or More Times (%)

	Full	Male	Female
Physical victimization	***6.9***	***9.0***	***4.9***
Punched, kicked or beaten up	5.3	7.6	3.1
Hurt physically	3.9	4.0	3.7
Verbal victimization	***20.1***	***20.2***	***20.0***
Made fun of	17.7	17.4	17.9
Called or swore	7.2	8.0	6.4
Relational victimization	***27.1***	***25.8***	***28.3***
Refused to talk	5.0	4.5	5.6
Friend trouble	8.9	8.5	9.2
stared at	21.0	20.1	21.9
Attacks on property	***8.9***	***7.7***	***10.0***
Theft	6.1	4.9	7.2
Property damage	5.1	4.9	5.3
Victimization	***37.9***	***35.8***	***39.9***
Observations	971	480	491

3.3.2 Measures of Outcomes

The literature section shows that bullying victimization is associated with different outcomes. I examine the consequences of being bullied on physical health directly and indirectly measured by self-rated health and alcohol use; mental health measured by distress and subjective well-being.

Self-rated health. Health status is a crucial factor of well-being. Many indicators, such as nutritional status and mental illness, can be used to measure health. Self-rated health is considered as incorporating physical, social, emotional, and mental aspects of well-being and is distinct from measures of well-being such as life satisfaction (Fosse and Haas, 2009; Joffer et al., 2016). Evidence has shown that victimization is associated with self-perceived health (Abada et al., 2008; Boynton-Jarrett et al., 2008; Gobina et al., 2008). Self-rated health in the Young Live survey is assessed based on an international health question that is extensively used to assess health in cross-national studies. The question asked the youth to rate their general health on a 5-point Likert scale, with which 1 indicates very poor and 5 very good health status. The response is then dichotomized as 0 indicating poor health status and 1 denoting good health based on the mean value of the response. The mean value of self-perceived health is 3.36; hence health status take 0 if the child reported their health between 0 to 3 and take 1 if the child reported their health between 4 to 5.

Subjective well-being. Subjective well-being is defined as a good quality of life and a state of satisfaction with life that people evaluate their own lives and reflect their affective reactions to experiences (OECD, 2013; Bourdillon and Boyden, 2014). It is distinct from objective well-being which refers to facts about their lives. Subjective well-being can refer to life satisfaction. Young Lives uses Cantril's ladder of life to assess subjective well-being (Cantril et al., 1965). Individuals were asked the question 'where on the ladder do you feel you personally stand at present time?' and the responses ranged from 1-9, higher scores indicated better subjective well-being. The responses are further recoded as 0 indicating low subjective well-being and 1 denoting high subjective well-being based on the mean value of the response. The mean value of the response is 5.59, then the values of responses between 1 and 5 are recoded as 0 and the values between 6 and 9 are recoded as 1.

Alcohol consumption. Alcohol consumption is a health risk behavior that often begins in youth and affects other risky behaviors, health, social, and economic problems (WHO, 2019; Crookston et al., 2014). The initiation of drinking alcohol at a young age is a predictor of substance use problems later in life. Alcohol consumption in this study is assessed based on the youth's response to the question in the self-administered questionnaire: 'How often do you usually drink alcohol?'. Alcohol consumption is defined as a dummy variable equal to 1 for those who reported alcohol use at least once a month.

Emotional and mental distress. Young Lives uses the five-item Emotional Difficulties subscale of the Strengths and Difficulties Questionnaire (SDQ) to assess distress (Goodman and Scott, 1999). The questions are in the self-administered questionnaire in which individuals were read statements and asked whether they strongly disagreed, disagreed, agreed or strongly agreed with the statement. The statements include 'I worry a lot'; 'I get a lot of headaches, stomach aches, or sickness'; 'I am often unhappy, downhearted, or tearful'; 'I am nervous in new situations', and 'I have many fears or am easily scared'. The scale is the average of these items.

3.3.3 Summary Statistics

Table 3.3 shows the summary statistics by victimization status. Adolescents who reported victimization are slightly more likely to be girls, from the Kinh Ethic group and smaller households.[1] There tends to be no difference between victimized and non-victimized children by socio-economic status including household wealth index, mother's and father's education levels and child age. However, there appear to be significant differences in their outcomes between victimized and non-victimized children. Children who faced bullying victimization tend to have worse health, use more alcohol and experience more emotional and mental distresses than their non-victimized counterparts. This suggests that victims are not randomly selected in terms of observable and unobservable factors.

3.4 Conventional Regressions and Endogeneity Issue

This section presents traditional ordinary least squares (OLS) and logistic regressions to provide preliminary evidence and discuss problems of measurement errors and endogeneity in this approach that motivate my approach.

I regress the effects of bullying (B) on the outcomes of interest (Y_m) by standard OLS and logistic methods. Our outcomes of interest include subjective well-being or life satisfaction,

[1] Although boys are much more likely to be physically abused, girls tend to be more likely to suffer relational victimization and attacks on their property. Therefore, in terms of general victimization, girls are slightly more likely to suffer victimization than boys (See Table 3.2).

Table 3.3: Summary Statistics by Victimization Status

Variables	Full sample (1)	Non-victimized (2)	Victimized[a] (3)	Difference (2) - (3)
Female	0.506	0.489	0.533	-0.043*
	(0.500)	(0.500)	(0.500)	
Urban	0.198	0.186	0.217	-0.032
	(0.398)	(0.389)	(0.413)	
Ethnic	0.870	0.854	0.897	-0.043**
	(0.336)	(0.353)	(0.305)	
Child age (in years)	15.052	15.046	15.063	-0.017
	(0.321)	(0.324)	(0.315)	
Wealth index	0.623	0.618	0.631	-0.014
	(0.185)	(0.187)	(0.181)	
Mother's education	6.143	6.123	6.177	-0.054
	(3.900)	(3.996)	(3.741)	
Father's education	6.715	6.675	6.780	-0.105
	(4.104)	(4.106)	(4.105)	
Household size	4.541	4.609	4.429	0.179**
	(1.356)	(1.338)	(1.379)	
Birth order	2.202	2.189	2.223	-0.034
	(1.334)	(1.307)	(1.379)	
Mother or father alive	0.933	0.941	0.922	0.019
	(0.249)	(0.237)	(0.269)	
Outcome variables:[b]				
Subjective well-being	0.397	0.393	0.405	-0.012
	(0.490)	(0.489)	(0.492)	
Self-rated health	0.502	0.528	0.458	0.070**
	(0.500)	(0.500)	(0.499)	
Alcohol drinking	0.281	0.223	0.375	-0.152***
	(0.450)	(0.417)	(0.485)	
Emotional distress	1.727	1.636	1.873	-0.236***
	(0.431)	(0.422)	(0.406)	
Observations	971	603	368	971

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively from a means t-test between non-victimized and victimized children.

[a] The construction of the victimization variable is detailed in Section 3.3.1 and Appendix Table A.1.

[b] Definitions, statements and computation of outcomes are discussed in Section 3.3.2 and Appendix A.1.

self-perceived health, alcohol consumption and distress. Bullying variables in childhood include the overall bullying victimization status and the four types of bullying: physical victimization, verbal victimization, relational victimization and attacks on property. I want to examine the effects of being bullied (B) on these outcomes, the model, therefore, takes the form:

$$Y_m = \alpha X_m + \beta B + u_{Y_m} \tag{3.1}$$

Where X_m is a vector of observed controls. B are the overall victimization indicator and indicators of the four types of bullying victims. u_{Y_m} is an error term. Table 3.4 shows the estimation results with the overall bullying victimization variable and the result suggests that bullying victimization is negatively associated with health and positively correlated with the likelihood of drinking and distress, while there seems to be no significant association between victimization and subjective well-being. Table 3.5 presents the impacts of four different types of victimization and shows interesting features of the conventional approach: only six out of sixteen results for four types of victimization show significant correlations with the outcomes of interest.

Although these results are very informative, this approach ignores two fundamental issues that can confound the results. First, it is strongly convinced that victimization is endogenous. There is likely reverse causality between Y_m and B, and they are jointly determined by observable and unobservable confounding variables. Additionally, victimization is not randomly selected; victims of bullying may be systematically different from non-victims in some unobservable and observable factors that can affect outcomes. Therefore, these factors can confound the consequence of victimization. Second, this approach ignores measurement error issues in terms of bullying victimization by using imperfect proxies (different observable bullying variables) for bullying victimization. Victimization is self-reported and subjective, it is likely to suffer from measurement errors. Different observed measures are not good proxies for true bullying. Measurement errors in B can be correlated with the error terms, u_{Y_D}, in Equation 3.1. Therefore, the evidence provided by this approach is limited. The alternative way to address these problems is using instrumental variables, structural models and latent factors for bullying instead of their observed measures. My approach uses instrumental variables and a structural model with factor analysis to overcome these challenges. In particular, bullying victimization is measured with errors, and I deal with this issue by latent factor models. Bullying victimization can be endogenously determined by individual observable and unobservable characteristics and I use an instrumental variable approach that is adapted to the latent factor structure of the model to address this issue.

3.5 Empirical Strategy

This section introduces a model with a latent factor structure and endogenous victimization. The approach adapts Heckman et al. (2006) and Cunha et al. (2010) to the analysis of victimization.

3.5.1 Identification of Factors

Following the approach of Heckman et al. (2006) and Cunha et al. (2010), I assume that bullying victimization is a latent or true factor instead of their observed measures that, in turn, determine outcomes. Besides the bullying victimization factor, I capture children's family background characteristics by an unobservable latent factor referred as the family factor. There are two types of observable measures, continuous and binary.

For the victimization measures, let $\mathcal{B}_i$ be a latent bullying factor estimated from the observable measures of bullying victimization that impact the observable measures and B_{ij}, j = {P, V, R, A} denote a 4x1 vectors of the observable victimization measures for individual i including physical victimization (P), verbal victimization (V), relational victimization (R) and attacks on property (A). Bullying victimization measures are binary and we only observe

Table 3.4: Conventional Regressions: Association between Overall Victimization Indicator and Outcomes

	Subjective well-being	Health	Drinking	Distress
Intercept	1.821 (2.516)	-1.858 (2.398)	-2.259 (3.164)	0.254 (0.450)
Female	0.270** (0.120)	0.006 (0.113)	-0.963*** (0.124)	0.148*** (0.021)
Urban	-0.528*** (0.168)	-0.038 (0.168)	-0.573*** (0.214)	-0.099*** (0.028)
Ethnic group	0.498** (0.231)	-0.106 (0.194)	-0.274 (0.195)	0.125*** (0.041)
Child age (in years)	-0.350** (0.173)	0.121 (0.160)	0.093 (0.213)	0.091*** (0.030)
Wealth index	2.610*** (0.445)	0.799** (0.399)	0.061 (0.573)	-0.041 (0.083)
Mother's education	0.025 (0.022)	-0.035* (0.019)	-0.018 (0.025)	-0.011*** (0.004)
Father's education	0.068*** (0.021)	0.010 (0.017)	0.063*** (0.021)	0.000 (0.003)
Household size	-0.030 (0.052)	0.017 (0.045)	-0.058 (0.054)	-0.001 (0.009)
Birth order	-0.036 (0.045)	-0.124*** (0.040)	0.091* (0.047)	0.009 (0.008)
Mother or father alive	0.554* (0.291)	0.107 (0.217)	0.052 (0.233)	-0.081 (0.050)
Bullying victimization	-0.035 (0.116)	-0.284** (0.119)	0.813*** (0.123)	0.221*** (0.021)
No. of obs.	895	895	887	887

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table 3.5: Conventional Regressions: Association between Different Types of Victimization and Outcomes

	Subjective well-being	Health	Drinking	Distress
Intercept	2.052	-1.450	-3.248	0.016
	(2.604)	(2.409)	(3.174)	(0.462)
Female	0.254**	-0.003	-0.924***	0.154***
	(0.119)	(0.117)	(0.126)	(0.022)
Urban	-0.518***	-0.032	-0.590***	-0.103***
	(0.167)	(0.171)	(0.210)	(0.027)
Ethnic group	0.503**	-0.112	-0.252	0.126***
	(0.230)	(0.198)	(0.200)	(0.040)
Child age (in years)	-0.361**	0.099	0.144	0.103***
	(0.178)	(0.161)	(0.212)	(0.030)
Wealth index	2.575***	0.762*	0.313	0.014
	(0.450)	(0.404)	(0.551)	(0.080)
Mother's education	0.025	-0.036*	-0.020	-0.011***
	(0.023)	(0.019)	(0.026)	(0.003)
Father's education	0.071***	0.012	0.061***	-0.000
	(0.022)	(0.017)	(0.022)	(0.003)
Household size	-0.034	0.017	-0.057	-0.001
	(0.051)	(0.045)	(0.055)	(0.009)
Birth order	-0.035	-0.119***	0.085*	0.007
	(0.046)	(0.041)	(0.048)	(0.008)
Mother or father alive	0.531*	0.065	0.174	-0.056
	(0.294)	(0.217)	(0.246)	(0.048)
Physical victimization	-0.275	-0.165	0.708***	0.145***
	(0.280)	(0.241)	(0.264)	(0.049)
Verbal victimization	-0.170	-0.088	0.152	0.044
	(0.154)	(0.169)	(0.201)	(0.027)
Relational victimization	-0.023	-0.379***	0.607***	0.199***
	(0.144)	(0.129)	(0.167)	(0.026)
Attacks on property	0.167	0.079	0.280	0.144***
	(0.208)	(0.200)	(0.220)	(0.039)
No. of obs.	895	895	887	887

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

the measure $B_{ij} = 1$ if $B^*_{ij} > 0$ and $B_{ij} = 0$ otherwise and it is defined as follows:

$$B^*_{ij} = \alpha_j + \beta_j \mathcal{B}_i + u_{ij} \tag{3.2}$$

Where α_j are intercepts, β_j are factor loadings and u_{ij} are measurement error terms.

The observable measures B_{ij} can be viewed as error-ridden indicators of the underlying latent bullying victimization factor and these measures are imperfect proxies for the latent factor. $\mathcal{B}$ is the bullying victimization factor that can be considered as an error-free measure. Measurement errors u_j in this equation reflect that the observable measures are imperfect proxies for the underlying factors.

u_{ij} are assumed to be independent across the measures and all the other errors in the model and logistically distributed, the probability of observing these measures conditional on the unobserved factors is

$$Pr(B_{ij}|\mathcal{B}_i) = \frac{exp(\alpha_j + \beta_j \mathcal{B}_i)^{B_i}}{1 + exp(\alpha_j + \beta_j \mathcal{B}_i)} \tag{3.3}$$

Similarly, for the family background characteristics measures, let $\mathcal{P}_i$ be the family factor and P_{ik} be the observable measures of family background characteristics including wealth index, mother's education, father's education, household size, birth order and whether either parent are alive or not. If observable measures are continuous, then these measures are modelled as a linear function of the factor:

$$P_{ik} = \alpha_k + \beta_k \mathcal{P}_i + v_{ik} \tag{3.4}$$

Where P_{ik} are the continuous observable measures, $\mathcal{P}_i$ is the family factor. v_{ik} is an error term that is not explained by the family factor and is independent across the measures and all the other errors in the model. v_{ik} is assumed to have a normal distributions with mean zero and variance σ_k^2, then the probability of the continuous measures is

$$Pr(P_{ik}|\mathcal{P}_i) = \frac{1}{\sqrt{2\sigma_k^2\pi}} exp\left(-\frac{(P_{ik} - \alpha_k - \beta_k\mathcal{P}_i)^2}{2\sigma_k^2}\right) \tag{3.5}$$

If observable measures are binary, the family factor is extracted from the following equation:

$$P_{ik}^* = \alpha_k + \beta_k\mathcal{P}_i + v_{ik} \tag{3.6}$$

The binary observable measures $P_{ijk} = 1$ if $P_{ik}^* > 0$ and $B_{ij} = 0$ otherwise. u_ij are independent across the family background characteristics measures and all the other errors in the model and logistically distributed. Conditional on the unobservable factor, the probability of these measures is:

$$Pr(P_{ik}|\mathcal{P}_i) = \frac{(exp(\alpha_j + \beta_j\mathcal{P}_i))^{P_{ij}}}{1 + exp(\alpha_j + \beta_j\mathcal{P}_i)} \tag{3.7}$$

From Equations 3.3, 3.5 and 3.7, the likelihood of all the observed measures conditional on $\mathcal{B}_i$ and $\mathcal{P}_i$ is

$$L(B_i, P_i|\mathcal{B}_i, \mathcal{P}_i) = \prod_{j=1}^{J} Pr(B_{ij}|\mathcal{B}_\rangle) \prod_{k=1}^{K} Pr(P_{ij}|\mathcal{P}_i) \tag{3.8}$$

For simplicity, let $\theta_i = \{\mathcal{B}_i, \mathcal{P}_i\}$, then Equation 3.8 can be rewitten as:

$$L(B_i, P_i|\theta_i) = \prod_{j=1}^{J} Pr(B_{ij}|\theta_i) \prod_{k=1}^{K} Pr(P_{ij}|\theta_i) \tag{3.9}$$

The log-likelihood function is:

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^{n} lnL(B_i, P_i) \\ &= \sum_{i=1}^{n} ln\left(\int L(B_i, P_i|\theta)dF(\theta)\right) \\ &= \sum_{i=1}^{n} ln\left(\int L(B_i, P_i|\theta)f(\theta)d\theta\right) \end{aligned} \tag{3.10}$$

Because of the unobservable nature of the factors, the log-likelihood function is constructed by integrating over the distributions of the unobservable factors. I do not impose normality on the latent factors. Instead, I assume that the joint distribution of the latent factors follows a mixture of normals. This assumption is to guarantee enough flexibility for the underlying distribution and it imposes few assumptions on the distributions. I allow for a mixture of C = 8 normals. Therefore, the estimated parameters include the parameters of the normals with mean μ and covariance Ω and mixture probability τ. Then, the probability density function of the factor is $f(\theta) = \sum_{c=1}^{8} \tau_c f(\theta|\mu_c, \Omega_c)$.

The core of a factor model is that each observable measure is a function of a latent/true variable(s) and a measurement error. Each observable measure includes the amount of each common, unobservable factor it loads onto and its measurement error.

I use a latent factor model developed by Cunha et al. (2010) to extract the true, unobservable latent factors from the observable measures and remove the measurement errors. They show that we can non-parametrically identify the factor distributions for the nonlinear measurement system. I follow their framework to identify the latent factors from the observable measures. Identification requires at least $2k + 1$ measures for k factors. I use four dedicated measures loading only onto the bullying factor and six measures loading only onto the family factor in our measurement system. Further requirements for identification are to set the scale and location for the measures. I set the loading for the physical victimization and wealth index for the bullying and family factors respectively equal to one and the remaining coefficients are interpreted in proportion to the normalized coefficients. The constants for these measures are equal to zero.

3.5.2 Outcomes

I am interested in explaining the effects of being bullied on health and psychosocial outcomes. The outcome of interest m for m = 1, 2, . . . , M of person i, Y_i^*, is determined by the bullying victimization factor $\mathcal{B}_i$, family factor $\mathcal{P}_i$ and a set of observable variables X_i

that impact the outcome:

$$Y_{im}^{*} = \beta_{xm} X_{im} + \beta_{\mathcal{B}m} \mathcal{B}_i + \beta_{\mathcal{P}m} \mathcal{P}_i + u_{Y_{im}} \tag{3.11}$$

Outcomes Y_{im}^{*} include subjective well-being or life satisfaction, self-rated health, alcohol consumption and distress. There are two types of outcomes, continuous and binary.

If outcome Y_{im} is continuous, then $Y_{im} = Y_{im}^{*}$, where Y_{im} is the observed continuous outcome of person i and $u_{Y_{im}}$ follow a normal distribution with mean zero and variance σ_m^2, then the probability of this continuous outcome is

$$Pr(Y_{im}|X_{im}, \mathcal{B}_i, \mathcal{P}_i) = \frac{1}{\sqrt{2\sigma_m^2 \pi}} exp\left(-\frac{(Y_{im} - \beta_{xm} X_{im} - \beta_{\mathcal{B}m}\mathcal{B}_i - \beta_{\mathcal{P}m}\mathcal{P}_i}{2\sigma_m^2}\right) \tag{3.12}$$

If outcome Y_{im} is binary, then Y_i^{*} is a latent variable and the observable outcome variable Y_i can be considered as an indicator with $Y_i = 1$ only if $Y_i^{*} > 0$ and $Y_i = 0$ otherwise. $u_{Y_{im}}$ is assumed to be distributed according to a logistic distribution, then the probability of this binary outcome is:

$$Pr(Y_{im}|X_{im}, \mathcal{B}_i, \mathcal{P}_i) = \frac{[exp(\beta_{xm} X_{im} + \beta_{\mathcal{B}m}\mathcal{B}_i + \beta_{\mathcal{P}m}\mathcal{P}_i)]^{Y_{im}}}{1 + exp(\beta_{im} X_{im} + \beta_{\mathcal{B}m}\mathcal{B}_i + \beta_{\mathcal{P}m}\mathcal{P}_i)} \tag{3.13}$$

The likelihood of all the observed outcomes is then:

$$L(Y_i|X_i, \mathcal{B}_i, \mathcal{P}_i) = \prod_{m=1}^{M} Pr(Y_{im}|X_{im}, \mathcal{B}_i, \mathcal{P}_i) \tag{3.14}$$

And the log-likelihood function is

$$\begin{aligned} \mathcal{L}_{\substack{Outcomes\ without \\ instruments}} &= \sum_{i=1}^{n} ln\left(\int\int L(Y_i|X_i, \mathcal{B}_i, \mathcal{P}_i) f(\mathcal{B}) f(\mathcal{P}) d(\mathcal{B}) d(\mathcal{P})\right) \\ &= \sum_{i=1}^{n} ln \int L(Y_i|X_i, \theta) f(\theta) d\theta \end{aligned} \tag{3.15}$$

3.5.3 Instrumental Variables

The challenge for identifying Equation 3.11 is that victimization is endogenous in terms of individual observable and unobservable characteristics. As discussed in the Literature Review (Section 3.2) and Summary Statistics (Section 3.3), victimization is not randomly selected in terms of observable and unobservable factors. To address the possible endogeneity of victimization, I use an instrumental variable approach adapted in a latent factor model by instrumenting victimization with the number of the child's friends being physical bullied and the percentage of children in the Young Lives survey clusters being physically punished by their parents. To be a valid instrument, it requires relevance and exogeneity. It must affect victimization but not directly affect the outcomes of interest. My instruments are inspired by Carrell and Hoekstra (2010) and Eriksen et al. (2014) that domestic violence influences children and their peers, and the number of a child's friends being physically bullied would increase the likelihood of that child being bullied. The first instrument affects bullying victimization as it accounts for the supply of bullying in school environment. The second one relates family emotional trauma with troubled behaviors in school and captures the fact that children from troubled family are more likely to have behavioral challenges (Wolfe et al., 2003; Eriksen et al., 2014). It is reasonable to assume that the two instrumental variables affect the outcomes of interest through their effects on bullying victimization. I discuss more about the relevance in the result section.

I implement a two-stage instrument variable estimation as follows:

The first stage:

$$\mathcal{B}_i = \beta_x X_i + \beta_p Troubledfamily_i + \beta_f Troubledfriend_i + u_{\mathcal{B}_i} \tag{3.16}$$

Where $Troubledfamily_i$ is the percentage of children in the Young Lives survey clusters being physically punished by their parents, $Troubledfriends_i$ is the number of the child's friends being physically bullied and $u_{\mathcal{B}_i}$ is an error term.

Where $u_{\mathcal{B}_i}$ follows a normal distribution with mean zero and variance $\sigma^2_{\mathcal{B}}$, then

$$L(\mathcal{B}_i|X_i, prbeat_i, frbeat_i) = \frac{1}{\sqrt{2\sigma^2_{\mathcal{B}}\pi}} exp\left(-\frac{(\mathcal{B}_i - \beta_x X_i - \beta_p Troubledfamily_i - \beta_f Troubledfriends_i)^2}{2\sigma^2_{\mathcal{B}}}\right) \tag{3.17}$$

Second-stage:

$$Y^*_{im} = \beta_{xm} X_{im} + \beta_{\mathcal{B}m}\widehat{\mathcal{B}}_i + \beta_{\mathcal{P}m}\mathcal{P}_i + u_{Y_{im}} \tag{3.18}$$

Applying the same procedure as in Section 3.5.2 with the likelihood function 3.17 for the endogenous victimization and the new outcome equation 3.18, the log-likelihood with instrumental variables is

$$\mathcal{L}_{\substack{Outcomes\,with \\ instruments}} = \sum_{i=1}^{n} ln\left(\int\int L(Y_i|X_i, \widehat{\mathcal{B}}_i, \mathcal{P}_i) L(\mathcal{B}_i|X_i, prbeat_i, frbeat_i) f(\mathcal{B}) f(\mathcal{P}) d\mathcal{B} d\mathcal{P}\right) \tag{3.19}$$

3.5.4 Estimation

I first estimate the system of Equations 3.2, 3.4 and 3.6 with the log-likelihood function 3.10 by maximum likelihood method. I implement the estimation using the minorization-maximization algorithm described in Chapter 1 to maximize Equation 3.10.

Once the parameters of the measurement system and distribution of the factors, $f(\mathcal{B})$ $f(\mathcal{P})$ or $f(\theta)$, are identified, I can estimate the outcome models 3.11, 3.16 and 3.18 by maximizing the log-likelihood function 3.15 and 3.19 . I draws R values of θ from the conditional distributions of θ, then the log-likelihood of the outcome Equations 3.15 and 3.19 become:

$$\begin{aligned} \mathcal{L}_{\substack{Outcomes\,without \\ instruments}} &= \sum_{i=1}^{n} \frac{1}{R} \sum_{r=1}^{R} ln(L(Y_i|X_i, \mathcal{B}_i, \mathcal{P}_i)) \\ &= \sum_{i=1}^{n} \frac{1}{R} \sum_{r=1}^{R} ln(L(Y_i|X_i, \theta)) \end{aligned} \tag{3.20}$$

and

$$\mathcal{L}_{\substack{Outcomes\ with \\ instruments}} = \sum_{i=1}^{n} \frac{1}{R} \sum_{r=1}^{R} ln(L(Y_i|X_i, \widehat{\mathcal{B}}_i, \mathcal{P}_i)L(\mathcal{B}_i|X_i, prbeat_i, frbeat_i)) \tag{3.21}$$

Since $\mathcal{B}$, $\mathcal{P}$ or θ are treated as observable data, outcome Equations 3.11, 3.16 and 3.18 can be estimated by standard OLS and logistic regression methods.

3.6 Model Results

In this section, I first present and discuss the characteristics of the measurement system. This system identify the distributions of the latent factors (bullying victimization and family factors) and the estimated parameter of the measurement system. I then present the determinants of bullying victimization. Lastly, I present and discuss the consequences of victimization on health and psychosocial outcomes.

3.6.1 Measurement System

Table 3.6 shows the estimates of the measurement system. Besides the loadings, I also calculate and report the average marginal effects (AME) of one standard deviation increase in the factors. The results show that the loadings that explain the contributions of the factors to the measures are large and statistically different from zero at the 1% significance level. The latent bullying victimization factor is more likely associated with verbal bullying and property attack, while the latent family factor is more likely related to parental education. The negative signs show the negative impacts of the factors on corresponding variables. In particular, a one standard deviation increase in the bullying victimization factor would, on average, increase the probability of having physical, verbal, relational and property attack bullying by 22.1%, 22.6%, 15.6% and 33.8% respectively. Increasing the family factor by a one standard deviation is associated with an average increase in wealth index, mother's education and father's education by 0.22, 3.16, 3.12 units respectively, while this factor is negatively correlated with the number of household members and birth order.

Table 3.6: Estimated Parameters of Measurement System[a]

Factor	Measures	Data type	Intercepts	Loadings	AME[b]	Signal
Bullying Victimization	Physical	Binary	0	1	0.221*** (0.006)	–
	Verbal	Binary	5.347*** (0.117)	7.727*** (0.024)	0.226*** (0.012)	–
	Relational	Binary	5.456*** (0.128)	6.900*** (0.030)	0.156*** (0.012)	–
	Property	Binary	1.929*** (0.133)	4.753*** (0.030)	0.338*** (0.013)	–
Family	Wealth index	Continuous	0	1	0.122*** (0.004)	0.430*** (0.017)
	Mother's education	Continuous	-9.978*** (0.571)	25.989*** (0.919)	3.164*** (0.094)	0.659*** (0.029)
	Father's education	Continuous	-9.160*** (0.517)	25.592*** (0.794)	3.116*** (0.100)	0.577*** (0.029)
	Hosuehold size	Continuous	5.793*** (0.261)	-2.018*** (0.404)	-0.246*** (0.049)	0.033*** (0.012)
	Birth order	Continuous	4.298*** (0.233)	-3.379*** (0.349)	-0.411*** (0.042)	0.095*** (0.017)
	Mother and/or father alive	Binary	1.231*** (0.457)	2.307*** (0.768)	0.016*** (0.005)	–

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

[a] The factor moments, including factor means, factor standard deviations, correlation and mixture component means, are provided in Appendix B.

[b] Average Marginal Effects of Factors.

3.6.2 The Determinants of Victimization and its Consequences on Outcomes

Table 3.7 presents the first-stage results. The number of the child's friends being physical bullied and the percentage of children being physically punished by their parents have a positive and significant effect on being bullied. One unit change in the number of a child's friends being physical bullied would induce an increase in victimization by 0.55 standard deviations and one percent change in the percentage of children being physically punished by their parents would increase bullying victimization by 0.2 standard deviations.

Table 3.7: Determinants of Bullying Victimization

Variables	Bullied
Intercept	-2.923***
	(0.890)
Female	0.105**
	(0.042)
Urban	0.082
	(0.054)
Ethnic group	0.126*
	(0.066)
Child age (in years)	0.049
	(0.060)
Troubled family	0.553***
	(0.198)
Troubled friends	0.200***
	(0.016)
First Stage F-Statistics	244
No. of obs.	971

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table 3.8 presents the second-stage estimates for the consequences of bully victimization on the different outcomes. Besides reporting the estimated coefficients, the average marginal effects of one standard deviation increase in bullying victimization and family factors, holding the other variable constant are also reported. The results show that while bullying victimization does not affect self-rated health, it have statistically significant effects on life satisfaction, alcohol use and distress. My findings indicate that increasing the bullying factor by one standard deviation would reduce the probability of having a good life on average by 8.7 percentage points, which represent an substantial decrease of about 21.75% over the baseline probability. In the same way, a one standard deviation increase in the bullying victimization factor would increase the incidence of drinking alcohol by 13.5 percentage points, a significant increase of 45.6%. A one standard deviation increase in the bullying victimization factor would increase the distress index by 0.284, equivalent an increase of 16.4% relative to the baseline value. Although estimation approaches are different, these results are aligned with Sarzosa and Urzúa (2021) and are similar in magnitude.

These findings differ from the results found in the reduced form regressions in Section 3.4 which ignore the endogeneity and the measurement errors for bullying victimization reports and show that no more than two types of victimization affect the outcomes. Appendix C presents the results for the specification that corrects measurement errors, but without the instruments. The results indicate that the consequences of bullying victimization on the outcomes are weaker and it shows endogeneity biases and the importance of instrumental variables.

As discussed in the literature, previous studies show negative effect of victimization on health and psychosocial outcomes, but provided limited evidence since they ignore the endogeneity and measurement error issues that cause different forms of bias. Studies considering these issues are rare. It makes comparisons between findings difficult. Our findings contribute substantially to the literature on peer victimization and health and psychosocial outcomes in this regard.

To understand the size and significance of consequence of the bullying victimization, the model needs to be simulated from the estimated results above. In this sense, the bullying victimization and family factors are randomly drawn from the estimated distributions of these factors in the first steps, these draws are paired with individuals and their controls, and estimated parameters are used to get expected outcomes as a function of the victimization factor in terms of deciles of its distribution. This way, we can see the consequences of victimization on the outcomes of interest. Figure 3.1 presents the results of theses exercises.

Table 3.8: Consequences of Bullying Victimization on Health and Psychosocial Outcomes

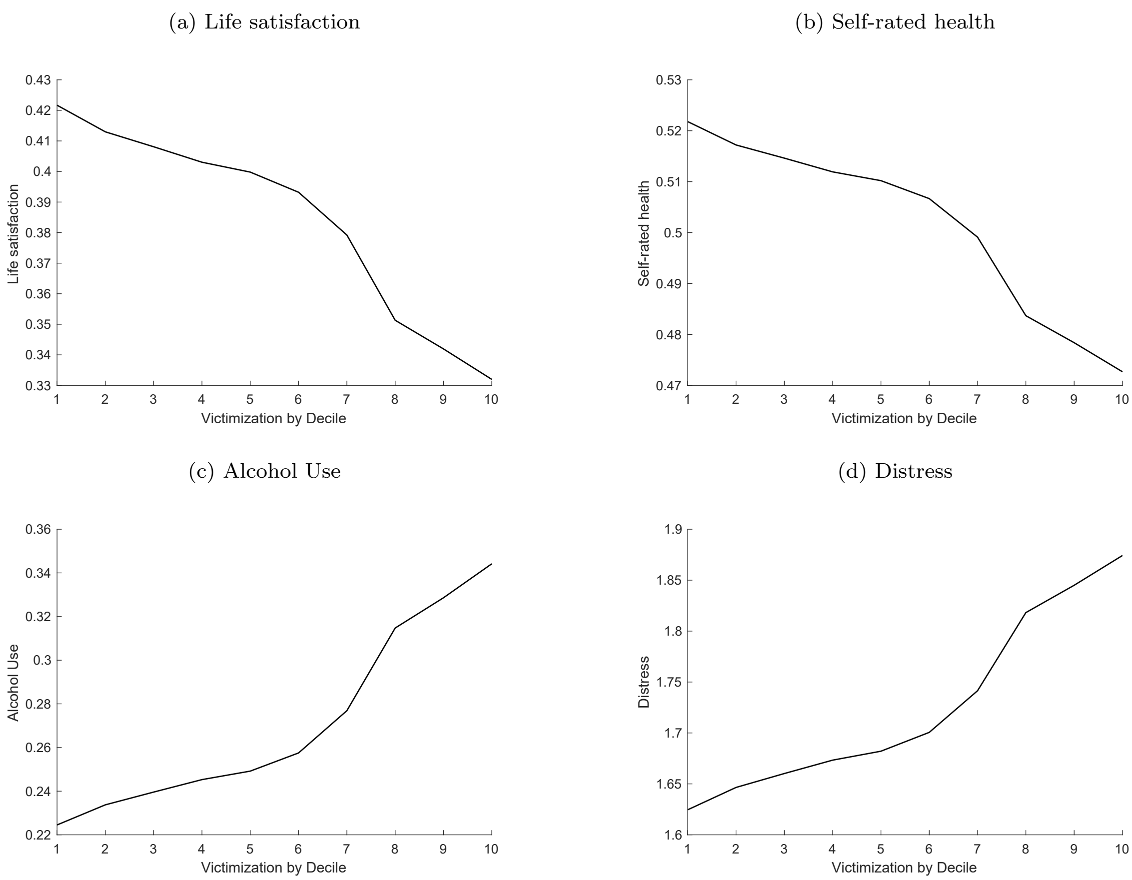
Variables	Life satisfaction	Self-rated health	Alcohol use	Distress
Intercept	-0.846	-2.164	-1.702	0.774
	(2.441)	(2.253)	(3.002)	(0.474)
Female	0.306***	0.045	-0.956***	0.129***
	(0.112)	(0.106)	(0.113)	(0.022)
Urban	-0.074	0.065	-0.492***	-0.088***
	(0.137)	(0.148)	(0.171)	(0.023)
Ethnic group	0.840***	-0.034	-0.220	0.106***
	(0.225)	(0.168)	(0.170)	(0.039)
Child age (in years)	-0.255	0.097	0.150	0.103***
	(0.160)	(0.147)	(0.194)	(0.031)
Bully	-0.226**	-0.116	0.354***	0.149***
	(0.096)	(0.083)	(0.103)	(0.019)
Family	4.343***	0.587	0.389	-0.493***
	(0.503)	(0.401)	(0.540)	(0.082)
Bully AME	-0.087***	-0.054	0.135***	0.284***
	(0.033)	(0.038)	(0.035)	(0.036)
Family AME	0.114***	0.018	0.009	-0.060***
	(0.012)	(0.012)	(0.012)	(0.010)
Baseline Probability/average value	0.400	0.502	0.296	1.728
No. of obs.	969	969	960	961

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

It shows negative gradients between victimization and positive outcomes (life satisfaction and health) and positive gradients between being bullied and adverse outcomes (alcohol use and distress). Figure 3.1a shows that the probability of being satisfied with life moves from 42% to 33% across the deciles of victimization. Likewise, although I did not find significant consequences of being bullied on health, Figure 3.1b demonstrates that probability of having a good health changes from 52% in the first decile of victimization to 47% with highly-victimised children. Figure 3.1c shows that the incidence of drinking alcohol increase from

23% to 34% across the distribution of victimization. Similarly, the distress index moves from 1.63 in the first decile to 1.87 in the last decile of victimization (Figure 3.1d).

Figure 3.1: Outcomes by Deciles of the Victimization Factor Distribution



3.7 Conclusion

I use a structural model combined with an instrumental variable strategy to deal with the endogeneity and measurement error issues of bullying to examine the consequences of peer victimization on a range of health, risky behavior and well-being indicators. The findings indicate that peer victimization strongly affects subjective well-being, alcohol consumption, and emotional and mental distress of children. My results are consistent with evidence from both developed and developing countries that bullying has strong consequences on health risks and psychosocial outcomes. I do not find evidence of associations between bullying victimization and self-rated health.

Adolescence is a critical period of development in which youth is extremely sensitive to bullying acts, stress, risky behaviors, physical and mental health, and well-being indicators that can affect the developmental trajectories of individuals. Bullying victims suffer long-lasting consequences in terms of health and psychosocial development over the life course. My research provides scarce evidence about the effects of victimization on various health and psychosocial outcomes, especially in low-resource settings. My results about solid and consistent associations between peer victimization and health risky behaviors and well-being of children highlight the need to increase awareness of, identify and recognize different types of bullying as a serious issue and have interventions to prevent these modifiable behaviors. It is also critical to mobilize protective resources and efforts and develop adequate education policies to curb school victimization, especially in developing countries.

Appendix A: Variable Description

Table A.1: Description of Variables

Variables	Description
Covariates	
Female	Binary variable equal to 1 for girls and 0 for boys
Urban	Binary variable equal to 1 if the child's household resides in urban areas and 0 otherwise
Ethnic group	Binary variable equal to 1 for Kinh ethic group and 0 for the other ethic groups
Child age	Child age in years
Troubled family	The percentage of children in the Young Lives survey clusters being physically punished by their parents in the last 12 months
Troubled friend	The number of a child's best friends being physical bullied
Overall victimization	Binary variable equal to 1 if a child has been victimized of any kinds in the 9-item Social and Health Assessment Peer Victimization Scale. See Section 3.3.1 and Table 3.1
Bullying victimization factor	
Physical victimization	Binary variable equal to 1 if a child has experienced the following bullying acts at least two times: 1) punched, kicked or beat up; 2) hurt physically in some other way and 0 otherwise.

Continued on the next page

Table A.1 – Description of Variables *Continued*

Variables	Description
Verbal victimization	Binary variable equal to 1 if a child has experienced the following bullying acts at least two times: 1) made fun of for some reason; 2) called names or swore at and 0 otherwise.
Relational victimization	Binary variable equal to 1 if a child has experienced the following bullying acts at least two times: 1) refused to talk to you or made other people not talk to you; 2) cried to get you into trouble with your friends; 3) made you uncomfortable by staring at you for a long time and 0 otherwise.
Attacks on Property	Binary variable equal to 1 if a child has experienced the following bullying acts at least two times: 1) took something without permission or stole things from you; 2) tried to break or damaged something of yours and 0 otherwise.
Family factor	
Wealth Index	The wealth index is a composite measure of living standards, it is the average of the three sub-indexes: consumer durable, housing quality and access to service indexes. It takes values from 0 to 1, a higher value reflect a wealthier household.
Mother's education	Mother's years of education
Father's education	Father's years of education
Household size	Number of household members living in the household of the child
Border	Birth order of the child in the family

Continued on the next page

Table A.1 – Description of Variables *Continued*

Variables	Description
Momdadlive	Indicate whether both mother and father live in the householde or not. It equals 1 if both parents live in the child's household and 0 otherwise.
Outcomes	
Self-rated health	This variable is based on interviewer-administered question asking youth to rate their general health on a 5-point Likert scale with which 1 indicates very poor and 5 very good health status. The response is then dichotomized as 0 indicating poor health status if youth reported their health below the mean value of 3.36 and 1 denoting good health the reported response value above the mean value.
Subjective well-being	This variable is based on the question 'where on the ladder do you feel you personally stand at present time?' and Responses range from 1-9, higher scores indicated better subjective well-being. The responses are further recoded as 0 indicating low subjective well-being and 1 denoting high subjective well-being based on the mean value of the response: 5.59.
Alcohol consumption	Dummy variable equal to 1 for those who reported alcohol use at least once a month and 0 otherwise. This variable is based on youth's response to the question in the self-administered questionnaire: 'How often do you usually drink alcohol?'.

Continued on the next page

Table A.1 – Description of Variables *Continued*

Variables	**Description**
Emotional and mental distress	Distress index is the average of the five items of the Emotional Difficulties subscale of the Strengths and Difficulties Questionnaire (SDQ) in the self-administered questionnaire: 'I worry a lot'; 'I get a lot of headaches, stomach aches, or sickness'; 'I am often unhappy, downhearted, or tearful'; 'I am nervous in new situations', and 'I have many fears or are easily scared'.

Appendix B: Factor Distribution

Table B.1: Factor Means, Standard Deviations and Correlation

	Bully	Family
Factore means	-2.862	0.620
	(0.067)	(0.005)
Factor Standard Deviations	1.910	0.122
	(0.029)	(0.004)
Factor Correlation:		
Bully	1	–
Family	0.033	1
	(0.032)	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Table B.2: Mixture Component Means

	Bully	Family	Type share
Type 1	-4.491	0.616	0.572
	(0.057)	(0.006)	(0.014)
Type 2	-0.719	0.677	0.035
	(0.007)	(0.009)	(0.003)
Type 3	-0.730	0.648	0.208
	(0.005)	(0.007)	(0.010)
Type 4	0.833	0.607	0.009
	(0.469)	(0.003)	(0.004)
Type 5	-0.693	0.841	0.033
	(0.014)	(0.015)	(0.006)
Type 6	-0.657	0.440	0.071
	(0.019)	(0.004)	(0.011)
Type 7	-0.733	0.633	0.028
	(0.006)	(0.009)	(0.002)
Type 8	-0.727	0.617	0.045
	(0.008)	(0.007)	(0.003)

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Appendix C: Model Estimates without Endogenous Bullying Victimization

Table c.1: Consequences of Bullying Victimization on Health and Psychosocial Outcomes without Instrumental Variables

	Life satisfaction	Self-rated health	Alcohol use	Distress
Intercept	0.466	-1.927	-2.578	0.274
	(2.413)	(2.191)	(2.950)	(0.440)
Female	0.268**	0.037	-0.952***	0.144***
	(0.107)	(0.107)	(0.118)	(0.020)
Urban	-0.102	0.060	-0.482***	-0.076***
	(0.136)	(0.149)	(0.173)	(0.022)
Ethnic group	0.771***	-0.047	-0.195	0.134***
	(0.220)	(0.166)	(0.161)	(0.035)
Child age (in years)	-0.292*	0.093	0.172	0.115***
	(0.160)	(0.147)	(0.193)	(0.029)
Bully	0.001	-0.068***	0.205***	0.054***
	(0.024)	(0.024)	(0.027)	(0.005)
Family	4.326***	0.567	0.478	-0.478***
	(0.502)	(0.404)	(0.538)	(0.080)
Bully AME	0.000	-0.032***	0.078***	0.104***
	(0.010)	(0.011)	(0.010)	(0.009)
Family AME	0.118***	0.017	0.011	-0.058***
	(0.013)	(0.012)	(0.012)	(0.009)
Baseline Probability/Average Value	0.397	0.502	0.281	1.727
No. of obs.	969	969	960	961

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Bibliography

Abada, T., F. Hou, and B. Ram (2008). The effects of harassment and victimization on self-rated health and mental health among Canadian adolescents. *Social Science & Medicine 67*(4), 557–567.

Acosta, P., N. Muller, and M. A. Sarzosa (2015). *Beyond qualifications: returns to cognitive and socio-emotional skills in Colombia.* The World Bank.

Agostinelli, F. and M. Wiswall (2016). Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. Volume 4, Chapter 1, pp. 1–181. Elsevier.

Arseneault, L., L. Bowes, and S. Shakoor (2010). Bullying victimization in youths and mental health problems:'much ado about nothing'? *Psychological medicine 40*(5), 717–729.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., F. Cunha, and P. Jervis (2019). Subjective parental beliefs. their measurement and role. Technical report, National Bureau of Economic Research.

Attanasio, O., C. Meghir, E. Nix, and F. Salvati (2017). Human capital growth and poverty:

Evidence from Ethiopia and Peru. *Review of Economic Dynamics 25*, 234–259. Special Issue on Human Capital and Inequality.

Attanasio, O. P. (2015). The determinants of human capital formation during the early years of life: Theory, measurement, and policies. *Journal of the European Economic Association 13*(6), 949–997.

Aucejo, E. and J. James (2021). The path to college education: The role of math and verbal skills. *Journal of Political Economy 129*(10), 2905–2946.

Bodewig, C., R. Badiani-Magnusson, K. Macdonald, D. Newhouse, and J. Rutkowski (2014). *Skilling up Vietnam: Preparing the workforce for a modern market economy.* The World Bank.

Borghans, L., A. L. Duckworth, J. Heckman, and B. ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4).

Boulton, M. J. and K. Underwood (1993). Bully/victim problems among middle school children. *European Education 25*(3), 18–37.

Bourdillon, M. and J. Boyden (2014). *Growing up in poverty: findings from Young Lives.* Springer.

Boyes, M. E., L. Bowes, L. D. Cluver, C. L. Ward, and N. A. Badcock (2014). Bullying victimisation, internalising symptoms, and conduct problems in south african children and adolescents: A longitudinal investigation. *Journal of abnormal child psychology 42*(8), 1313–1324.

Boynton-Jarrett, R., L. M. Ryan, L. F. Berkman, and R. J. Wright (2008). Cumulative violence exposure and self-rated health: longitudinal study of adolescents in the united states. *Pediatrics 122*(5), 961–970.

Brown, D. W., L. Riley, A. Butchart, and L. Kann (2008). Bullying among youth from eight african countries and associations with adverse health behaviors.

Cantril, H. et al. (1965). Pattern of human concerns.

Carbone-Lopez, K., F.-A. Esbensen, and B. T. Brick (2010). Correlates and consequences of peer victimization: Gender differences in direct and indirect forms of bullying. *Youth violence and juvenile justice 8*(4), 332–350.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics 2*(1), 211–28.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276. PMID: 26828106.

Cawley, J., J. Heckman, and E. Vytlacil (2001). Three observations on wages and measured cognitive ability. *Labour Economics 8*(4), 419 – 442.

Cluver, L. and M. Orkin (2009). Cumulative risk and aids-orphanhood: Interactions of stigma, bullying and poverty on child mental health in south africa. *Social science & medicine 69*(8), 1186–1193.

Coneus, K., M. Laucht, and K. Reuß (2012). The role of parental investments for cognitive and noncognitive skill formation—evidence for the first 11 years of life. *Economics and Human Biology 10*(2), 189–209.

Crookston, B., R. Forste, C. McClellan, A. Georgiadis, and T. Heaton (2014). Factors associated with cognitive achievement in late childhood and adolescence: The young lives cohort study of children in ethiopia, india, peru, and vietnam. *BMC pediatrics 14*, 253.

Crookston, B. T., R. M. Merrill, S. Hedges, C. Lister, J. H. West, and P. C. Hall (2014). Victimization of Peruvian adolescents and health risk behaviors: young lives cohort. *BMC public health 14*(1), 1–7.

Cunha, F. and J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*(2-3), 320–364.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunningham, W., M. P. Torrado, and M. Sarzosa (2016). *Cognitive and non-cognitive skills for the Peruvian labor market: Addressing measurement error through latent skills estimations*. The World Bank.

Del Boca, D., C. Flinn, and M. Wiswall (2013). Household choices and child development. *The Review of Economic Studies 81*(1), 137–185.

Dercon, S. and P. Krishnan (2009). Poverty and the psychosocial competencies of children: evidence from the young lives sample in four developing countries. *Children Youth and Environments 19*(2), 138–163.

Dercon, S. and A. Sánchez (2013). Height in mid childhood and psychosocial competencies in late childhood: Evidence from four developing countries. *Economics and Human Biology 11*(4), 426 – 432.

Díaz, J. J., O. Arias, and D. V. Tudela (2012). Does perseverance pay as much as being smart? the returns to cognitive and non-cognitive skills in urban Peru. *Unpublished paper, World Bank, Washington, DC*.

Doyle, O., C. P. Harmon, J. J. Heckman, and R. E. Tremblay (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology 7*(1), 1–6.

Drago, F. (2011). Self-esteem and earnings. *Journal of Economic Psychology 32*(3), 480 – 488.

Due, P. and B. E. Holstein (2008). Bullying victimization among 13 to 15 year old school children: Results from two comparative studies in 66 countries and regions. *International journal of adolescent medicine and health 20*(2), 209–222.

Duncan, G., C. Dowsett, A. Claessens, K. Magnuson, A. Huston, P. Klebanov, L. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, H. Sexton, and C. Japel (2007). School readiness and later achievement. *Developmental psychology 43*, 1428–46.

Engle, P. L., M. M. Black, J. R. Behrman, M. C. De Mello, P. J. Gertler, L. Kapiriri, R. Martorell, M. E. Young, I. C. D. S. Group, et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The lancet 369*(9557), 229–242.

Eriksen, T. L. M., H. S. Nielsen, and M. Simonsen (2014). Bullying in elementary school. *Journal of Human Resources 49*(4), 839–871.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, pp. 287–302. Elsevier.

Fosse, N. E. and S. A. Haas (2009). Validity and stability of self-reported health among adolescents in a longitudinal, nationally representative survey. *Pediatrics 123*(3), e496–e501.

GSO (2016). Nang suat lao dong Viet Nam: Thuc trang va giai phap [Vietnam's labor productivity: Situations and solutions].

Glewwe, P., Q. Huang, and A. Park (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural China. *Journal of Economic Behavior & Organization 134*, 141 – 164.

Gobina, I., A. Zaborskis, I. Pudule, I. Kalnins, and A. Villerusa (2008). Bullying and subjective health among adolescents at schools in Latvia and Lithuania. *International journal of public health 53*(5), 272–276.

Goodman, R. and S. Scott (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of abnormal child psychology 27*(1), 17–24.

Green, D. and W. C. Riddell (2003). Literacy and earnings: An investigation of the interaction of cognitive and unobserved skills in earnings generation. *Labour Economics 10*, 165–184.

Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume 4 of *Handbook of Public Economics*, Chapter 30, pp. 2045–2141. Elsevier.

Hanushek, E. A. (2009). *The economic value of education and cognitive skills*, Chapter 3. Routledge.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from piaac. *European Economic Review 73*, 103 – 130.

Hanushek, E. A. and L. Woessmann (2008). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–68.

Hanushek, E. A. and L. Zhang (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital 3*(2), 107–143.

Hawker, D. S. and M. J. Boulton (2000). Twenty years' research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. *The Journal of Child Psychology and Psychiatry and Allied Disciplines 41*(4), 441–455.

Heckman, J., J. Humphries, S. Urzúa, and G. Veramendi (2011). The effects of educational choices on labor market, health, and social outcomes. *Human Capital and Economic Opportunity Working Paper*.

Heckman, J. J. (2008). The case for investing in disadvantaged young children. *CESifo DICE Report 6*(2), 3–8.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Helmers, C. and M. Patnam (2011). The formation and evolution of childhood skill acquisition: Evidence from India. *Journal of Development Economics 95*(2), 252–266.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

James, J. (2017). MM algorithm for general mixed multinomial logit models. *Journal of Applied Econometrics 32*(4), 841–857.

Joffer, J., L. Jerdén, A. Öhman, and R. Flacking (2016). Exploring self-rated health among adolescents: a think-aloud study. *BMC public health 16*(1), 1–10.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1), 141–151.

Karlsson, E., A. Stickley, F. Lindblad, M. Schwab-Stone, and V. Ruchkin (2014). Risk and protective factors for peer victimization: a 1-year follow-up study of urban american students. *European child & adolescent psychiatry 23*(9), 773–781.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of cognitive neuroscience 16*(8), 1412–1425.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006). Economic, neurobiological, and behavioral perspectives on building america's future workforce. *Proceedings of the national Academy of Sciences 103*(27), 10155–10162.

Kottelenberg, M. J. and S. F. Lehrer (2019). How skills and parental valuation of education influence human capital acquisition and early labor market return to human capital in canada. *Journal of Labor Economics 37*(S2), S735–S778.

Krishnan, P. and S. Krutikova (2013). Non-cognitive skill formation in poor neighbourhoods of urban India. *Labour Economics 24*, 68 – 85.

Ladd, G. W., I. Ettekal, and B. Kochenderfer-Ladd (2017). Peer victimization trajectories from kindergarten through high school: Differential pathways for children's school engagement and achievement? *Journal of Educational Psychology 109*(6), 826.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review 10*(2), 179–214.

Lereya, S. T., W. E. Copeland, E. J. Costello, and D. Wolke (2015). Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *The Lancet Psychiatry 2*(6), 524–531.

Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–28.

Lister, C., R. M. Merrill, D. Vance, J. H. West, P. C. Hall, and B. T. Crookston (2015a). Predictors of peer victimization among peruvian adolescents in the young lives cohort. *International journal of adolescent medicine and health 27*(1), 85–91.

Lister, C. E., R. M. Merrill, D. L. Vance, J. H. West, P. C. Hall, and B. T. Crookston (2015b). Victimization among Peruvian adolescents: Insights into mental/emotional health from the young lives study. *Journal of school health 85*(7), 433–440.

Long, M. C., D. Goldhaber, and N. Huntington-Klein (2015). Do completed college majors respond to changes in wages? *Economics of Education Review 49*, 1 – 14.

Malhi, P., B. Bharti, and M. Sidhu (2014). Aggression in schools: Psychosocial outcomes of bullying among Indian adolescents. *The Indian Journal of Pediatrics 81*(11), 1171–1176.

Murnane, R. J., J. B. Willett, M. Braatz, and Y. Duhaldeborde (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? evidence from the NLSY. *Economics of Education Review 20*(4), 311 – 320.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management 19*(4), 547–568.

Mynard, H. and S. Joseph (2000). Development of the multidimensional peer-victimization scale. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression 26*(2), 169–178.

Nguyen, A. J. et al. (2016). A latent class approach to understanding experiences of bullying victimization among youth in four low-resource settings.

Nguyen, N. (2008, 01). An assessment of the young lives sampling approach in vietnam. *University of Oxford, Open Access publications from University of Oxford*.

Nordman, C. J., L. Sarr, and S. Sharma (2015). Cognitive, non-Cognitive skills and gender wage gaps: Evidence from linked employer-employee data in Bangladesh. IZA Discussion Papers 9132, Institute of Labor Economics (IZA).

OECD (2013). *OECD guidelines on measuring subjective well-being.* OECD Publishing, Paris.

Olweus, D. (1993). Bullying at school: What we know and what we can do. *Malden, MA: Blackwell Publishing*.

Olweus, D. (1996). Bully/victim problems in school. *Prospects 26*(2), 331–359.

Olweus, D. (1997, 12). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education 12*, 495–510.

O'Neill, J. (1990). The role of human capital in earnings differences between black and white men. *Journal of economic Perspectives 4*(4), 25–45.

Ouellet-Morin, I., C. L. Odgers, A. Danese, L. Bowes, S. Shakoor, A. S. Papadopoulos, A. Caspi, T. E. Moffitt, and L. Arseneault (2011). Blunted cortisol responses to stress signal social and behavioral problems among maltreated/bullied 12-year-old children. *Biological psychiatry 70*(11), 1016–1023.

Prada, M. F. and S. Urzúa (2017). One size does not fit all: Multiple dimensions of ability, college attendance, and earnings. *Journal of Labor Economics 35*(4), 953–991.

Rigby, K. (2002). *New perspectives on bullying.* Jessica Kingsley Publishers.

Rivers, I. and P. K. Smith (1994). Types of bullying behaviour and their correlates. *Aggressive behavior 20*(5), 359–368.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton University Press.

Roseth, V. V., A. Valerio, and M. Gutierrez (2016). *Education, skills, and labor market outcomes: Results from Large-Scale Adult Skills Surveys in urban areas in 12 countries.* World Bank.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied 80*(1), 1.

Ruchkin, V., M. Schwab-Stone, and R. Vermeiren (2004). Social and health assessment (saha): psychometric development summary. *New Haven: Yale University*.

Sahn, D. E. and K. M. Villa (2015). The role of personality, cognition and shocks in determining age of entry into labor market, sector of employment, and within sector earnings. (330-2016-13805), 83.

Sánchez, A. (2013). The structural relationship between nutrition, cognitive and non-cognitive skills: evidence from four developing countries. Young Lives Working Paper 111.

Sánchez, A. (2017). The structural relationship between early nutrition, cognitive skills and non-cognitive skills in four developing countries. *Economics and Human Biology 27*, 33–54.

Sánchez, A. and A. Singh (2018). Accessing higher education in developing countries: Panel data analysis from India, Peru, and Vietnam. *World Development 109*, 261 – 278.

Sarzosa, M. and S. Urzúa (2021). Bullying among adolescents: The role of skills. *Quantitative Economics 12*(3), 945–980.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Singh, A. (2019). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*. jvz033.

Smith, P. K. and P. Brain (2000). Bullying in schools: Lessons from two decades of research. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression 26*(1), 1–9.

Smith, P. K., L. Talamelli, H. Cowie, P. Naylor, and P. Chauhan (2004). Profiles of non-victims, escaped victims, continuing victims and new victims of school bullying. *British journal of educational psychology 74*(4), 565–581.

Stadler, C., J. Feifel, S. Rohrmann, R. Vermeiren, and F. Poustka (2010). Peer-victimization and mental health problems in adolescents: are parental and school support protective? *Child Psychiatry & Human Development 41*(4), 371–386.

Takizawa, R., B. Maughan, and L. Arseneault (2014). Adult health outcomes of childhood bullying victimization: evidence from a five-decade longitudinal british birth cohort. *American journal of psychiatry 171*(7), 777–784.

Thiel, H. and S. L. Thomsen (2013). Noncognitive skills in economics: Models, measurement, and empirical evidence. *Research in Economics 67*(2), 189 – 214.

UNESCO (2019). *Behind the numbers: Ending school violence and bullying*. United Nations Educational, Scientific and Cultural Organization.

Urzúa, S. (2008). Racial labor marketgaps: The role of abilities and schooling choices. *The Journal of Human Resources 43*(4), 919–971.

VandenBos, G. R. (2007). *APA dictionary of psychology*. American Psychological Association.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted jamaican children: prospective cohort study. *The lancet 366*(9499), 1804–1807.

WHO (2019). *Global status report on alcohol and health 2018*. World Health Organization.

Wolfe, D. A., C. V. Crooks, V. Lee, A. McIntyre-Smith, and P. G. Jaffe (2003). The effects of children's exposure to domestic violence: A meta-analysis and critique. *Clinical child and family psychology review 6*(3), 171–187.

The data used in this thesis comes from Young Lives, a 20-year study of childhood poverty and transitions to adulthood in Ethiopia, India, Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Foreign, Commonwealth and Development Office and a number of further funders. The views expressed here are those of the author. They are not necessarily those of Young Lives, the University of Oxford, FCDO or other funders.